UNIVERSITY OF NAIROBI

GENETIC ANALYSIS OF RESPIRATORY HUMAN ADENOVIRUSES AMONG PEDIATRIC PATIENTS ATTENDING NEW NYANZA PROVINCIAL GENERAL HOSPITAL IN THE PERIOD JUNE 2010 TO JUNE 2012

MSc. Thesis by:

Mbui M. Fredrick

Reg No: H56/67898/2011

Department of Biochemistry

School of Medicine, College of Health Sciences

University of Nairobi

A thesis submitted in partial fulfillment of the requirement for the award of the degree of Masters of Science in Biochemistry of University of Nairobi

NOVEMBER 2014

DECLARATION

This is my original work and it has not been presented in any other University

Name **Mbui M. Fredrick**

Reg.No: H56/67898/2011

Signature……..

Date……..

This work has been submitted for examination with our approval as University of Nairobi Supervisors.

Prof. Wallace D. Bulimo

Department of Biochemistry

School of Medicine

University of Nairobi

Signature……..

Date……..

Rachel Achilla

DEID Human Influenza Program

US Army Medical Research Unit - Kenya

Signature……..

Date…….

ENDORSEMENT

This work has been submitted for examination with the approval of Department of Biochemistry University of Nairobi.

The Chairman

Department of Biochemistry

School of Medicine

University of Nairobi

Signature……..

Date……..

DEDICATION

This work is dedicated to my dear parents for the sacrifice they made to put me through school and for always believing in my academic potential.

ACKNOWLEDGEMENT

To Dr. Bulimo, I am very grateful for your exceptional supervision and guidance as I prepared the proposal for this study and while carrying out the laboratory experiments and data analyses. I appreciate the sacrifice you made on your busy schedule so that you can create time and guide me while compiling this document.

To Ms Rachel Achilla I am also grateful for your supervision and guidance while carrying out this project.

My appreciation also goes to the entire staff of National Influenza Center laboratory for your invaluable assistance as I carried out the serological and molecular biology procedures.

I appreciate Global Emerging Infectious surveillance (GEIS) for generously funding this project.

To the citizens and the taxpayers of the Federal Republic of Germany through DAAD, I am humbled by your generosity in kind and financial support. I sincerely thank you for sponsoring my Masters course through the DAAD In-country master's scholarship program.

I appreciate Prof Marcus Frohme and Prof Heike Pospisil of the Technische Hochschule Wildau *University of Applied Sciences* for welcoming me to your institution. I am grateful to you and your bioinformatics department staff members for training me on the use of bioinformatics in biological data analyses. I greatly relied on those skills while analyzing data in this project.

Lastly, I thank the almighty God, my parents, brother, sisters and my fiancée Virginia for your constant encouragement without which this work would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMES

Adv	Adenovirus
ARD	Acute Respiratory Disease
BLOSUM	Block Substitution Matrix
CAR	Cox-sackie and Adenovirus Receptors
CD46	Cluster of differentiation 46
CPE	Cytopathic Effect
dDNTPs	di-deoxynucleotide triphosphates
DFA	Direct Fluorescent Antibody
DNA	Deoxyribonucleic Acid
DEID	Department of Emerging Infectious Disease
DMEM	Dulbecco's Modified Eagle Medium
EMBOSS	European Molecular Biology Online Software Suit
HAdV	Human Adenovirus
IFA	Immunofluorescence Assay
Kb	Kilo base pair
KEMRI	Kenya Medical Research Institute
MCMC	Markov chain Monte Carlo
MLP	Major-late promoter
ORF	open reading frame
PCR	Polymerase Chain Reaction
RGD	Arginine-Glycine-Asparagine
USAMRU-K	United States Army Medical Research Unit – Kenya
WRAIR	Walter Reed Army Institute of Research

ABSTRACT

Human adenoviruses (HAdVs) are common pathogens associated with diseases affecting the respiratory tract, gastro-intestinal tract as well as various organs like the liver, kidney and the brain. Currently there are 60 human adenovirus serotypes classified into 7 species A to G on the basis of serology, genome sequencing and phylogenomics. HAdV species B, C and E are mainly implicated in respiratory tract infections whereas the other species are associated with gastro-intestinal, genitourinary, and ocular infections. The respiratory HAdV species play a significant role in pediatric infections accounting for 10% of overall respiratory illnesses and 5%–11% of pneumonia cases. The burden of diseases due to respiratory adenoviruses in Kenya has not been studied. There is no documented data on respiratory human adenovirus species and serotypes circulating in the country.

The aim of this study was to characterize respiratory human adenoviruses using serological and molecular approaches. Specifically the study sought to determine the species and serotypes of HAdVs that were associated with pediatric respiratory infections in New Nyanza Provincial General Hospital in the period of June 2010 to June 2012.

HAdVs were isolated from 16 archived nasopharyngeal swab patient specimens after inoculation into cultured Hep2 cells. After cytopathic effect observation, presence of adenovirus was confirmed through immunofluorescence assay. Virus DNA was extracted from the isolates followed by PCR amplification of the hyper-variable region 7 located in loop 2 region of hexon gene. The amplicons were sequenced using the Sanger dideoxy termination method followed by bioinformatics analyses of the nucleotide sequences. The hexon nucleotides sequences were used to classify the HAdV isolates into various species and serotypes by use of phylogenomics.

Nucleotide sequences of hexon loop-2 fragment of approximately 500 base pairs were obtained. Multiple sequence alignment of the loop-2 sequences with reference sequences obtained from other parts of the world revealed the location of the hyper-variable region at 1330-1400bp; a region characterized by several nucleotide substitutions, insertions and deletions. Phylogenomics analysis showed that during the study period, respiratory HAdV species B and C were associated with respiratory infections among pediatric patients attending New Nyanza Provincial General Hospital accounting for approximately 1% of the overall respiratory viruses. There were no cases of infections caused by respiratory HAdV E implying that this species was not in circulation during the study period. HAdV C was the predominant species accounting for 68.75% of the reported cases with serotype distribution as HAdV C1-25%, HAdV C2-25%, HAdV C5-6.25%, and HAdV C6-12.5%. HAdV B serotype 7 was the most prevalent serotype at 31.25%. HAdV C5 and HAdV B7 sequences were found to be under positive selection pressure indicating that these viruses are undergoing an evolutionary process which signifies instability in their genomes.

Characterization of respiratory human adenoviruses that circulated at New Nyanza Provincial General hospital during the study period revealed that species B and C were present but not E. There was significant genetic variation in the hexon gene of the HAdVs seen at this site compared to those from other parts of the world implying continuing evolution of respiratory HAdVs. To gain a complete understanding of this evolutionary process, whole genome sequencing of these viruses is called for in order to determine genetic stability and uniqueness of these viruses.

CHAPTER ONE

1.0 INTRODUCTION

Adenovirus is a double stranded DNA virus first isolated in 1953 (Rowe, 1955) from adenoid tissue–derived cell cultures, hence the name. Human adenovirus (HAdV) belongs to the genus *Mastadenovirus* in the Adenoviridiae family of viruses. It is a non enveloped virus with an icosahedral structure and a genome size of between 26-45kb. There are 60 known human adenovirus serotypes classified into 7 species (subgroups A to G) based on serological assays, genome sequencing and phylogenomics (http://www.vmri.hu/~harrach/AdVtaxlong.htm; Kaneko *et al*., 2009; Robinson *et al.*, 2013; Walsh *et al*., 2010). Human adenovirus species B is further described into two subspecies; B1 which includes HAdV type 3, 7, 16, 21, 50 and B2 comprising of HAdV types 11, 14, 34, 35, and 55 (Russell, 2009).

These viruses cause a broad spectrum of diseases including pharyngitis, acute respiratory diseases (ARD), pneumonia, pharyngoconjunctival fever, epidemic keratoconjunctivitis genitourinary tract infections and gastroenteritis. Most of these infections are self limiting but severe or lethal infections can occur especially in infants (Cevenini *et al*., 1985; Cusi, *et al*., 1986) young children and immune compromised patients (Munoz *et al.,* 1998). Adenoviruses have also been a major concern in overcrowded military recruitment camps. They have been associated with respiratory illnesses in military training camps and are major cause of morbidity. Adenoviruses have also been associated with death of military personnel (Potter *et al.,* 2012).

Each species of adenovirus has been associated with different spectra of diseases and severity of infection and therefore rapid identification of a particular virulent serotype can help in prevention and disease management strategies. Species B, C and E are usually associated with respiratory infections whereas the other species are associated with gastrointestinal, genitourinary and ocular infections (Sharma *et al.*, 2009). Table 1 below illustrates the disease conditions associated with particular species and serotypes of human adenoviruses. Adenovirus species A-G and their respective serotypes are given together with disease conditions they are mainly associated with. 13

Table 1 Classification of human adenoviruses into species and serotype specific to various infections. Adopted and updated from Anurag Sharma et.al 2009, and Robinson et.al 20

Species	Infection	Serotypes
A	Meningitis	12,18,31
B	Respiratory ,Ocular and Renal urinary tract	3,7,11,14,16,21,34,35,50,55
C	Respiratory, Hepatitis	1,2,5,6,57
D	Ocular, Meningitis	8,9,10,13,15,17,19,20, 22-30,32,33,36-39, 42-49,51,53,54,56,58,59,60
E	Respiratory	4
F	Gastrointestinal	40,41
G	Gastrointestinal	52

Subspecies B1 serotype 3, 7, and 21 are responsible for most of the adenovirus lower respiratory tract infections epidemics (Moro *et al.*, 2009). Conversely, species C serotypes 1, 2, 5 & 6 and species E serotype 4 are common causes of pediatric upper and lower respiratory tract infections (Erdman *et al.*, 2002; Kajon *et al.*, 1996; Singh-Naz *et al.*, 1993). In otherwise healthy children most adenovirus respiratory infections are mild and indistinguishable from other viral infections. However fatality has been associated

with specific serotypes such as species B serotypes 3, 7, and 14 (Kim *et al*., 2003; Lewis *et al*., 2009; O'Flanagan *et al.*, 2011). Serotype identification is therefore critical for epidemiological surveillance, detection of new strains and understanding of HAdV pathogenesis of clinical isolates. Initially, diagnosis and determination of adenovirus type involved virus isolation in cell culture followed by antibody/antigen detection by immunofluorescence assay and visualization by electron microscopy (Leland *et al.,* 2007). However there are now innovative and advanced molecular methods for rapid detection of different adenovirus serotypes which are based on the nucleotide sequencing of adenovirus genome (Lu *et al.*, 2006). In 2006, Lu and Erdman demonstrated that almost all adenovirus serotypes excluding serotypes 15 and 17 could be identified by using PCR and nucleotide sequencing of adenovirus hyper variable region of hexon gene. HAdVs have also been identified using other nucleotide sequences of other genes including virus-associated (VA) RNA gene (Kidd *et al.*, 1993), the polypeptide IX (pIX) gene (Akalu *et al*,, 1998) or the penton and fiber genes (Xu *et al*., 2000).

In Kenya there is limited data on respiratory human adenovirus infections including their serotype distribution. Molecular characterization data of respiratory human adenovirus is crucial in patient care and management of respiratory diseases since certain species and serotypes have been associated with specific disease syndromes, epidemiological setting and demographic risk groups. Furthermore, molecular changes in adenovirus genome including homologous recombination are a major driving force in molecular evolution of human adenovirus and emergence of new adenovirus pathogens (Robinson *et al.*, 2011). Adenovirus is also a popular vaccine vector (Chen *et al*., 2010), and basic understanding of circulating adenovirus serotypes in the Kenyan population is important to any

researcher interested in formulating adenovirus vectored vaccines for the Kenyan population. This study carried out genetic analysis by PCR, sequencing and phylogenomics of respiratory human adenoviruses isolated from pediatric patients attending New Nyanza Provincial General Hospital in the period June 2010 to June 2012. The data generated from this study contributes to knowledge about respiratory adenoviruses circulating in Kenya.

1.2.0 LITERATURE REVIEW

Adenoviruses are medium sized (80–120nm) in diameter, non-enveloped viruses with an icosahedral structure. The linear double stranded DNA genome is found inside a nucleocapsid whose composition is mainly hexon, penton and fiber proteins as shown in Fig. 1 below.(MicrobeWiki, 2013)

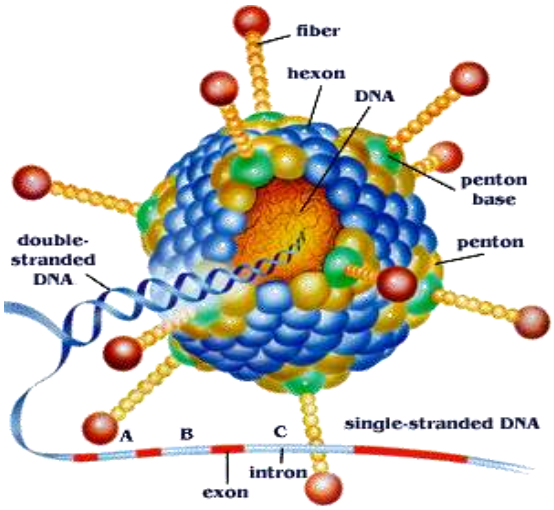


Figure 1. Adenovirus structure. The figure demonstrates the icosahedral shape of adenovirus showing the location of the major capsid proteins enclosing the double stranded DNA genome

1.2.1 Epidemiology of respiratory human adenoviruses

Respiratory adenoviruses infects a wide variety of vertebrates host cells besides human beings including the fowls, sheep, fish, amphibians and reptiles (Sharma *et al*., 2009). Human adenovirus infections occur throughout the year and are not seasonal although they have been reported to occur more commonly in the template regions from late winter to early summer (Lewis *et al*., 2009). Certain environmental conditions like overcrowded institutions e.g. pediatric ward and military training camps may also encourage seasonal outbreaks of the respiratory adenoviruses. The virus is resilient and can survive for a long time on environmental surfaces. Nosocomial transmission of adenovirus leading to

outbreaks is common in pediatric wards and overcrowded military training camps. Transmission occurs through respiratory droplets, fomites and ingestion with a relatively short incubation period of 2-14 days (Kim *et al.*, 2003; Lewis *et al.*, 2009; Marc *et al.*, 2012; O'Flanagan et al., 2011).

Globally, adenovirus infections accounts for 5–15% of upper respiratory infections and 5% of lower respiratory infections in pediatric respiratory infections (Marc *et al*., 2012). In Kenya the epidemiology of adenovirus has not been documented. However, adenoviruses have been associated with acute respiratory diseases (ARD) which affect both pediatric and adult population.

1.2.2.0 Methods of adenovirus detection

Study of viruses requires isolation of virus particles in cell cultures which provide an alternative to earlier methods of virus isolation which included use of embryonated eggs and laboratory animals (Leland *et al.,* 2007). Cell culture offers advantage over traditional methods in that contamination can be controlled by incorporating antibiotics in culture media, decreased use of experimental animals and an opportunity to work in a clean air working environment i.e. biosafety cabinets. In addition, viruses growing within susceptible cells will reach high titers and culture tubes/flasks are easy to manipulate. There are diverse cell types which can be propagated *in vitro* in flasks and test tubes and this provides living hosts that many human viruses can infect (Leland *et al.,* 2007).

Due to technological development there are now advanced methods of detecting viral infections such as the use of monoclonal antibodies and use of molecular diagnostics that detects viral DNA or RNA through amplification by PCR. These methods are sensitive

and highly specific and do not require lengthy incubation period and do not involve a lot of technical expertise unlike isolation in cell culture (Leland *et al.,* 2007).

1.2.2.1 Virus isolation in cultured cells

Successful virus isolation depends on appropriate collection, transport and processing of clinical samples which ensures preservation of virus titer and viral infectivity prior to inoculation in cell culture. Sample collection method is crucial and in general clinical samples collected from sites such as the skin, and the genital urinary tract and which are usually contaminated with a wide range of micro organism are collected with a polyester swab and placed in Viral Transport Media (VTM) (Leland *et al.,* 2007). VTM contains antibiotics, a suitable buffered salt solution, a proteinaceous component such as albumin, gelatin or serum, and a pH indicator e.g. phenol red. Respiratory tract samples include sputum and nasopharyngeal or oral pharyngeal swabs also collected in VTM.

Processing of nasopharyngeal swabs specimens usually involves vortex of transport medium tube and discarding the swab (Leland *et al.,* 2007). The liquid medium is centrifuged and supernatant which contains the virus is used to inoculate the cell culture. Bacteria, fungi cells, blood, mucus fibers etc settles at the bottom as a pellet due to centrifugal force (Leland *et al.,* 2007).

Inoculation involves simply adding about 0.1 or 0.2 ml of the processed inoculum to the culture tube. The inoculated tube is incubated for 30-90minutes at 37^{o}C after which excess inoculum is discarded prior to addition of fresh culture medium. Inoculated cell culture tubes are incubated for 10 to 14 days depending on the virus type and source of specimen. Over the incubation period the cell culture monolayer is screened by

microscopic examination on a daily basis for detection of virus proliferation. Degenerative changes in monolayer cells collectively termed cytopathic effect (CPE) provide evidence of viral presence (Leland *et al.* 2007). Typical cytopathic effects of common viruses are shown in Fig. 2 below.

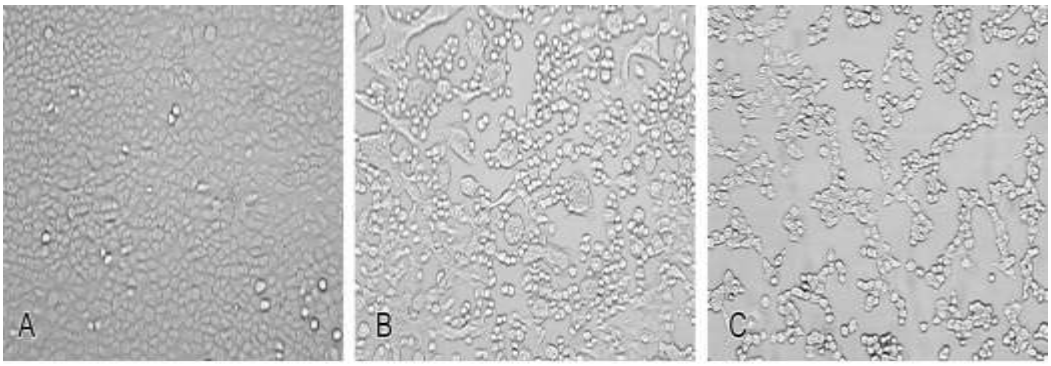


Figure 2. Uninfected cell cultures and cell cultures showing CPE of commonly isolated viruses. (A) Uninfected A549 cells (B) HSV-2 in A549 (C) Adenovirus in A549 cells. Adopted from Diane *et al*., 2007

1.2.2.2 Immunofluorescence assays

Virus positive cell cultures by CPE are confirmed by reaction of antibody of known specificity with viral antigens expressed in the infected cells (Leland *et al.,* 2007). This confirmatory testing is accomplished through immunofluorescence technique which use fluorescein isothiocyanate (FITC) labeled monoclonal antibodies of known specificity. Fluorescence microscope is used to view binding of monoclonal antibodies to viral proteins which is captured as a fluorescence signal. This process is fast and usually takes 1-2 hours (Leland *et al.,* 2007). Nowadays, serological determination of viruses has been replaced by molecular methods which are more efficient and highly specific.

1.2.2.3 Molecular methods of adenovirus detection

Serological determination of adenovirus type usually takes long (1-2 weeks) which limits its clinical value especially in the management of immune compromised patients (Leland *et al.,* 2007). Studies have shown that specific adenovirus serotypes are associated with manifestation and severity of disease presentation and therefore more expeditious methods of screening patient's samples are preferable (Wu, *et al.,*1990). Such methods involve PCR followed by further description of the amplicons through bioinformatics. This allows identification of virus at species and type level in a relatively short period and with minimum labor expense. Therefore, genomics and bioinformatics techniques are very important research tools in advancement of research involving adenoviruses.

1.2.4.0 Adenovirus lifecycle and genetics

Adenovirus entry into the cell involves interaction of the knob domain of fiber protein with the host cell receptors which includes the coxsackie and adenovirus receptors (CAR) and CD46 (Howitt *et al.,* 2003). The receptors mediate internalization of the virus through the process of endocytosis via clathrin coated pits. After successful entry into host cell, virus capsid dissociates and the viral genome is translocated to the nucleus through nuclear pores. Subsequently, viral genes are expressed and new virus particles are generated (Lauer *et al.,* 2004).

Adenovirus lifecycle consist of the early and late phases separated by DNA replication event termed intermediate phase. Expression of adenovirus genome is therefore organized into early, intermediate and late transcriptional genes.

The early genes consist of a set of first transcription units which are the E1A, E1B, E2, E3 and E4 genes. The E1A gene is the first transcription unit to be expressed following adenovirus infection. The expressed proteins functions in regulating host cells, modulating both viral and host cellular gene expression (Flint *et al.,* 1997). E1B codes for proteins which inhibits cellular defense mechanism and regulates viral late gene expression (Yew *et al.,* 1994). E2 transcription unit encodes proteins required for viral DNA replication. These are the 'terminal protein' precursor, DNA polymerase and DNA binding protein which functions along with host cellular proteins for efficient viral DNA replication (de Jong *et al.,* 2003). E3 transcription region encodes proteins that antagonize host immune response (Wold *et al.,* 1991). These proteins are not required for efficient *in vitro* growth of the virus. The E3 region is only found in members of genus *Mastadenovirus* and is therefore of particular interest to scientists as a target site for insertion of foreign gene constructs in gene therapy (Russell, 2009). Protein products of E4 transcription unit performs a wide range of functions including viral RNA export and stabilization and regulation of cellular transcription factors (Leppard, 1997)

The intermediate phase separates the early and late phases in adenovirus lifecycle and involves active DNA replication. It occurs once the early genes have synthesized adequate virus proteins, replication machinery and replication substrates. The intermediate gene region encodes two proteins Protein IX and Protein Iva2 which play a critical role in viral DNA packaging and as transcription activators of the late genes (Zhang *et al.*, 2003).

The late genes transcription unit is expressed as a single primary transcript from a single promoter site known as major-late promoter (MLP) and various distinct mRNA products

are obtained though truncation of the primary transcript at multiple polyA signals. The transcripts are grouped into five categories i.e. L1-L5. L1 transcript encodes a 52 kDA protein and protein IIIa which acts as a scaffold for capsid assembly during the assembly of viral particles. L4 transcript encodes four proteins, 100kDa protein, 22kDa protein, 33kDa protein and protein VIII. The functions of these proteins are not known (Young, 2003). The remaining three late genes transcripts encode the diagnostically useful antigens in adenovirus; the penton, hexon and fiber structural proteins of adenovirus capsid (Madisch *et al.*, 2005). Fig. 3 below shows the location of these major capsid proteins in adenovirus icosahedral structure (http://biommed.lsu.edu/onlinefiles/BIOL4190/LECT-5-ADENOVIRUS/LECT-5-SLIDES-B&W-ADENOVIRUS.pdf).

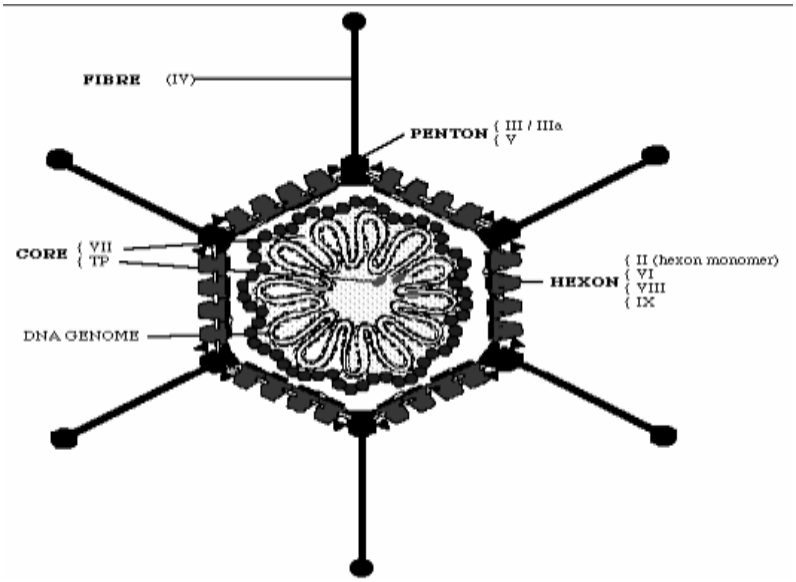


Figure 3. Morphological structure of adenoviruses illustrating adenovirus major capsid proteins (hexon penton and fiber). Adopted from Louisiana State University School of Veterinary Medicine in Baton Rouge, LA. Slide #11

Adenovirus capsid consists of 252 capsomeres (12 pentons, 240 hexons). Pentons makes up the vertexes and possess projecting fibers. Penton protein is encoded by L2 gene

transcript and contains regions that bind to host integrins via a conserved RGD (Arginine-Glycine-Asparagine) sequence to trigger virus internalization (Wickham *et al.,* 1993). Fiber protein is encoded by L5 gene transcript and it mediates the attachment of virus to cellular receptors and has been shown to possess hemagglutinin activity (Madisch *et al*., 2005). Its N- terminal domain attaches non-covalently to the penton base protein while the C terminal "knob" domain binds host cells. A study by Howitt et al (Howitt *et al.,* 2003) showed important key residues required for binding of the knob domain to the Cox-sackie and adenovirus receptors' CAR. These residues are Asparagine 415, Proline 417 and Proline 418.

L3 gene transcript encodes hexon protein the major structural component of adenovirus capsid. It makes up about 63% of the total virion mass and for this reason, this protein has been extensively studied by use of X-ray crystallography, molecular modeling and bioinformatics (Rux *et al.,* 2003). The genome region coding the L3 gene transcript is approximately 2.9 kb and several studies have shown that it consist of several hyper-variable regions (HVR) (Adhikary *et al*., 2004; Crawford *et al.,*1996; Ebner *et al*., 2005; Madisch *et al*., 2005). In their study on comparison of hexon sequence in human adenoviruses subgroup C, Kinloch et.al for the first time demonstrated that hexon gene could be divided into three different zones (1-393, 394-1410, 1411-2910) as shown in Fig. 4 (Kinloch *et al.,* 1984).

Zones 1 and 3 in the hexon gene of HAdV-C2 and HAdV-C5 had 95% and 89.5% nucleotide identity respectively with nucleotide identity of the gene diverging in zone 2. It has since been determined that zone 2 is recognized as the ε determinant which contains the type specific epitopes of the hexon gene (Crawford *et al.,* 1996). It consists

of loop 1 and loop 2 regions of hexon protein also illustrated in Fig. 4. Loop 1 consists of HVR-1 to HVR-6 while loop 2 has HVR-7.

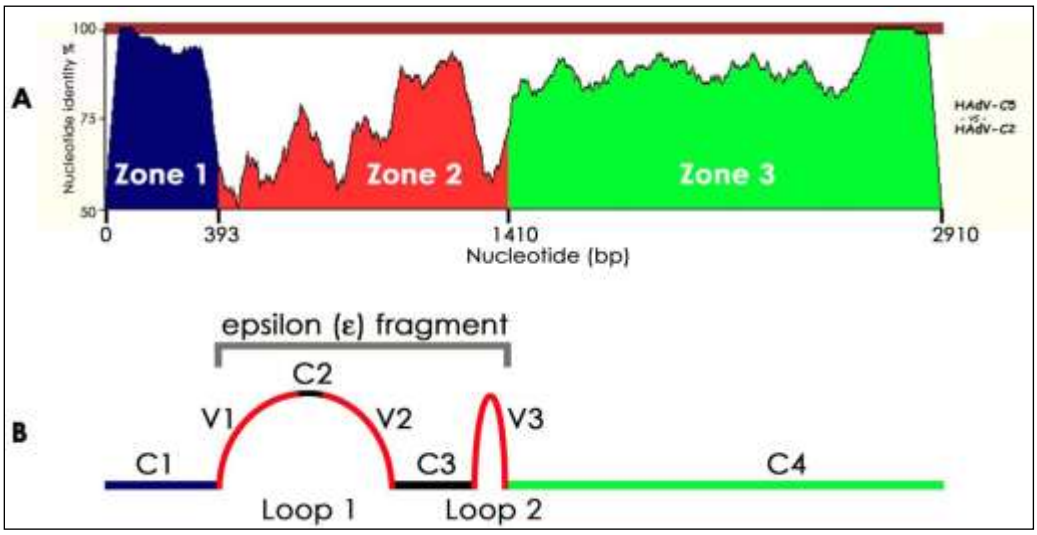


Figure 4. Analytical description of the hexon gene and hexon protein. (A) Nucleotide homology between the hexon coding sequences of HAdV-C2 and HAdV-C5. (B) Schematic illustration of the hexon protein. Variable regions V1, V2, and V3 are located in loops 1 and 2 and are outwardly oriented on the surface of the adenovirus particle. The epsilon (ε) fragment includes regions V1, C2, V2, C3, and V3. Adopted from Sarah Torres *et al.,* 2010

The hyper variable regions are determined from multiple sequence alignments of adenoviruses hexon gene. The hyper variable regions 1-6 are demonstrated in Fig. 5 from a study carried out by Madisch et al in 2005 which demonstrated the importance of the hyper variable regions in molecular taxonomy of adenoviruses (Madisch *et al.*, 2005). Figure 5 illustrates one of the Simplot diagrams obtained after carrying out similarity plots analysis of the ε determinant deduced amino acid sequence from HAdV-D serotypes that were observed to cluster in phylogenetic analysis.

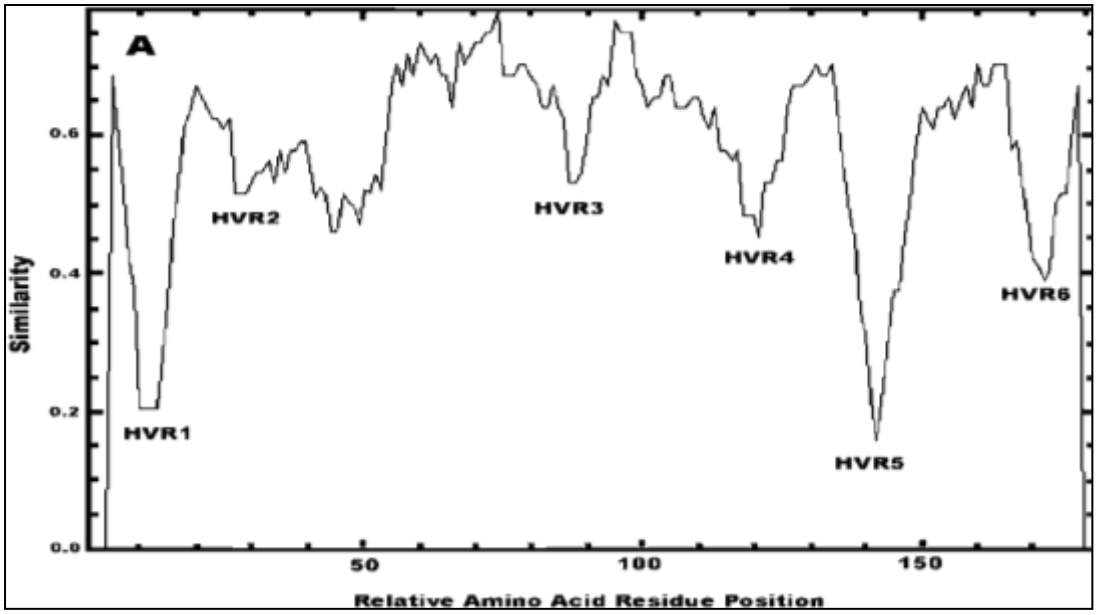


Figure 5 Similarity plot of ε fragment amino acid sequences from HAdV-D serotypes 23, 15, 29, 42, 20, 9 and 32. Adopted from Madisch et al 2005

The study (Madisch *et al.*, 2005) proposed a molecular classification method of classifying adenoviruses that relies on phylogenetic analysis of nucleotide and deduced amino acid sequences of the epsilon (ε) determinant which constitute Loop 1 and Loop 2 fragments of hexon. Fig. 6 illustrates some of the results the researchers obtained from phylogenetic analysis of loop 1 and loop 2 sequences obtained from members of HAdV species A-F. Their study demonstrated an overall similarity in phylogenetic trees reconstructed from nucleotide and deduced amino acid sequences of Loop1 and Loop2 regions. Loop2 region was proposed as the simplest yet specific molecular approach in molecular taxonomy of adenovirus and hence should be used as a standard procedure in molecular classification of adenoviruses. Similar findings have been reported by Helen Sarantis et al. (Sarantis *et al.,* 2004) who designed a PCR assay with a primer pair targeting conserved regions within Loop 2 HVR7 region that was able to detect all

serotypes of human adenoviruses known at that time with high sensitivity. Consequently, ε determinant region of hexon gene makes hexon gene a good molecular chronometer in evolutionary studies of adenoviruses.

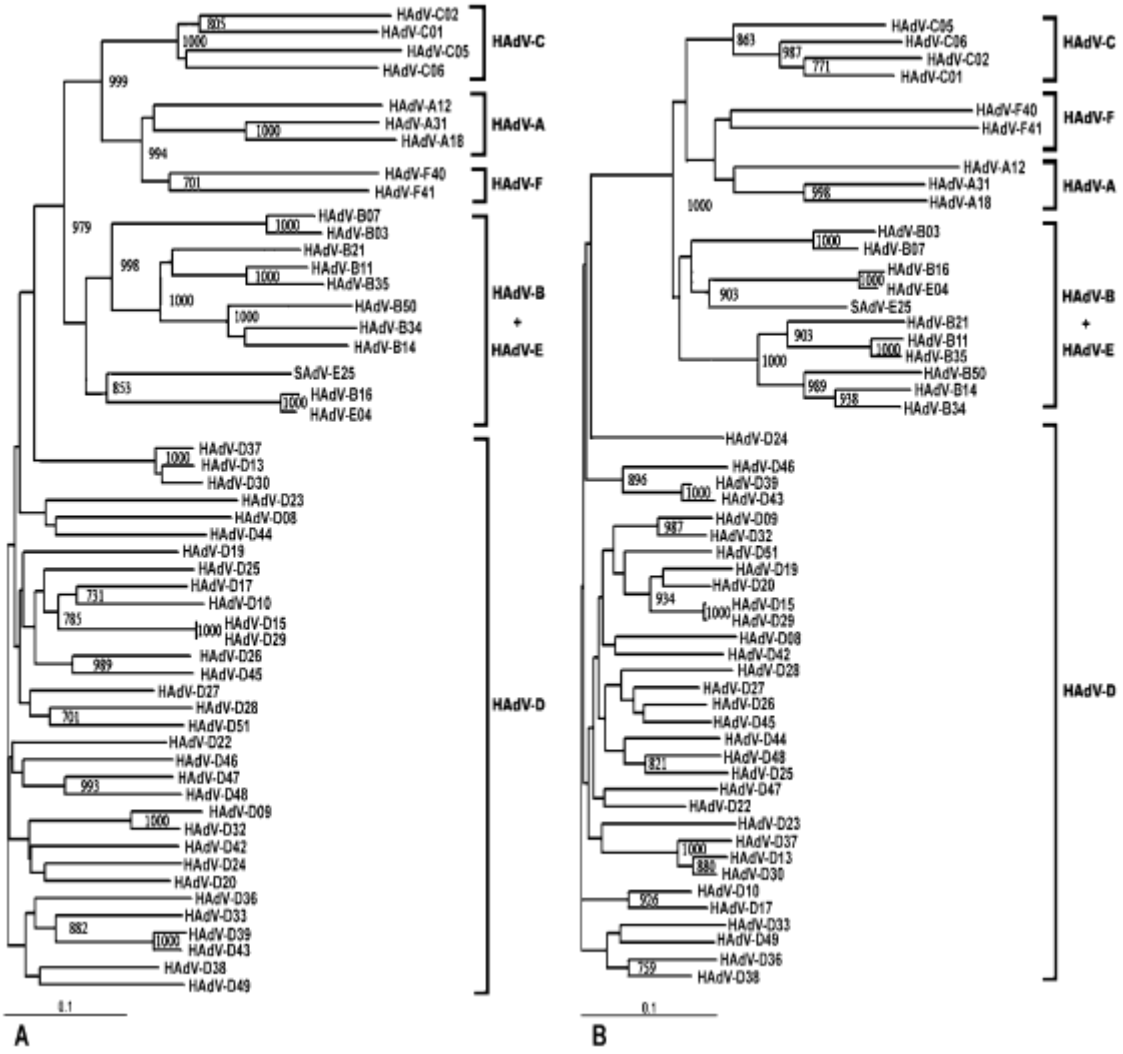


Figure 6. Illustration of phylogenetic analysis of adenoviruses using nucleotide sequences of (A) ε determinant Loop 1 and (B) ε determinant Loop 2. The reference sequence for phylogenetic analysis was HAdV C1 ε determinant sequences. The aligned sequences from various HAdV species were obtained from GenBank. The two phylogenetic trees display overall similarity in clustering of the HAdV species. Adopted from Madisch et al. 2005

1.2.4.1 Molecular taxonomy of adenoviruses

Madisch et al. 2005 proposed a method where loop 2 nucleotides and amino acids divergence are used as a basis for typing adenoviruses or identifying novel human adenovirus strains (Madisch *et al.*, 2005). This was informed by the finding that loop 2 nucleotide sequence pair wise identity scores among adenoviruses results into two groups of heterogeneous serotypes of the same species and heterogeneous serotypes from different species. Their study demonstrated the lowest nucleotide divergence in loop 2 region among serotypes of the same species as 2.5%. Therefore, the researchers suggested that if an isolate has a nucleotide divergence of less than 2.5% compared to the next homologous prototype already identified and deposited in GenBank then such an isolate can be said to have been successfully typed. This interpretation may also be presented from a sequence identity perspective. Thus two HAdVs with ≥97.5% nucleotide sequence identities are deemed to be of the same serotype.

The study (Madisch *et al.*, 2005) also proposed that if analysis of nucleotide divergence is unsuccessful, then deduced amino acids sequences should be used to calculate the divergence rate. Consequently, the isolated virus deduced amino acids sequence is aligned with sequences already deposited in the data bank and if loop 2 amino acids divergence compared to the next homologous sequence in the data bank is less than 1.2%, typing is accomplished. Isolation of a new adenovirus serotype is suspected if amino acids divergence of above 1.2% is observed.

Currently there are 60 types of human adenoviruses (Robinson *et al.*, 2013). Serotypes 1-51 were classified based on serological assays and further classified into six species (A to F) based on ability to agglutinate various types of erythrocytes. The term ‘type’ is

therefore used as an inclusive term referring to the initial 51 serotypes determined using traditional serology as well as the additional nine 'genotypes' characterized using genomics and bioinformatics (Torres *et al.,* 2010).

Powerful bioinformatics tools which include Multiple Sequence Comparison by Log-Expectation (MUSCLE), Clustal, and Basic Local Alignment Search Tool (BLAST) for sequence alignment; and MrBayes, and phylogeny inference program (PHYLIP), for inferring phylogenetic relationship of sequences have been used on whole genomes (Altschul *et al.,* 1990; Altschul *et al*., 1997; Edgar, 2004; Felsenstein, 1993; Huelsenbeck *et al.,* 2001; Sprengel *et al.,* 1994). Spengel et.al 1994 used Clustal V in a ground breaking comparative analyses that enabled researchers to view overall similarity of two genomes on the nucleotide level in one figure (Sprengel et al., 1994). This comparative analysis led to discovery that HAdV-A12 and HAdV-C5 genomes have the highest nucleotide identity in a 5000-8000bp region which codes for DNA polymerase. With advancement in sequencing technology and bioinformatics software development, we now have a rich collection of bioinformatics resources that have been very useful in research involving adenoviruses and other human pathogens.

1.2.4.2 Recombination in adenoviruses

Recombination among adenoviruses has also been well documented (Shenk, 2001.). It is thought that recombination resulting from genetic exchange of parts of hexon penton or fiber genes among members of same or different species is a key drive to emergence of new strains of adenoviruses (Walsh *et al.*, 2010). HAdV55 is thought to have emerged from recombination involving HAdV B11 and HAdV B14 hexon genes (Walsh *et al.*, 2010). HAdV E4 tends to cluster with HAdV B16 in phylogenetic analyses (Fig. 6 section 1.2.4.0) and it is suspected that HAdV E4 the sole serotype of species E could have resulted from recombination event involving strains of HAdV B serotype 16 (Shenk, 2001.).

1.2.5.0 Genetic Mutations

Mutations or nucleotide substitutions in protein coding genes can be categorized as synonymous or non synonymous (Nei *et al.,* 1986). Synonymous mutations do not result in change of amino acids sequence in the protein and are usually invisible to natural selection. Non synonymous mutations on the other hand results in change in the protein's amino acids sequence and are an indicator of strong selective pressure in a particular gene (Nei *et al.,* 1986). Therefore estimation of synonymous and non synonymous substitution rates is important in understanding the dynamics of molecular evolution of gene sequences.

Due to the degeneracy of the genetic code there are fewer opportunities for non synonymous changes than for synonymous changes implying that the rate of synonymous mutations is greater than the rate of non synonymous substitutions. However under

certain conditions, non synonymous substitution may be accelerated by positive evolutionary selection. It is therefore important to examine the number of synonymous differences per synonymous sites (dS or Ks) and the number of non synonymous differences per non synonymous sites (dN or Ka). The dN/dS ratio is expected to be 1.0 for aligned genes that show no selection pressure relative to each other. dN/dS ratio of less than 1 implies purifying selection, and a ratio greater than 1 implies positive selection (Nei *et al.*,1986).

While estimating mutations in a pair of homologous sequences presenting only one nucleotide difference in the aligned codons, the number of synonymous and non synonymous substitutions may be obtained by simply counting the number of silent versus non silent amino acid changes (Nei *et al.,* 1986). However in a pair of sequences presenting two or three nucleotide differences in the aligned codons, distinction between synonymous and non synonymous substitutions is not easy to calculate and statistical estimation methods are needed (Nei *et al.,* 1986). There are various methods of estimating synonymous and non synonymous substitutions and their effect on selection which includes the following:

1.2.5.1 Nei and Gojobori method

This method uses the number of synonymous and non synonymous substitutions and the number of potentially synonymous and non synonymous sites. It's based on the Jukes-Cantor model of nucleotide substitutions (Nei *et al.,* 1986).

1.2.5.2 Kumar method

Kumar method analyzes mutations by comparing the numbers of synonymous and non-synonymous substitutions per site where, the ratio dN/dS=φ measures the difference between the two rates (Kumar *et al.,* 2004).

1.2.5.3 Tajima's method

In Tajima's method, the pattern of nucleotide difference is examined as measured by the difference between π (observed average pair wise nucleotide diversity) and θ (expected nucleotide diversity under neutrality derived from the number of segregating sites (Tajima, 1993)

1.2.6.0 Evolutionary distances of sequences

The evolutionary distances between pair of sequences are usually measured by the number of nucleotide or amino acid substitutions between them. Evolutionary distances are fundamental in the study of molecular evolution of genes, phylogenetic reconstructions and in estimating the divergence times of homologous sequences (Kumar *et al.*, 2004). Various statistical methods are used to estimate evolutionary distances. They constitute mathematical algorithms designed for predicting nucleotide or amino acid differences/change expected in a sequence. They include:

1.2.6.1 The P-distance (nucleotides or amino acids)

This distance is the proportion *(p)* of nucleotide or amino acid sites at which two sequences being compared are different. It is obtained by dividing the total number of nucleotide or amino acids differences by the total number of nucleotides compared. It does not make correction for multiple substitutions at the same site, transitional and

transversional substitution rate biases or differences in evolutionary rates among the sites (Kumar et al., 2004).

Po= nd/n where *Po* is the proportion of nucleotide or amino acid while *nd/n* is the number of positions at which the sequences differs divided by their lengths.

1.2.6.2 Jukes and Cantor, "one-parameter model" denoted as "1-p":

This model assumes that the rate of nucleotide substitution is the same for all pairs of the four nucleotides A, T, C and G (Jukes *et al.,* 1969). The multiple hit correction equation for this model produces a maximum likelihood estimate of the number of nucleotide substitutions between two sequences. It assumes an equality of substitutions rates among sites, equal nucleotide frequencies, and it does not correct for higher rate of transitional substitutions as compared to transversional substitutions.

$d = -3/4 \log_e (1-4/3p)$ where *p* is the proportional of sites with different nucleotides

1.2.6.3 Kimura's 2-parameter model, denoted as "2-p" :

The model assumes that the rate of transitional nucleotide substitution is often higher than that of transversional substitution (Kumar *et al.,* 2004). Transitional substitution refers to nucleotide substitution involving either purines or pyrimidines while in transversional substitution a purine is substituted with pyrimidine or vise versa.

$$d = -(1/2) * Ln (1 - 2 * P - Q) - (1/4) * Log (1 - 2 * Q)$$

P is the proportion of transitional differences, *Q* is the proportion of transversional differences. *P* and *Q* are respectively calculated over synonymous and non synonymous differences.

1.3 PROBLEM STATEMENT

Adenoviruses play a significant role in pediatric infections accounting for up to 15% of overall respiratory illnesses (Marc *et al.,* 2012). The burden of diseases due to respiratory viruses including adenoviruses in Kenya has not been well studied. Furthermore, whereas anecdotal evidence indicates that respiratory HAdV are a prevalent cause of respiratory illnesses in Kenya, the viruses have not been molecularly characterized into their species and serotypes. Efforts are currently being made to enhance these studies at the National Influenza Center (NIC) in Nairobi Kenya. The NIC has a mandate to carry out surveillance of influenza and other viral respiratory pathogens. Earlier on, some human adenovirus strains were isolated at the NIC but they were never genetically characterized. This study sought to isolate and characterize using molecular methods the various strains of respiratory human adenoviruses isolated from pediatric patients specimens received from the New Nyanza Provincial General Hospital in the period June 2010 to June 2012. Some strains of adenoviruses are implicated in fatal pneumonia and further studies will better inform the clinicians on the types of respiratory adenoviruses circulating in this region to enable better patient care and disease management.

1.4 JUSTIFICATION AND SIGNIFICANCE OF THE RESEARCH

Human adenoviruses are recognized causative agents of acute respiratory diseases which occasionally can result in severe or lethal infections (Kim *et al*., 2003; Lewis *et al*., 2009; O'Flanagan *et al*., 2011). Adenoviruses are also vectors for several vaccines including Adv5 in the Prime Boost HIV vaccine in Kenya (Kibuuka *et al*., 2010), and knowledge of adenovirus prevalence is important as it will impact on the efficacy of such vaccines. There is very little information about human adenovirus in Kenya and specifically, there is no documented data on respiratory human adenovirus species and serotypes circulating in the country. We do not know the genetic composition of the isolated adenoviruses and how they compare with other adenovirus isolates implicated in similar infections in other parts of the world. Knowledge of the various species and serotypes implicated in respiratory diseases is necessary for better patient care and disease management. Furthermore, serotype identification is crucial for epidemiological surveillance, detection of new strains and better understanding of human adenoviruses. To address some of these challenges genetic analysis of isolated respiratory adenoviruses from pediatric out-patients presenting with respiratory infections in New Nyanza Provincial General Hospital in the period June 2010 to June 2012 were carried out.

1.5 RESEARCH QUESTIONS

Respiratory adenoviruses are a significant group of viruses implicated in acute respiratory diseases in Kenya. However there is no documented data of confirmed respiratory adenovirus species implicated in these infections. In addition molecular sub typing of the respiratory species has not been carried out. This study provides a genetic overview of various respiratory adenovirus serotypes implicated in respiratory diseases in western Kenya in the period June 2010 to June 2012. The focus of this study is therefore based on the general question:

1. Were adenoviruses associated with pediatric respiratory infections in New Nyanza Provincial General Hospital in the period June 2010-June 2012?

From the general question, the study sought to answer the following specific questions.

1. Which species of respiratory adenovirus was the predominant cause of illness among pediatric patients visiting the New Nyanza Provincial General Hospital and hence circulating in western region of Kenya?

2. Which respiratory adenovirus serotypes were associated with pediatric infections in New Nyanza Provincial General Hospital in the study period?

1.6 OBJECTIVES

1.6.1 Main objective

To genetically characterize respiratory adenoviruses isolated from pediatric patients attending New Nyanza Provincial General Hospital in the period June 2010 to June 2012, using serological and molecular approaches

1.6.2 Specific objectives

1. To confirm the presence of adenoviruses from archived virus stocks isolated from patient nasopharyngeal specimens using cell culture and immunofluorescence assays.
2. To determine the nucleotide sequences of the hexon gene segments of the adenovirus isolates using PCR amplification and the Sanger method.
3. To carry out analyses of the hexon gene nucleotide sequences and classification of the adenoviruses obtained using the bioinformatics techniques

CHAPTER TWO

2.0 MATERIALS AND METHODS

2.1 Study Sites

The study site for this work was New Nyanza Provincial General Hospital, one of the sites included in the United States Army Medical Research Unit in Kenya (USAMRU-K) National Influenza surveillance network. The hospital serves patients mainly from the Western region of Kenya.

2.2 Study Design

Retrospective study design was used and it involved use of archived nasopharyngeal patient's samples collected over a period of two years in an ongoing program under the ethical review committee approved "Influenza Surveillance in Kenya" protocol, KEMRI SSC#981 and WRAIR#1267.

2.3 Study population

The study population consisted of patients aged ≥2 months to 17 years of age from New Nyanza Provincial General Hospital. In the period June 2010 –June 2012 a total of 1879 nasopharyngeal patient's samples from New Nyanza Provincial General Hospital were received at National Influenza Centre (NIC). These samples had previously been screened for various respiratory viruses including adenoviruses in routine surveillance work at the NIC. Since the goal of the study was to analyze adenovirus cases in pediatric patients, only isolates obtained from patients aged ≥2 months and not exceeding 17 years of age were considered. For that reason, the study population was constituted 17 virus

stocks from reported adenovirus cases that matched the case definition of the study and that were collected within the stated study period.

2.4 Inclusion criteria and exclusion criteria

Only isolates that were obtained from paediatric patients aged 2 months to 17 years and were positive for HAdV were included in this study. Isolates from paediatric patients below 2 months or above 17 years were excluded.

2.5 Ethical Considerations

The study was a sub-protocol of the "Influenza Surveillance in Kenya", WRAIR#1267/SSC # 981 protocol. The study was granted ethical and scientific clearance by the Kenya Medical Research Institute (KEMRI) and the Walter Reed Army Institute of Research (WRAIR) committees under protocol number SSC#2499/WRAIR#2014. In the parent protocol, informed consents had been obtained from parents or guardians of children involved in the study. Those subjects aged between 12-17 years had also assented that their samples could be used for future studies for identification of other respiratory viruses. Archived virus stocks used in this study were given unique codes that do not give patient information. This information was stored and accessed by the principal investigator and his supervisors only.

2.6.0 Laboratory procedures

2.6.1 Sample collection

Nasopharyngeal throat swabs samples had been collected from pediatric patients attending the New Nyanza Provincial General Hospital as part of the USAMRU-K Influenza like Illnesses surveillance protocol (KEMRI SSC#981; WRAIR#1267).

2.6.2.0 Virus Isolation

2.6.2.1 Growth and maintenance media preparation

50 ml of growth media was prepared for culturing Hep2 cells which are a HeLa cell contaminant (Scherer *et al.*, 1953), (Masters, 2002). The media component were; 1% L-glutamate, 1%Penicillin/streptomycin, 1% Fungizone, 10% heat inactivated Fetal Bovine Serum, 0.1%Gentamycin, and DMEM topped up to 50ml (all reagents from Gibco® Life Technologies™ Grand Island NY 14072 USA. This media was used to grow Hep2 cells monolayer confluent for virus stocks inoculation.

Maintenance media consisted of the following components. 1%L-glutamate, 1%Penicillin/streptomycin, 1% Fungizone, 2%Fetal Bovine Serum, 0.1%Gentamycin, DMEM topped up to 50ml. The media was used to maintain a steady growth of virus inoculated Hep2 cells during incubation period.

2.6.2.2 Preparation of Hep2 cells

Hep2 cell culture monolayer was grown using the Walter Reed Army Institute of Research -USAMRU-K Standard Operating Procedure for Hep2 cell line proliferation as described below. 18 culture tubes were prepared using Hep2 cells (ATCC, VA, USA. Ref

number: HB-8065™) passage 6 in T75 culture flask (Corning Incorporated, Corning, NY 14831 USA). The culture media in T75 flask was poured off and the cell monolayer briefly washed twice by adding 2ml of Trypsin/EDTA solution (Gibco® Life technologies Corporation, Grand Island, NY, 14072, USA) and a final 3ml of Trypsin/EDTA solution incubated to allow the cell monolayer to detach. Cell clumps were re-suspended by gently pipetting against the flask bottom surface in order to ensure detachment of all the cells. In order to prevent cell lysis by trypsin/EDTA, 8ml of the growth media containing fetal bovine serum; an inhibitor of trypsin was added. 2ml aliquot of the resulting cell mixture was mixed with a further 20ml of growth media and dispensed in slanted culture tubes (Nunclon™, Surface, DK-4000, Roskilde, Denmark) 1ml per tube. The tubes were incubated loosely capped at 33-37^{o}C 5% CO_2 (Sanyo Incubator, IR Safe 1300, Michael Drive, Suite A, Wood Dale, IL, 60191, USA) for 3 days prior to inoculation. The tubes were observed daily under an inverted microscope (Olympus, Shinjuku Monolith, 2-3-1 Nishi-Shinjuku, Shinjuku-ku, Tokyo 163-0914, Japan) for 70-90% monolayer confluent required for virus inoculation.

2.6.2.3 Virus inoculation

Virus inoculation involved 17 archived adenovirus stocks that had been stored at -80^{o}C. The virus stocks are given in table 2 showing unique identification that does not give patient information. The virus stocks were allowed to thaw at room temperature prior to inoculation. Standard procedure for Hep

Growth media in the 70-90% confluent monolayer of Hep2 cells in culture tubes was poured off and a volume of 100µl of each isolate inoculated into each tube. The tubes were incubated at 33- 37^{o}C (Thermo Scientific, 81, Wyman Street, Waltham, MA 02454

USA) for 60 minutes to allow virus attachment to receptors. 1 ml of maintenance media was then added to each culture tube and incubated at 33-37^{o}C 5% CO_2. After a period of 48 hours, the media was discarded and replaced with fresh maintenance media. The inoculated Hep2 cells were incubated for additional 10 days with daily observation for cytopathic effect (CPE) using an inverted microscope (Olympus, Shinjuku Monolith, 2-3-1 Nishi-Shinjuku, Shinjuku-ku, Tokyo 163-0914, Japan). In culture tubes showing clear CPE, the supernatants were collected through centrifugation in a bench centrifuge (Centrifuge 5810 R Eppendorf AG Barkhausenweg 1 22339 Hamburg Germany) at 3000rpm for 10 minutes and stored for further characterization. The pellet was used for immunofluorescence assay to confirm adenovirus presence.

Table 2 Identities of archived virus stocks used in the study. Virus stocks are labeled with unique codes corresponding to patient's nasopharyngeal specimens. NNY (New Nyanza) refers to the study site where the samples were obtained followed by the year and month the samples was collected and sample serial number.

Sample Unique ID
NNY-10-06-016
NNY-10-06-018
NNY-10-06-032
NNY-10-06-073
NNY-10-07-009
NNY-10-07-033
NNY-11-09-001
NNY-10-11-022
NNY-10-11-034
NNY-10-11-040
NNY-11-03-077
NNY-11-07-001
NNY-11-10-044
NNY-11-11-013
NNY-11-12-004
NNY-12-03-027
NNY-12-03-029

2.6.3 Immunofluorescence assay to confirm Adenovirus

Light Diagnostics Adenovirus direct fluorescent antibody (DFA) kit (Millipore, Temecula, CA, 92590, USA) was used for immunofluorescence assays detection of adenovirus in the supernatants. 3ml of sterile Tween Phosphate Buffered Saline (Millipore, Temecula, CA, 92590, USA) was added to the pellet and centrifuged in a bench centrifuge (Centrifuge 5810 R Eppendorf AG Barkhausenweg 1 22339 Hamburg Germany) at 3000rpm for 10 minutes. The supernatant was discarded and 25µl of the pellet placed in microscope glass slides (Cell-line/Thermoscientific). The slides were left overnight to allow sufficient time for virus immobilization. Virus immobilized microscope glass slides were fixed in acetone (SIGMA-Aldrich) for 10 minutes and air dried prior to application of adenovirus specific monoclonal antibodies. Adenovirus specific monoclonal antibodies tagged with a fluorescent dye fluorescein isothiocyanate (FITC) were applied to virus immobilized on the microscope glass slides and incubated for 30 minutes. The glass slides were washed with freshly prepared Tween 20 Phosphate Buffered Saline (Thermo Scientific, 81, Wyman Street, Waltham, MA 02454 USA) Tween solution and air dried. Mounting fluid was applied and the glass slide covered with cover slides. The slides were then observed under ultra violet (UV) light on fluorescence microscope (Olympus BX51 Olympus, Shinjuku Monolith, 2-3-1 Nishi-Shinjuku, Shinjuku-ku, Tokyo 163-0914, Japan); an apple green fluorescence on immobilized cells indicated adenovirus presence while red colour indicated a negative outcome.

2.6.4 DNA Extraction

Adenovirus DNA was extracted from supernatants of the 17 viral stocks using the QIAamp DNA mini kit (Qiagen, Valencia CA, USA). In this extraction protocol 20µl of Proteinase K, 200µl of the cell culture supernatants and 200µl of viral lysis buffer were placed in 1.5ml centrifuge tubes and mixed by vortex for 15 seconds. The mixtures were incubated at 56^{o}C for 10 minutes in a Thermo Mixer comfort Incubator (Eppendorf AG Barkhausenweg 1 22339 Hamburg Germany). The mixtures were then centrifuged for 30 seconds at 8000rpm 4^{o} C in a bench centrifuge (Centrifuge 5415 Eppendorf AG Barkhausenweg 1 22339 Hamburg Germany). 500µl absolute ethanol was added to each tube, mixed by vortex for 15 seconds and the mixtures centrifuged for 30 seconds at 8000rpm 4^{o} C. The mixtures were then transferred to the spin columns and spun for 60 seconds at 8000rpm 4^{o} C and flow-through discarded. 500µl of first wash buffer AW1 was added to each column and centrifuged for 60 seconds at 8000rpm 4^{o} C and the flow-through discarded. 500µl of second wash buffer AW2 was added to each column and centrifuged at 13200rpm 4^{o} C for 3 minutes. The flow-through was discarded and the 2ml collection tubes replaced for a dry spin at 13200rpm 4^{o} C for 60 seconds. Viral DNA was eluted using 100µl of elution buffer EA by spinning at 8000rpm 4^{o} C for 60 seconds. The extracted DNA was stored at -20^{o}C for further processing.

2.6.5 Polymerase Chain Reaction (PCR) of the hexon gene loop2 fragment.

Extracted DNA from the 17 culture supernatants were amplified by conventional PCR using a primer set that amplify the hexon hyper-variable region 7 located in loop2 region. The forward and reverse primer sequences were adopted from published study (Sarantis *et al.*, 2004). Primer sequences are given in Table 3 below.

Table 3 Primers Sequence for hexon amplification

HVR7 up (forward)	5′CTG ATG TAC TAC AAC AGC ACT GGC AAC ATG GG 3′
HVR7 down (reverse)	5′GCG TTG CGG TGG TGG TTA AAT GGG TTT AC 3′

The reaction mixture contained 1X PCR Buffer, 3µM MgSO4, 200µM concentrations of each deoxynucleotide triphosphate, 0.5µM concentrations of each primer, 2.5 U of High Fidelity *Taq* DNA polymerase (all reagents from Invitrogen, Carlsbad CA, 92008 USA), and 5 µl of DNA template in a total volume of 50µl. PCR was carried out in a Gene Amp PCR Systems 9700 Thermo cycler (Life Technologies, 3175 Staley Road, Grand Island, NY, 14072, USA) with the following reaction parameters; an initial denaturation at 94°C for 2 minutes, followed by 30 cycles of denaturation at 94°C for 1 minute, annealing at 54°C for 1 minute, and elongation step at 72°C for 3 minutes, with a final extension at 72°C for 7 minutes. The PCR parameters were adopted and optimized from (Daniel *et al.* 2004). Virus amplicons were separated using 1% agarose gel electrophoresis containing 1µg/ml of ethidium bromide for 40 minutes at 70 V. The amplicon bands were visualized under ultra violet light in a Transilluminator Alpha Imager HP (Weltevreden Park, Johannesburg, 1715 S.A).

2.6.6.0 Nucleotide Sequencing

2.6.6.1 ExoSAP-IT PCR products purification and quantification

The amplicons from PCR were purified using ExoSAP-IT protocol (USB® ExoSAP-IT® PCR Product Cleanup Product numbers 78200/01/02/05/50). 2µl of ExoSAP were placed in 1.5 ml PCR tubes and mixed with 10µl of PCR amplified viral DNA. The components were mixed briefly by voltex followed by spinning for 30 seconds. The mixture was placed in a preheated Gene Amp PCR Systems 9700 Thermo cycler (Life Technologies, 3175 Staley Road, Grand Island, NY, 14072, USA) with the following parameter; incubation at 37^{o}C for 15 minutes followed by inactivation step at 80^{o}C for 15 minutes. Purified amplicons were quantified using Nanodrop 2000c spectrophotometer (Thermo Scientific, 81, Wyman Street, Waltham, MA 02454 USA), in a spectrophotometric procedure where absorbance of nucleic acids is measured at 260nm wavelength. 1µl of DNA was used in this measurement and concentration recorded in ng/µl. The ExoSAP purified viral DNA amplicons were used in Sanger Sequencing BigDye Reaction as described below.

2.6.6.2 Sanger Sequencing BigDye Reaction

The BigDye® Direct Cycle Sequencing Kit protocol (Applied Biosystems, Austin, TX ,18744 USA) was adopted for sequencing of ExoSAP purified viral DNA amplicons. The BigDye master mix was prepared in two separate reaction volumes for the HVR7-Forward primer and HVR7-Reverse primer. Each reaction mixture contained 2µl BigDyeTerminator V1.1, V3.1 5X Sequencing Buffer and 1µl BigDye Terminator V3.1 cycle Sequencing RR-100 (Applied Biosystems, Austin, TX ,18744 USA) 1µl of 4µM

primer, 4µl of Nuclease free water (Promega Madison WI USA) and 2µl of ExoSAP purified DNA amplicons. The 10µl reaction volume BigDye PCR was carried out in a Thermo cycler-Gene Amp PCR Systems 9700 (Life Technologies, 3175 Staley Road, Grand Island, NY, 14072, USA) with the following reaction parameters; an initial denaturation at 95°C for 5 minutes, followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds, and elongation step at 68°C for 2 minutes 30 seconds, with a final extension at 68°C for 3 minutes. The BigDye PCR products were purified by Sephadex purification method as described below.

2.6.6.3 Sephadex purification and resolution of BigDye PCR extension products

The BigDye PCR products were purified by Sephadex purification method adopted from published protocol (Sequencing using capillary electrophoresis on the ABI Prism 3700.pdf, n.d.) as described below. Sephadex G-50 powder (SIGMA-Aldrich) was measured using a plate loader and transferred to two multi-screen filter plates (Multiscreen™, Millipore, Ireland). 300µl of distilled water was added to each well on the multi-screen filter plates and let to set for at least 3 hours. Two 96 well optical PCR plates (MicroAmp™, Foster City, CA, USA) were aligned with the multi-screen plates and centrifuged at 910 X g for 5 minutes (Centrifuge 5810 R Eppendorf) to remove excess water. Another set of 96 well optical PCR plates separately labeled for the HVR7 forward and HVR7 reverse primers BigDye reaction products were aligned with the multi-screen filter plates. BigDye PCR products from each reaction were transferred to the aligned multi-screen filter plate-96 well optical PCR plates and centrifuged at 910 X g for 8 minutes. 10µl of double distilled water was added to each well of the 96 well optical PCR plates with Sephadex purified BigDye PCR products. The plates were assembled by

placing septa over the wells and loaded in the auto sampler cassette of preheated ABI 3500xL Genetic Analyzer (Applied Biosystems, Carlsbad CA, USA) and allowed to run for resolution of BigDye PCR extension products. DNA sequence Contigs from ABI 3500xLGenetic Analyzer were assembled and edited by DNA Baser Sequence Assembler version 10 (BioSoft) available at http://www.dnabaser.com and saved in FASTA format.

2.6.7.0 Bioinformatics analysis

2.6.7.1 Nucleotide sequence identity search, translation and submission to GenBank

To determine the identity of the assembled nucleotide sequences, an online bioinformatics resource tool – Basic Local Alignment Search Tool (BLAST) was used (Altschul *et al.*, 1990; Altschul *et al.*, 1997) available at http://www.ncbi.nlm.nih.gov/Blastn.cgi. The nucleotide sequences were used to query GenBank non-redundant nucleotide sequences database for similar sequences. The nucleotide sequences were translated into protein code using the Expasy translate tool (Artimo *et al.*, 2012) available on Swiss Institute of Bioinformatics resource portal http://web.expasy.org/tools/translate/.

The nucleotide sequences from this study were deposited in GenBank database under the following accession numbers: KJ527469, KJ527470, KJ527471, KJ527472, KJ527473, KJ527474, KJ527475, KJ527476, KJ527477, KJ527478, KJ527479, KJ527480, KJ527481, KJ527482, KJ527483, and KJ527484.

2.6.7.2 Multiple sequence alignment and Phylogenetic analysis

The GenBank nucleotide sequences displaying the highest sequence identities to each of the sixteen HAdV sequences were fetched from GenBank prior to multiple sequence alignment. The sequences fetched had accession numbers: KC747631.1 for HAdV C serotype 1 isolated in Korea; AF515814.1 for HAdV B serotype 7 isolated in China; HQ535662.1 for HAdV C serotype 6 isolated in United Kingdom; AF053085.1 for HAdV B serotype 7 isolated in Japan; AB436561.1 for HAdV C serotype 2 isolated in Japan; FJ943599.1 for HAdV C serotype 1 isolated in Germany; KC747641.1 for HAdV B serotype 5 isolated in Korea and JQ407712.1 HAdV C isolated in France. Reference hexon sequences representing each serotype of respiratory human adenovirus species B,C and E were also included in this analysis. These reference hexon coding sequences of HAdV species B, C and E included GenBank accesion numbers AJ293903.1 for HAdV C serotype 2 representing species C isolated in Germany, AY008279.1 for HAdV B serotype 21 representing species B1 isolated in the USA, KC689915.1 for HAdV B serotype 7 representing species B2 and KC551973.1 for HAdV 55 both isolated from China. Others were AB330115.1 for HAdV B serotype 34, AB330116.1 for HAdV B serotype 35, AB330131.1 for HAdV B serotype 50, AB330097.1 for HAdV B serotype 16, AB330095.1 for HAdV B serotype 14, AB330092.1 for HAdV B serotype 11, AB330088.1 for HAdV B serotype 7, AB330087.1 for HAdV C serotype 6, AB330086.1 for HAdV C serotype 5, AB330085.1 for HAdV E serotype 4, AB330084.1 for HAdV B serotype 3, and AB330082.1 for HAdV C serotype 1 all isolated from Japan. K01264.1 representing bovine adenovirus 3 was used as the outgroup in this analysis.

Multiple sequence alignment was perfomed using Muscle version 3.8.3.1(Edgar, 2004) and edited and visualised by GeneDoc version 2.6.002 (Nicholus *et al*.,1997) Phylogenetic analysis was perfomed using bayesian inference of phylogeny executed in MrBayes version 3.2.2 (Huelsenbeck *et al.,* 2001). The General Time reversible (GTR) model (Tavaré, 1986) with gamma distributed rate variation across site and a proportion of invariable sites was used. For phylogenetic reconstruction, 1,000,000 Markov chain Monte Carlo (MCMC) generations were performed with a sampling frequency after every 1000 trees. The results of phylogenetic analyses were visualized with FigTree version 1.4.0 (Andrew Rambaut, 2007).

2.6.7.3 Location of hyper variable region 7

To locate the hyper variable region 7 within the hexon gene using the nucleotide sequences, a multiple sequence alignment of the sequences was built using the HAdV2 hexon gene (GenBank accesion number AJ293903.1). The alignment was perfomed using Muscle version 3.8.3.1(Edgar, 2004) and visualised by GeneDoc version 2.6.002 (Nicholus *et al.,* 1997)

2.6.7.4 Amino acid sequence substitutions in HAdV loop2 sequences

To demonstrate amino acids substitutions in the hyper-variable and the fairly conserved regions of hexon loop 2 sequences, a multiple sequence alignment of translated sequences was perfomed using Muscle version 3.8.3.1 (Edgar, 2004) and visualised by GeneDoc version 2.6.002 (Nicholus *et al.,* 1997).

2.6.7.5 Analysis of synonymous and non synonymous mutations

Synonymous and non synonymous mutations were calculated using MEGA version 5.2 (Tamura *et al.*, 2011). The analysis involved computation of the overall mean distance in a set of aligned sequences of the same serotype together with corresponding reference sequences. Two substitution models were used separately for comparison purpose; the Nei-Gojobori (Jukes-cantor) method and the Kumar (Kimura-2-parameter) method. Non synonymous- synonymous ratio (dN/dS) of <1 for sequences of the same serotype was used to imply a purifying selection of the sequences. Likewise dN/dS ratio of >1 implied that the sequences were under a positive selection pressure.

CHAPTER THREE

3.0 RESULTS

3.1 Detection of cytopathic effect (CPE) in Hep2 cells putatively infected with HAdV.

Following inoculation of the 17 adenovirus virus stocks and subsequent growth in Hep2 cells, presence of adenovirus was detected by observation of CPE. The CPE consisted of rounding and detachment of the cells from the substratum. Sixteen virus stocks (94%) exhibited CPE after 7 days of inoculation. Fig. 7 below shows an example of CPE observed from Hep2 cell inoculated with one of the virus stocks. From a single virus stock ID number NNY-10-11-040, CPE in the Hep2 cells was only observed after an extended incubation of 10 days post inoculation (table 4).

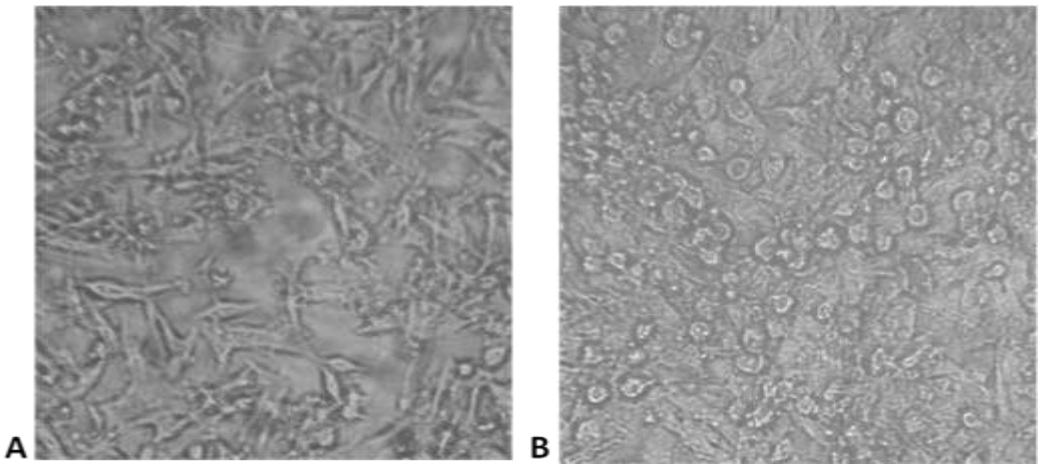


Figure 7. Cytopathic effects (CPE) on cultured Hep2 cells upon infection with a HAdV stock. Virus infection of the monolayer of Hep2 cells resulted in cell morphology changes. Fig 7-A: monolayer of uninfected Hep2 cells with no CPE. Fig 7-B: a monolayer of the same cells 7 days after inoculation with virus stock number NNY-10-06-016. Observations were made at magnification x200 using Olympus microscope, Model 299682 (Olympus, Tokyo, Japan).

3.2 Confirmation of presence of HAdV in Hep2 cells displaying CPE.

Presence of HAdV in the inoculums was confirmed by immunofluorescence assay. In this assay HAdV-specific monoclonal antibodies labeled with fluorescein isothiocyanate (FITC) - a fluorescent dye- emitted apple green fluorescence that was easily visualized by a fluorescence microscope under ultra violet light. The apple green fluorescence was observed in all the sixteen tubes reported in section 3.1 above where CPE had been observed by the seventh day of inoculation. No fluorescence was detected in the tube inoculated with virus stock NNY-10-11-040. A typical result of this result is shown in Fig. 8 (A-C), below. IFA results for all the 17 virus stocks used in this study are given in appendix I.

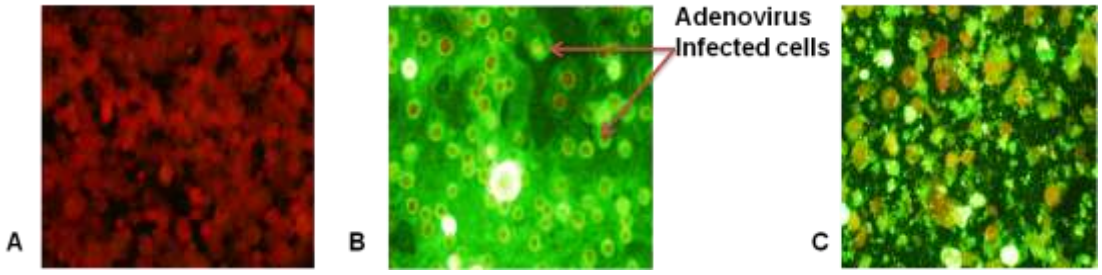


Figure 8. A typical Immunofluorescence assay (IFA) result of adenovirus infection of Hep-2 cells. Infected cells fluoresce with the apple green color while uninfected cells appear with a background dull red color. Fig 8-A: negative control slide. Fig 8-B: adenovirus positive control slide showing adenovirus specific monoclonal antibodies on infected cells. Fig 8-C: Hep2 cells infected with virus stock number NNY-10-06-016. The observations were made at Magnification, x400 using Olympus BX51 Fluorescence Microscope (Olympus, Tokyo, Japan).

Table 4 Summary of serological detection of adenovirus in Hep2 cell culture. Presence of virus by CPE observation is shown by a "YES" and absence by "NO". Adenovirus presence in culture by IFA is denoted by "+" and absence of virus is denoted by "-".

Sample I.D	Presence of CPE?		IFA
	Day 7 post inoculation	Day 10 post inoculation	
NNY-10-06-016	YES		+
NNY-10-06-018	YES		+
NNY-10-06-032	YES		+
NNY-10-06-073	YES		+
NNY-10-07-009	YES		+
NNY-10-07-033	YES		+
NNY-11-09-001	YES		+
NNY-10-11-022	YES		+
NNY-10-11-034	YES		+
NNY-10-11-040	NO	YES	-
NNY-11-03-077	YES		+
NNY-11-07-001	YES		+
NNY-11-10-044	YES		+
NNY-11-11-013	YES		+
NNY-11-12-004	YES		+
NNY-12-03-027	YES		+
NNY-12-03-029	YES		+

3.3 Determination of purity and concentration of viral DNA extracted from IFA-positive Hep2 cells.

The extracted DNA from each of HAdV isolates yielded an absorbance ratio at 260/280nm of more than 1.6 (table 5). Upon measurement of absorbance at 260nm, the concentrations from the virus DNA ranged between 245.4 to 858.4ŋg/µl (table 5).

Table 5 Purity and concentrations of DNA extracts from HAdV isolates. Concentration of each DNA sample is given in ŋg/µl.

Sample Unique ID	Absorbance Ratio 260/280nm	DNA concentration (ŋg/µl)
NNY-10-06-016	1.8	779.2
NNY-10-06-018	1.7	399.2
NNY-10-06-032	1.8	570.6
NNY-10-06-073	1.8	736.8
NNY-10-07-009	1.8	858.4
NNY-10-07-033	1.8	460.9
NNY-11-09-001	1.8	528.5
NNY-10-11-022	1.8	462.8
NNY-10-11-034	1.7	462.7
NNY-10-11-040	1.8	245.4
NNY-11-03-077	1.7	605.6
NNY-11-07-001	1.8	595.0
NNY-11-10-044	1.8	458.6
NNY-11-11-013	1.8	467.3
NNY-11-12-004	1.8	365.7
NNY-12-03-027	1.8	529.6
NNY-12-03-029	1.6	504.7

3.4 Detection and purification of the hexon gene amplicons.

PCR amplicons were detected as fluorescent bands after an agarose gel electrophoresis and exposure to UV light. The amplicon bands of approximately 500bp were seen in all cases where HAdV DNA and hexon gene-specific primers had been included (Fig. 9).

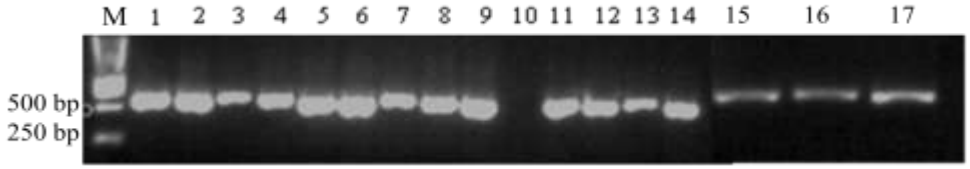


Figure 9. Gel electrophoresis of HVR 7 PCR amplicons. Lane M represents 100 bp DNA ladder (Promega Madison WI USA) used as molecular weight marker. PCR amplicons are shown in lanes 1-17 representing DNA samples used in this study 1.NNY-10-06-016, 2.NNY-10-06-018, 3.NNY-10-06-032, 4.NNY-10-06-073, 5.NNY-10-07-009, 6.NNY-10-07-033, 7. NNY-11-09-001, 8.NNY-10-11-022, 9.NNY-10-11-034, 10.NNY-10-11-040, 11.NNY-11-03-077, 12.NNY-11-07-001, 13.NNY-11-10-044, 14.NNY-11-11-013, 15.NNY-11-12-004, 16.NNY-12-03-027, 17.NNY-12-03-029

Upon purification of the PCR amplicons by ExoSAP-IT to remove unused primers, DNTPs, and DNA template, clean PCR amplicons were obtained (Fig. 10). The concentrations of the DNA of the purified amplicons ranged between 118.1-526.7 ɲg/µl (Table 6).

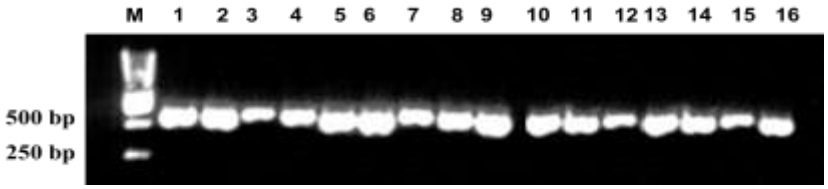


Figure 10. Gel electrophoresis of cleaned PCR amplicons. Lane M represents 100 bp DNA ladder (Promega Madison WI USA) used as molecular weight marker. Purified PCR amplicons are shown in lanes 1-16 representing DNA samples used in this study 1.NNY-10-06-016, 2.NNY-10-06-018, 3.NNY-10-06-032, 4.NNY-10-06-073, 5.NNY-10-07-009, 6.NNY-10-07-033, 7.NNY-11-09-001, 8.NNY-10-11-022, 9.NNY-10-11-034, 10.NNY-11-03-077, 11.NNY-11-07-001, 12.NNY-11-10-044, 13.NNY-11-11-013, 14.NNY-11-12-004, 15.NNY-12-03-027, 16. NNY-12-03-029

Table 6 Concentrations of purified pre-sequencing PCR amplicons. Concentration of each DNA sample is given in ng/µl.

Sample Unique ID	DNA concentration (ng/µl)
NNY-10-06-016	146.7
NNY-10-06-018	118.1
NNY-10-06-032	371.1
NNY-10-06-073	526.7
NNY-10-07-009	121.6
NNY-10-07-033	154.4
NNY-11-09-001	119.3
NNY-10-11-022	380.1
NNY-10-11-034	484.6
NNY-11-03-077	310.0
NNY-11-07-001	366.0
NNY-11-10-044	198.8
NNY-11-11-013	293.3
NNY-11-12-004	314.8
NNY-12-03-027	316.8
NNY-12-03-029	125.7

3.5 Nucleotide sequences of HAdV HVR7 isolates of the study

The ABI 3500XL genetic analyzer yielded plus and minus strand nucleotide sequences for each amplicons with very high signal/noise ratio (Fig. 11). Contig assembly yielded consensus nucleotide sequences for each amplicons. A portion of a representative contig assembled in DNAbaser is shown in Fig. 12.

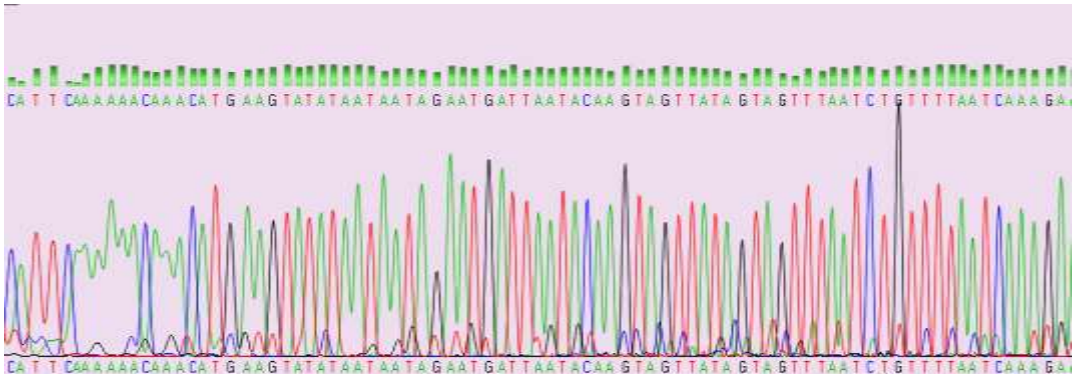


Figure 11. Chromatogram trace of a portion of one of the sequenced HAdV's loop-2 fragment. The background peak signal is very low compared to signal peaks of the called bases.

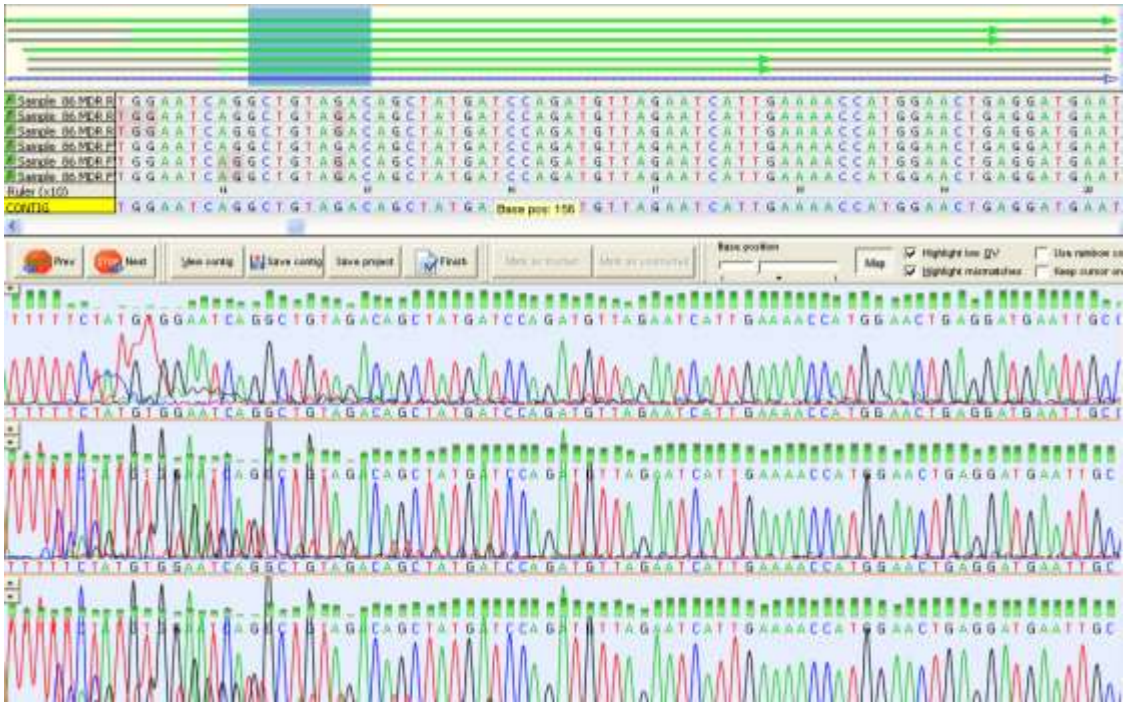


Figure 12. A portion of an assembled contig profile for one of the HAdV loop-2 fragment. The assembly was built using DNAbaser version 10 (Biosoft) available at http://www.dnabaser.com.

3.6 Nucleotide identity search analysis

In order to identify sequences in the GenBank similar to the assembled contigs, BLAST analyses were carried out. The results revealed that the sixteen sequences had very high similarities to human adenovirus sequences with expectation (E) values of zero. A representative BLAST search result is shown in Fig.13. A summary of the BLAST results using the nucleotide sequences for all the sixteen sequences is shown in table 7 below. The sixteen sequences had nucleotide identities of 98-99% compared to the next homologous reference sequence in the GenBank database.

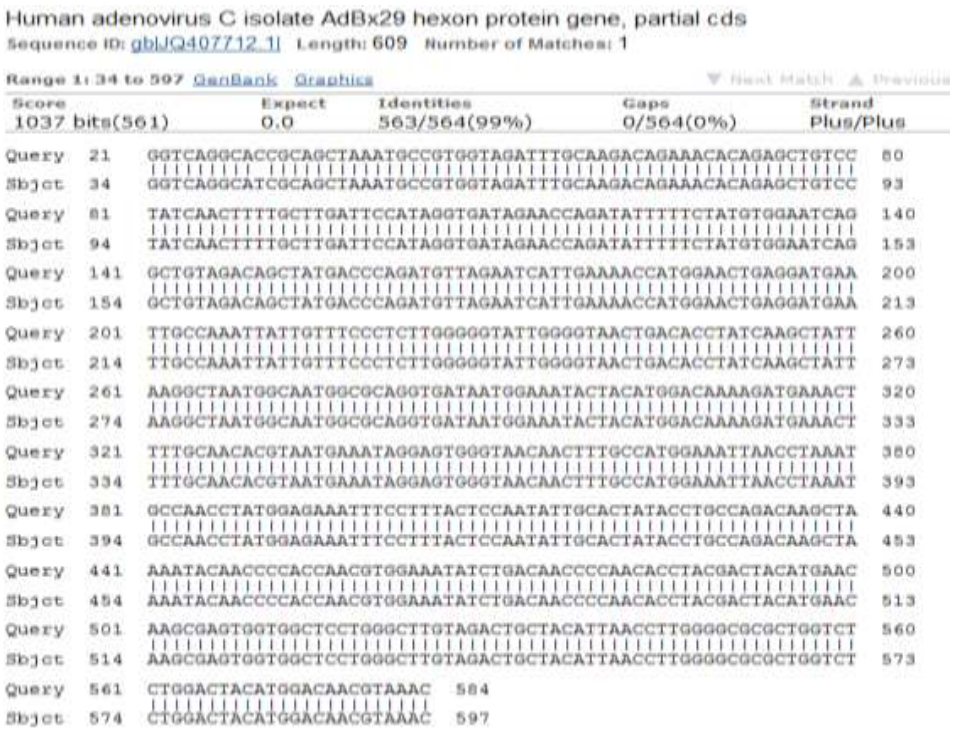


Figure 13. A representative result of nucleotide BLAST (nblast) of the sequenced HAdV loop-2 fragment. NNY-10-06-016 sequence is highly similar to a portion of the human adenovirus C isolate Adbx29 whose GenBank accession number is JQ407712.1 with a single mismatch in a 564 region of the loop2 and E value of 0.

Table 7 Summary of nucleotide BLAST (nblast) results of the sequenced isolates.

Sample Unique Identity	Accession Numbers showing the highest % identity score	Nucleotide Identity score (%)	E Value
01	KC747631.1	99	0.0
02	AF515814.1	98	0.0
03	HQ535662.1	99	0.0
04	HQ535662.1	99	0.0
05	AF053085.1	98	0.0
06	AF515814.1	98	0.0
07	AB436561.1	99	0.0
08	FJ943599.1	99	0.0
09	KC747631.1	99	0.0
10	AB436561.1	98	0.0
11	AF515814.1	98	0.0
12	KC747641.1	99	0.0
13	AF053085.1	99	0.0
14	AF053085.1	98	0.0
15	AB436561.1	98	0.0
16	JQ407712.1	99	0.0

3.7 Location of hyper variable region 7

The location of HVR7 within the loop 2 region of the hexon gene was determined by multiple sequence alignment using HAdV C2 (Acc. No. AJ293903.1) complete hexon coding sequences as a guideline. The analysis revealed that amplicons spanned nucleotide position 1104-1640 of the HAdV C2 hexon gene (Fig.14). The hyper variable region 7 (HVR7) is located at nucleotide position 1330-1400 within HAdV C2 hexon gene (Fig. 14). From these multiple sequence alignment, the location of HVR7 in all the sixteen sequences was clearly located (Fig. 14).

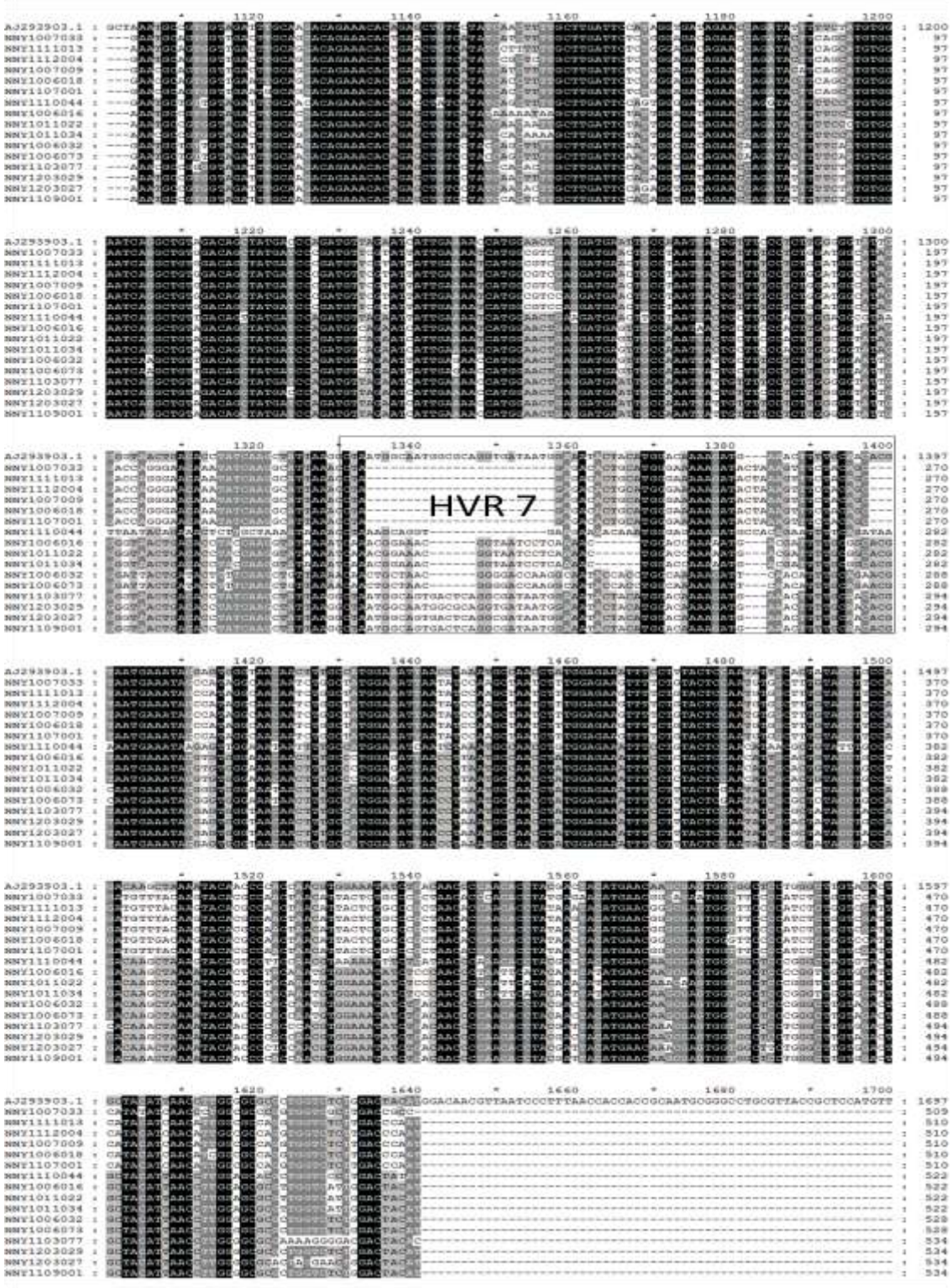


Figure 14. Multiple sequence alignment of the loop 2 region of the sixteen HAdV nucleotide sequences with HAdV 2 hexon gene. In this alignment, the hyper variable region is located in nucleotide position 1330-1400 based on the 2910bp coding sequences of the hexon gene of HAdV 2 (Acc No. AJ293903.1).

3.8. Phylogenetic analyses

Phylogenetic reconstruction using the sixteen HAdV sequences in this study together with reference sequences from respiratory HAdV species B, C and E yielded the phylogentic tree shown in Fig.16. The sixteen HAdV sequences grouped into two main genetic clusters belonging to HAdV B and C species. Six sequences (NNY-10-06-018, NNY-10-07-009, NNY-10-07-033, NNY-11-07-001, NNY-11-11-013, NNY-11-12-004) clustered among the HAdV B7 reference serotypes while the remaining ten clustered with HAdV C reference serotypes. The grouping in the HAdV C genetic clusters further split into four distinct clades representing serotype 1 (NNY-10-06-016, NNY-10-11-034, NNY-10-11-022) serotype 2 (NNY-12-03-029, NNY-11-03-077, NNY-12-03-027, NNY-11-09-001) serotype 5 (NNY-11-10-044) and serotype 6 (NNY-10-06-073, NNY-10-06-032). Overall, the HAdV distribution was HAdV C serotype 1 at 25%, HAdV C serotype 2 at 25%, HAdV C serotype 5 at 6.25%, and HAdV C serotype 6 at 12.5%. HAdV B serotype 7 accounted for 31.25% of the sequences.

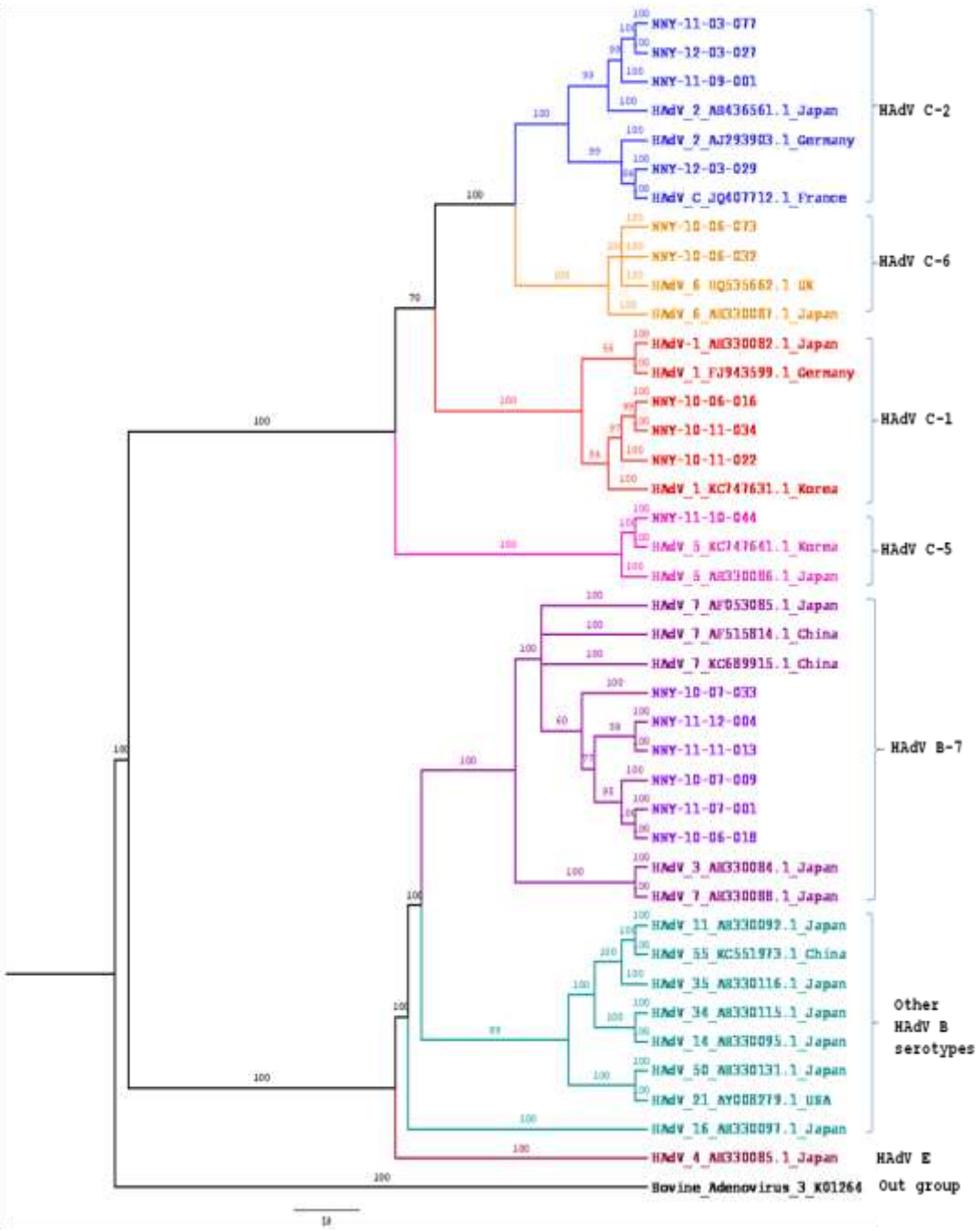


Figure 15. Phylogenetic analysis using the nucleotide sequences of the ε determinant loop2 from HAdVs under study. The phylogeny was inferred by Bayesian method from an alignment performed using Muscle3.8 and the tree diagram visualized using FigTree v1.4.0. Reference hexon sequences from each serotype of respiratory HAdV species B, C and E are identified with their GenBank accession numbers at the end of the branches. The HVR7 loop2 sequences are identified with unique codes as used in this study. Bovine Adenovirus 3 was used as an out-group. Sequences belonging to the same species or serotype clusters are highlighted with brackets. The posterior probability values are expressed as percentages at each branch.

3.9 Amino acid substitutions in HAdV loop2

The region of the hexon protein analyzed consisted of approximately 140 amino acids from position 399 to 545 (HAdV C2 hexon sequences numbering [Acc.No. AJ293903.1] (Fig.16). Thus the hyper variable region 7 is located within this analyzed region, being positioned at amino acid positions 433-467 according to HAdV C serotype 2. This region is characterized by amino acids substitutions, insertions and deletions (Fig.16).

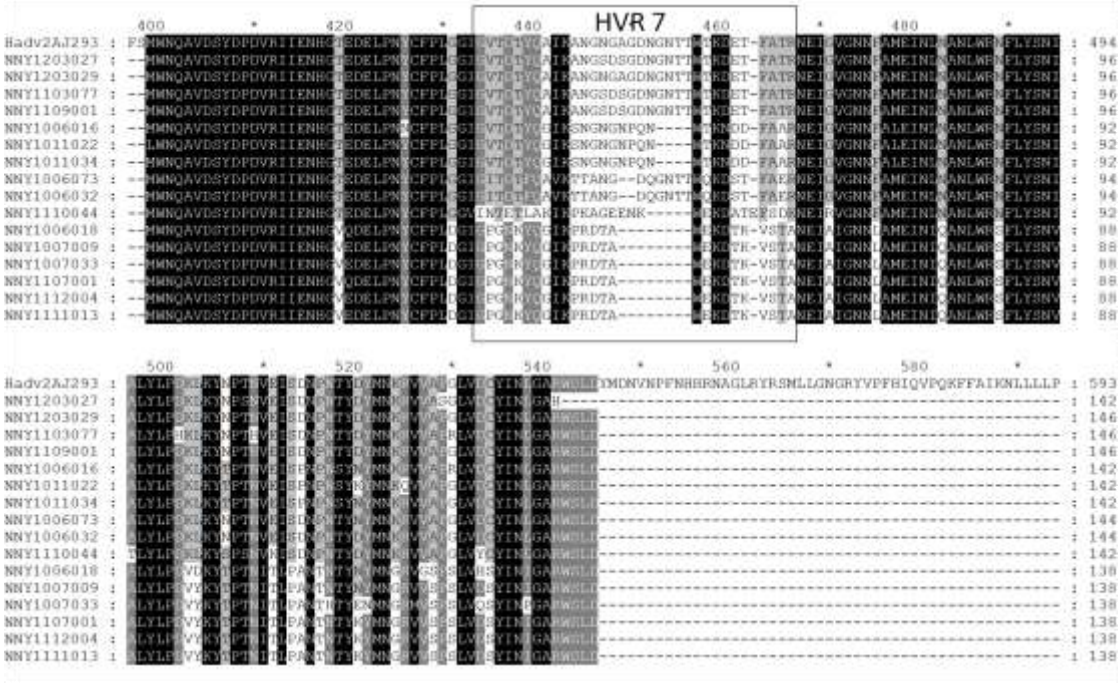


Figure 16. Multiple sequence alignment of all the sixteen HAdV loop-2 sequences deduced amino acid sequences. HAdV2 amino acid sequence numbering [accession number AJ293903.1] was used as the reference. The HVR7 is highlighted.

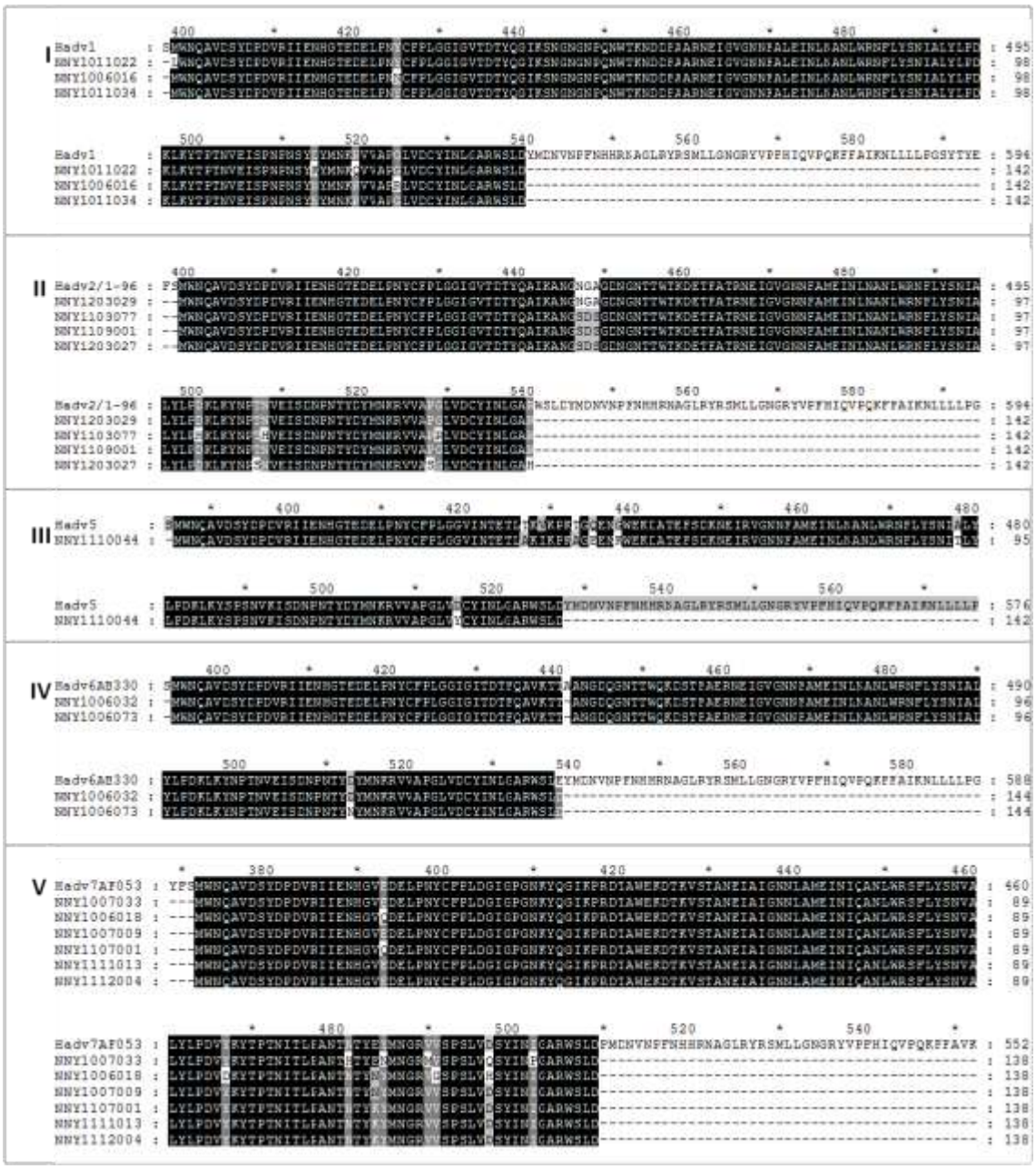


Figure 17 (I-V). Amino acid polymorphisms of the HAdV isolates within hexon loop-2 region in relationship to reference strains. The phylogenetically classified HAdV serotypes 1, 2, 5, 6 and 7 sequence clusters are presented in figures 17I, 17II, 17III, 17IV, and 18V respectively. Reference strains for each of the HAdV serotype cluster appear above the other sequence(s) in each of the alignments.

Amino acid polymorphisms in loop 2 of the hexon gene in each of the five genetic clusters representing the HAdV serotypes as determined by phylogenetic classification (Fig.16) are shown in Fig. 17 I-V.

Overall, compared to the reference strain HAdV C serotype 1 (Acc. No. AB330082.1), there were 5 polymorphic sites in the phylogenetic cluster representing serotype C-1 (Fig. 17-I). Individually, the substitutions were distributed as follows: three in NNY-10-11-022, three in NNY-06-016 and one in NNY-10-11-034. Compared to the reference strain, the protein sequence of the NNY-10-11-022 isolate had a leucine (L) to methionine (M) substitution at position 398. Two more amino acids substitutions were observed in this sequence at position 514, where aspartate (D) was replaced by lysine (K) and at position 519 where arginine (R) replaced glutamine (Q). There were three substitution events in isolate NNY-10-06-016. These included substitutions at positions 425 where tyrosine (Y) was substituted with asparagine (N), at position 514 where aspartate (D) was replaced by asparagine (N) and at position 524 where glycine (G) was substituted for by arginine (R). There was only one substitution in NNY-10-11-034 at position 514 involving a replacement of aspartate (D) with asparagine (N). Overall, position 514 was the most polymorphic site because all the Kenyan isolates differed from the reference strain at this site.

In the HAdV C serotype-2 genetic cluster, there were 9 positions at which 15 mutations were observed (Fig. 17-II). NNY-11-09-001, NNY-12-03-027 and NNY-11-03-077 exhibited three consecutive substitutions at positions 447, 448, and 449 (HAdV C serotype-2 Acc. No. AJ293903.1 numbering). At each of these three sites in the three sequences, the substitutions consisted of asparagine (N) to serine (S) at position 447,

glycine (G) to aspartate (D) at position 448 and alanine (A) to serine (S) at positions 449. There were two more substitutions in NNY-12-03-027 at positions 507 and 540 involving threonine (T) to serine (S) and arginine (R) to histidine (H) respectively. There was also a substitution of proline (P) to serine (S) at position 528 in this sequence. Three more substitutions were observed in NNY-11-03-077. These occurred at positions 500, 508 and 529 and involved replacement of aspartate (D) with histidine (H), asparagine (N) with histidine (H) and glycine (G) with arginine (R) respectively.

HAdV C serotype-5 genetic cluster contained a single test sequence NNY-11-10-044, and was compared with reference sequence of HAdV C serotype-5 accession number AB330086.1 (Fig. 17-III). There were 7 positions at which amino acid substitutions were observed. These included position 429 where valine (V) was replaced with isoleucine (I), positions 427 & 433 where Threonines (T) were substituted by alanines (A), while the remaining substitutions occurred at positions 435, 438, 478 and 515 and involved substitutions of glutamine (Q) with glutamate (E), glycine (G) with lysine (K), alanine (A) with threonine (T) and aspartate (D) with tyrosine (Y) respectively.

Compared to the reference strain HAdV C serotype 6 (Acc. No. AB330087.1), two of the Kenyan isolates i.e. NNY-10-06-032 and NNY-10-06-073 that grouped in the HAdV C serotype-6 genetic cluster had a deletion of alanine (A) at position 401 (Fig. 17-IV). In contrast, isolate NNY-10-06-073 had an additional mutation involving substitution of aspartate (D) by asparagine (N) at position 513.

Amongst the five phylogenetic clusters observed in this study, HAdV B serotype-7 cluster constituted the bulk of the sequences. There were 9 positions that exhibited amino

acid changes accounting for the 15 individual mutations in this cluster. NNY-11-07-001 had a substitution of glutamate (E) with glutamine (Q) at position 393 (HAdV B serotype-5 Acc. No. AF053085.1 numbering). NNY-11-07-001, NNY-11-12-004 and NNY-11-11-013 each had E484K substitutions involving replacement of glutamate (E) with lysine (K). NNY-10-06-018 and NNY-10-07-009 also had substitutions E484N at this position where the glutamate (E) residue was replaced by asparagine (N). NNY-10-06-018 had four more substitutions consisting of E393Q, Y467D, V491G and D497H. Five substitutions were observed in NNY-10-07-033; these included V490M, I502P, N481H, Y485N, and D497Q.

3.10 Hexon loop2 nucleotide sequence analyses for dN/dS

The synonymous (dS) and non-synonymous (dN) rates of nucleotide substitutions in the loop-2 sequences are given in table 8 below. The table also shows a summary of the dN/dS ratio as calculated following two substitution models; Nei-Gojobori and Kumar substitution models. Values of dS and dN for sequences belonging to the same serotype are given as computed using the two substitution models. Sequences in the HAdV C serotype 1and serotype 6 gene clusters had dN/dS ratio of 1. Sequences constituting the HAdV species C serotype-2 genetic clusters yielded dN/dS ratio of less than 1 while in HAdV species C serotype 5 and HAdV species B serotype 7 the ratio of dN/dS was greater than 1 (Table 8).

Table 8. Summary of nucleotide sequences distance estimation by synonymous and non-synonymous substitutions

	Nei-Gojobori substitution model			Kumar substitution models		
	dS	dN	dN/dS	dS	dN	dN/dS
HAdV C serotype 1 genetic cluster	0.2052	0.2746	1.0	0.2291	0.2523	1. 0
HAdV C serotype 2 genetic cluster	0.0208	0.0133	0.64	0.0335	0.0136	0.4
HAdV C serotype 5 genetic cluster	0.0067	0.0417	6.2	0.0026	0.046	17.7
HAdV C serotype 6 genetic cluster	0.0109	0.0118	1.0	0.01	0.0109	1.0
HAdV B serotype 7 genetic cluster	0.013	0.0283	2.2	0.0185	0.0263	1.4

CHAPTER FOUR

4.0 DISCUSSION

HAdVs are recognized disease causing pathogens especially in pediatric patients, immune compromised patients and in the military recruits living in close quarters (Potter *et al.*, 2012). Specific HAdV serotypes have been associated with particular clinical conditions (Sharma *et al*., 2009) and therefore an understanding of HAdV serotypes circulating in a population is crucial in planning and management of the disease caused by these viruses and ensuring timely therapeutic intervention. Adenoviral infections are cumbersome to treat just like other viral infections, and for that reason early diagnosis will aid in efficient supportive management of disease symptoms. Understanding the species and serotypes of adenoviruses in circulation will aid policymakers and clinicians in public health preparedness for disease due to these viruses. It will also assist clinicians in determining the most appropriate course of treatment to administer at the earliest time possible following the onset or outbreaks of the infection.

In this study serological and molecular approaches were used to identify and characterize respiratory human adenoviruses from nasopharyngeal specimens from pediatric patients presenting with influenza like illness. In the period of June 2010 to June 2012, 1879 nasopharyngeal swab specimens from symptomatic patients were collected from New Nyanza Provincial General Hospital through the influenza surveillance program at USAMRU-K. From these specimens, 17 cases infected with adenovirus were reported. Thus the prevalence of respiratory HAdV at this hospital during the study period was approximately 1%. This prevalence is very low compared to prevalence of infection by influenza viruses at the same hospital during the same time, which was at approximately

25% (NIC lab reports), showing that respiratory disease burden at this site due to respiratory HAdV was very low.

Whereas preliminary diagnosis had been performed on the patient samples to determine prevalence of infection by respiratory HAdV no further analyses had been carried out on the samples to provide information regarding serotypes and molecular characteristics of adenoviruses implicated in those cases. Serologic and molecular characteristics are important in understanding the epidemiology of any viral diseases. For this reason, this study sought to bridge this gap in the knowledge by systematically carrying out serological and molecular characterization of those 17 respiratory human adenoviruses.

The first part of this study involved re-inoculation and serological identity confirmation of the isolated adenovirus. The need to confirm the presence of adenovirus in the archived stocks was informed by the fact that virus viability is affected by handling and storage conditions (Leland *et al.,* 2007). Re-inoculation of the virus stocks in cell culture was also important because it amplified the viruses leading to increased virus titers, a prerequisite for obtaining adequate viral DNA for subsequent molecular characterization procedures. Surprisingly, CPE was observed in all but one of the inoculated culture after seven days. The failure of virus stock NNY-10-11-040 to generate CPE in culture within the first week of inoculation may have been due to either total absence of virus in the stock, hence wrong diagnosis of this patient having been infected with a respiratory HAdV, or very low viral load in the stock that required extended periods of incubation to allow the low virus load to amplify before CPE is established. Whereas CPE was eventually established on the 10^{th} day, this CPE was not due to respiratory HAdV because subsequent serological tests confirmed the absence of HAdV. The CPE

observed after the 10th day could have been caused by accumulation of metabolic waste products in the cell culture or by another virus that has the Hep2 cell line as its permissive host such as coxsackievirus. Due to unavailability of test reagents, this culture was not tested for the presence of coxsackievirus. In 16 cultures that were confirmed to have respiratory HAdV by IFA, disparity in the number of apple green staining of Hep2 cells was evident indicating that each of the virus stocks had varying viral loads.

For further downstream molecular analyses including PCR and nucleotide sequencing to be successful, a pure sample of virus DNA was required. Spectroscopic measurements in the UV range have been shown to be a very quick and accurate measure of purity of nucleic acid where the absorbance ration (due to presence of purine and pyrimidine rings in the DNA bases) of pure DNA at 260/280nm of approximately 1.8 generally accepted as "pure" for DNA (ThermoFisherScientificTechnicalsupport). In the current study, the ratio of absorbance at 260/280nm in nearly all the DNA extracts were ~1.8 indicating that the purification method yielded DNA that was amenable to further downstream manipulations. The concentrations from each virus stock ranged between 245.4 to 858.4ŋg/µl, again attesting to the variability in the viral loads between disparate virus stocks. From this consideration alone, it can be said that virus stock NNY-10-11-040 which yielded 245.4ŋg/µl of DNA had the lowest virus load and that stock virus NNY-10-07-009 that yielded 858.4ŋg/µl of DNA had the highest viral load. However, since Hep2 cells supports growth of a variety of other viruses besides adenovirus the DNA concentrations may not have solely been of adenovirus origin. The fact that all virus extracted DNAs showed only a single band with no smears when they were run on the

agarose gels confirms that the Qiagen kit used in this study has indeed been optimized for virus DNA extraction, suggesting that the entire DNA was of adenovirus origin. Since a minimum of 50ŋg/µl of DNA per sequencing reaction is required for a successful Sanger sequencing reaction, the purification process yielded adequate quantities for downstream sequencing because at least 50ŋg/µl of purified amplicon DNA was obtained per isolate. This was also confirmed by the successful nucleotide sequencing of the amplicons as discussed below.

BLAST analyses showed that all the sequences generated from the 16 virus stocks were of respiratory human adenoviruses with sequence homology of greater than 98% compared to the next homologous adenovirus sequence in GenBank with E values of 0.0. The low E-values, a statistical parameter that describes the number of hits one can "expect" to obtain by chance when searching a database, showed highly significant matches in nucleotides between the 16 isolates and database adenovirus entries. Thus, the high sequence homology confirmed the CPE and IFA results that indeed the virus stocks belonged to adenovirus group of viruses. Indeed, past studies have demonstrated that nucleotide homology of more than 97.5% is a valid criteria for determining the identity of adenovirus isolates (Madisch *et al.,* 2005).

Hexon epsilon fragment is one of the species and serotype specific determinant regions in adenovirus genome (Xu *et al.*, 2000). Other serotype-specific regions are found in the penton and fiber genes (Xu *et al*., 2000). Previous studies have demonstrated that analyzing either loop1 or loop2 regions of hexon epsilon determinant can lead to classification of adenoviruses into serotypes. Hexon loop2 region used in this study has been described as the most simple but sufficient approach for molecular typing of human

adenoviruses (Madisch *et al.*, 2005). Thus using the nucleotide sequences of loop 2 region in a phylogenetic reconstruction analysis, the 16 virus isolates clustered into distinct clades with high (100%) overall posterior probability values. The general overall high posterior probabilities values indicated that the viral sequences were adequately identified as belonging to a particular species or serotype. The posterior probability values are however low (e.g. 54%) particularly at nodes separating putative strains of the same viral isolate, indicating that strain separation may be more difficult to achieve using molecular phylogenetic analyses.

Phylogenetic analyses using the hexon loop 2 nucleotide sequences demonstrated that during the study period, HAdV C was the predominant species accounting for 68.75% of the reported cases. During the study period four serotypes of species C (i.e. serotypes C1, C2, C5 and C6) were identified in circulation. However, for HAdV B species, only a single serotype 7 was identified. Interestingly, despite being the only species B serotype in circulation during this period, HAdV 7 was the predominant serotype accounting for 31.25% of all isolates. The HAdV B serotype 7 has been associated with outbreaks and in certain cases fatalities and therefore prevalence of this serotype in the Kenyan population has public health significance (Kim *et al*., 2003; Lewis *et al*., 2009; O'Flanagan *et al*., 2011).

HAdV species E was not observed among the isolates and therefore was not in circulation during the study period. Previous studies have suggested that HAdV E might have emerged through recombination involving strains of HAdV B serotype 16 (Shenk, 2001.). Since only HAdV B7 serotype was present in the Kenyan isolates during the study period

without any other species B serotypes to recombine with, this could explain the absence of HAdV E among the Kenyan isolates.

As was expected, sequence analyses revealed the presence of the hyper variable region 7 within the fairly conserved hexon epsilon fragment across the entire HAdV serotypes. The hyper variable region is characterized by insertions and deletions of nucleotides/ amino acids which together with mutations seen in the fairly conserved segment of hexon gene contribute to the genetic diversity of the adenoviruses. The variability in the hyper variable region is a key immunological determinant in adenovirus driven by immune pressure since the hexon protein is outwardly oriented in adenovirus capsid. Therefore, any change occurring at nucleotide level that amounts to alteration in the amino acid sequence of the hexon protein has profound effect on the virus capacity to interact with its host receptor. Remarkable genetic variability in hexon protein as wells as in the other immunological significant proteins i.e. penton and fiber may result to emergence of new genotypes of adenoviruses.

Genetic diversity in a population is a good indicator of an ongoing evolutionary process. Therefore, analysis of synonymous and non-synonymous mutations in the isolates that were identified as belonging to the same serotype was carried out to determine if the sequences were under selection pressure. Under purifying selection pressure the dN/dS ratio is expected to be less than 1. The dN/dS analysis in sequences constituting HAdV C1 and HAdV C6 genetic clusters showed that those sequences were not under selection pressure. Selection pressure was observed in sequences constituting the HAdV C2, HAdV C5 and HAdV B7 genetic clusters. The HAdV C2 genetic cluster yielded dN/dS ratio of less than 1 indicating the sequences were under a purifying selection pressure.

The dN/dS ratio of sequences constituting the HAdV C5 and HAdV B7 genetic clusters was greater than 1 implying that these sequences were under positive selection pressure.

Normally genes exhibiting positive selection pressure indicate that the host organism is continually undergoing evolution in order to adapt to changes in the environment. Consequently, the Kenyan HAdV C5 and HAdV B7 were found to be undergoing an evolutionary process which signifies instability in their genomes. HAdV B7 was also the predominant serotype circulating during the study period. These serotypes are of great significance from a clinical perspective. HAdV B7 is one of the serotypes that has been implicated in adenovirus outbreaks and in some cases fatalities (Kim *et al.*, 2003; Lewis *et al.*, 2009; O'Flanagan *et al.*, 2011); while HAdV C5 is a popular vaccine vector for many test drugs including Kenya Prime Boost HIV vaccine that has previously been on clinical trials in Kenya (Kibuuka *et al.*, 2010). Therefore, there is need to continually carry out surveillance of these adenovirus serotypes in order to track their evolution. Drastic evolution of these serotypes could eventually result in emergence of more virulent adenoviruses which can be difficult to manage in case of severe outbreaks. Policy makers in the Kenyan health sector should consider introducing the adenovirus vaccine currently used by the US military in the Kenyan population especially among the most vulnerable in the general population like patients in pediatric wards and military recruits living in close quarters. The adenovirus used by the US military is a divalent vaccine with components of B7 and B14 serotypes (Potter *et al.*, 2012). A Kenyan specific vaccine may include C5 and B7 components. Adenovirus vaccination programs have been successful in reducing the prevalence of adenovirus infections in other parts of the world and therefore such efforts should be embraced here as well.

The main goal of this study was to identify and classify adenovirus isolates. This has been successfully achieved by sequencing the HVR 7 region of hexon gene; a well known adenovirus antigenic site important in adenovirus serotyping (Crawford *et al*.,1996). This work has for the first time demonstrated the genetic variations of the Kenyan respiratory human adenoviruses while at the same time confirmed the findings of a previous study (Sarantis *et al.*, 2004) that used hexon loop 2 sequences to detect and classify HAdVs that were known by that time.

4.1 STUDY LIMITATIONS

Whereas this study has contributed to the knowledge gap that existed in the molecular characteristics of respiratory human adenoviruses found in Kenya, like most studies, it had a few limitations. For example, only one of the three diagnostically useful genes in detection and genotyping of adenoviruses, the hexon gene, was used. Further studies utilizing nucleotide sequences of the penton and fiber genes, preferably full-length genes and even better full genome sequences of the adenoviruses will afford a more comprehensive picture of the genetic characteristics of the Kenyan respiratory human adenoviruses.

CHAPTER FIVE

5.0 CONCLUSION AND RECOMMENDATION

The nucleotide sequences of the hyper-variable loop 2 of the HAdV hexon gene have been applied in characterization of respiratory adenoviruses that circulated at New Nyanza Provincial General hospital during the study period into species B and C. Specifically, HAdV B serotype 7 and HAdV C serotypes 1, 2, 5 and 6 were found to have been in circulation during the study period. There was significant genetic instability in the hexon gene of the HAdV species B7 and HAdV species C5 seen at this site compared to those from other parts of the world implying that there is a continuing evolution of these Kenyan respiratory HAdVs. I therefore recommend that policy makers in Kenya health sector should consider introducing adenovirus vaccine containing HAdV serotypes 5 and 7 components to be administered to pediatrics in western Kenya since they are the most vulnerable in case of an outbreak.

To gain a complete understanding of the evolutionary process seen in isolated HAdVs, in future, researchers should carry out whole genome sequencing and analyses of these viruses to determine recombination events which are the major driving force in emergence of new adenovirus serotypes (Robinson *et al.*, 2013). Furthermore, the present study utilized human adenoviruses obtained from a single site in western Kenya whose findings were used as generalizations to represent the whole country and therefore future studies should include viruses drawn from the whole country to give a more comprehensive description of the HAdVs circulating in Kenya.

REFERENCES

*A*dhikary, A. K., Inada, T., Numaga, J., Suzuki, E., Ushijima, H., Banik, U.,Okabe, N. (2004). Characterisation of hexon and fibre genes of a novel strain of adenovirus involved in epidemic keratoconjunctivitis. *J Clin Pathol, 57*(1), 95-97.

Akalu, A., Seidel, W., Liebermann, H., Bauer, U., & Döhner, L. (1998). Rapid identification of subgenera of human adenovirus by serological and PCR assays. *J Virol Methods, 71*(2), 187-196.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410. doi: 10.1016/s0022-2836(05)80360-2

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-3402.

Andrew Rambaut. (2007). Figtree version 1.4.0. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/

Artimo, Panu, Jonnalagedda, Manohar, Arnold, Konstantin, Baratin, Delphine, Csardi, Gabor, de Castro, Edouard, . . . Stockinger, Heinz. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research, 40*(W1), W597-W603. doi: 10.1093/nar/gks400

BioSoft, Heracle. DNA Baser Sequence Assembler v10.

Cevenini, R., Varoli, O., Rumpianesi, F., Mazzaracchio, R., Nanetti, A., & La Placa, M. (1985). A two-year longitudinal study on the etiology of acute diarrhea in young children in Northern Italy. *Microbiologica, 8*(1), 51-58.

Chen, H., Xiang, Z. Q., Li, Y., Kurupati, R. K., Jia, B., Bian, A., . . . Ertl, H. C. (2010). Adenovirus-based vaccines: comparison of vectors from three species of adenoviridae. *J Virol, 84*(20), 10522-10532. doi: 10.1128/jvi.00450-10

Crawford-Miksza, L., & Schnurr, D. P. (1996). Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues. *J Virol, 70*(3), 1836-1844.

Cusi, M. G., Valensin, P. E., Barberi, A., Guglielmetti, P., & Rossolini, A. (1986). Viral childhood gastroenteritis. *Microbiologica, 9*(4), 471-478.

Daniel, A B., Metzgar, D., Luke T. D., Margaret, A. K. R, Jianguo, W., Christopher, W., Charles, T. L., Nikki E. F., Christian J. H., Gregory C.G.,& Kevin L.R., (2004). Molecular analysis of adenovirus isolates from vaccinated and unvaccinated young adults. *J Clin Microbiol, 42(4). 1686–1693* doi: *10.1128/jcm.42.4*

de Jong, R. N., van der Vliet, P. C., & Brenkman, A. B. (2003). Adenovirus DNA replication: protein priming, jumping back and the role of the DNA binding protein DBP. *Curr Top Microbiol Immunol, 272*, 187-211.

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*(20), 12635-12642. doi: 10.1128/jvi.79.20.12635-12642.2005

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics, 5*, 113. doi: 10.1186/1471-2105-5-113

Erdman, D. D., Xu, W., Gerber, S. I., Gray, G. C., Schnurr, D., Kajon, A. E., & Anderson, L. J. (2002). Molecular epidemiology of adenovirus type 7 in the United States, 1966-2000. *Emerg Infect Dis, 8*(3), 269-277. doi: 10.3201/eid0803.010190

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) (Version 3.5c): Distributed by the author.

Flint, J., & Shenk, T. (1997). Viral transactivating proteins. *Annu Rev Genet, 31*, 177-212. doi: 10.1146/annurev.genet.31.1.177

Howitt, J., Anderson, C. W., & Freimuth, P. (2003). Adenovirus interaction with its cellular receptor CAR. *Curr Top Microbiol Immunol, 272*, 331-364.

http://biommed.lsu.edu/onlinefiles/BIOL4190/LECT-5-ADENOVIRUS/LECT-5-SLIDES-B&W-ADENOVIRUS.pdf.

http://www.vmri.hu/~harrach/AdVtaxlong.htm.).

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics, 17*(8), 754-755.

Jukes, T. H. , & Cantor, C. R. . (1969). Evolution of protein molecules.

Kajon, A. E., Mistchenko, A. S., Videla, C., Hortal, M., Wadell, G., & Avendaño, 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. doi: 10.1002/(sici)1096-9071(199602)48:2<151::aid-jmv6>3.0.co;2-8

Kaneko, H., Iida, T., Ishiko, H., Ohguchi, T., Ariga, T., Tagawa, Y., . . . Suzutani, T. (2009). Analysis of the complete genome sequence of epidemic keratoconjunctivitis-related human adenovirus type 8, 19, 37 and a novel serotype. *J Gen Virol, 90*(Pt 6), 1471-1476. doi: 10.1099/vir.0.009225-0

Kibuuka, H., Kimutai, R., Maboko, L., Sawe, F., Schunk, M. S., Kroidl, A., . . . Robb, M. L. (2010). A phase 1/2 study of a multiclade HIV-1 DNA plasmid prime and recombinant adenovirus serotype 5 boost vaccine in HIV-Uninfected East Africans (RV 172). *J Infect Dis, 201*(4), 600-607. doi: 10.1086/650299

Kidd, A. H., & Tiemessen, C. T. (1993). Characterization of a single SA7-like VA RNA gene in subgroup F adenoviruses. *J Gen Virol, 74 (Pt 8)*, 1621-1626.

Kim, Y. J., Hong, J. Y., Lee, H. J., Shin, S. H., Kim, Y. K., Inada, T., . . . Piedra, P. A. (2003). Genome type analysis of adenovirus types 3 and 7 isolated during successive outbreaks of lower respiratory tract infections in children. *J Clin Microbiol, 41*(10), 4594-4599.

Kinloch, R., Mackay, N., & Mautner, V. (1984). Adenovirus hexon. Sequence comparison of subgroup C serotypes 2 and 5. *J Biol Chem, 259*(10), 6431-6436.

Kumar, S., Tamura, K., & Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform, 5*(2), 150-163.

Lauer, K. P., Llorente, I., Blair, E., Seto, J., Krasnov, V., Purkayastha, A., . . . 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. doi: 10.1099/vir.0.80118-0

Leland, D. S., & Ginocchio, C. C. (2007). Role of cell culture for virus detection in the age of technology. *Clin Microbiol Rev, 20*(1), 49-78. doi: 10.1128/cmr.00002-06

Leppard, K. N. (1997). E4 gene function in adenovirus, adenovirus vector and adeno-associated virus infections. *J Gen Virol, 78 (Pt 9)*, 2131-2138.

Lewis, P. F., Schmidt, M. A., Lu, X., Erdman, D. D., Campbell, M., Thomas, A., . . . Gilbert, D. (2009). A community-based outbreak of severe respiratory illness caused by human adenovirus serotype 14. *J Infect Dis, 199*(10), 1427-1434. doi: 10.1086/598521

Lu, X., & Erdman, D. D. (2006). Molecular typing of human adenoviruses by PCR and sequencing of a partial region of the hexon gene. *Arch Virol, 151*(8), 1587-1602. doi: 10.1007/s00705-005-0722-7

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. doi: 10.1128/jvi.79.24.15265-15276.2005

Marc Tebruegge, & Curtis, Nigel. (2012). Adenovirus: An Overview for Pediatric Infectious Diseases Specialists Clinical Prediction Rules *The Pediatric Infectious Disease Journal, Volume 31*,(Number 6,).

Masters, J. R. (2002). HeLa cells 50 years on: the good, the bad and the ugly. *Nat Rev Cancer, 2*(4), 315-319. doi: 10.1038/nrc775

MicrobeWiki. (2013). *Adenovirus-based Gene Therapy: a Promising Novel Cancer Therapy.* from https://microbewiki.kenyon.edu/index.php/Adenovirus-based_Gene_Therapy:_a_Promising_Novel_Cancer_Therapy.

Moro, M. R., Bonville, C. A., Suryadevara, M., Cummings, E., Faddoul, D., Kobayaa, H., . . . Domachowske, J. B. (2009). Clinical features, adenovirus types, and local production of inflammatory mediators in adenovirus infections. *Pediatr Infect Dis J, 28*(5), 376-380. doi: 10.1097/INF.0b013e31819075a6

Munoz, F. M., Piedra, P. A., & Demmler, G. J. (1998). Disseminated adenovirus disease in immunocompromised and immunocompetent children. *Clin Infect Dis, 27*(5), 1194-1200.

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol, 3*(5), 418-426.

Nicholus, Karl B, & Nicholus, Hugh B Jr. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments.

O'Flanagan, D., O'Donnell, J., Domegan, L., Fitzpatrick, F., Connell, J., Coughlan, S., . . . Carr, M. J. (2011). First reported cases of human adenovirus serotype 14p1 infection, Ireland, October 2009 to July 2010. *Euro Surveill, 16*(8).

Potter, R. N., Cantrell, J. A., Mallak, C. T., & Gaydos, J. C. (2012). Adenovirus-associated deaths in US military during postvaccination period, 1999-2010. *Emerg Infect Dis, 18*(3), 507-509. doi: 10.3201/eid1803.111238

Robinson, C. M., Seto, D., Jones, M. S., Dyer, D. W., & Chodosh, J. (2011). Molecular evolution of human species D adenoviruses. *Infect Genet Evol, 11*(6), 1208-1217. doi: 10.1016/j.meegid.2011.04.031

Robinson, C. M., Singh, G., Lee, J. Y., Dehghan, S., Rajaiya, J., Liu, E. B., . . . Chodosh, J. (2013). Molecular evolution of human adenoviruses. *Sci Rep, 3*, 1812. doi: 10.1038/srep01812

Rowe, W. P. (1955). Studies of the adenoidal-pharyngeal-conjunctival (APC) group of viruses. *Am. J. Hyg*.

Russell, W. C. (2009). Adenoviruses: update on structure and function. *J Gen Virol, 90*(Pt 1), 1-20. doi: 10.1099/vir.0.003087-0

Rux, J. J., Kuser, P. R., & Burnett, R. M. (2003). Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution x-ray crystallographic, molecular modeling, and sequence-based methods. *J Virol, 77*(17), 9553-9566.

Sarantis, H., Johnson, G., Brown, M., Petric, M., & Tellier, R. (2004). Comprehensive detection and serotyping of human adenoviruses by PCR and sequencing. *J Clin Microbiol, 42*(9), 3963-3969. doi: 10.1128/jcm.42.9.3963-3969.2004

Sequencing using capillary electrophoresis on the ABI Prism 3700 [Internet]. CBS Fungal Biodiversity Centre An institute of the Royal Netherlands Academy of Arts and Sciences; [cited 2014 Nov 14]. Available from: http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&sqi=2&ved=0CCsQFjAC&url=http%3A%2F%2Fwww.scedosporium-ecmm.com%2FProtocols%2FSequencing%2520using%2520capillary%2520electrophoresis%2520on%2520the%2520ABI%2520Prism%25203700.pdf&ei=BAVmVKGkOMvwaO-hgLAO&usg=AFQjCNEA6cYtIgMPnZbjbZK6R6Fb_U6aRA&sig2=QRhembl2mlJeS2ug35AHrA&bvm=bv.79400599,d.d2s

Scherer, W. F., Syverton, J. T., & Gey, G. O. (1953). Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. *J Exp Med, 97*(5), 695-710.

Sharma, A., Li, X., Bangari, D. S., & Mittal, S. K. (2009). Adenovirus receptors and their implications in gene delivery. *Virus Res, 143*(2), 184-194. doi: 10.1016/j.virusres.2009.02.010

Shenk, T. E (2001.). Adenoviridae: the viruses and their replication. *4*, 2265– 2300.

Singh-Naz, N., Brown, M., & Ganeshananthan, M. (1993). Nosocomial adenovirus infection: molecular epidemiology of an outbreak. *Pediatr Infect Dis J, 12*(11), 922-925.

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, Peterson D, Peterson N, Stercher G, Nei M, & S, Kumar. (2011). MEGA: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. . *Molecular Biology and Evolution 28:2731-2739*.

ThermoFisherScientificTechnicalsupport. Assessment of Nucleic Acid Purity. *T042 TECHNICAL BULLETIN NanoDrop Spectrophotometers*. www.nanodrop.com

Tavaré, S. 1986. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *Lectures on Mathematics in the Life Sciences.* American Mathematical Society.

Torres, S., Chodosh, J., Seto, D., & Jones, M. S. (2010). The revolution in viral genomics as exemplified by the bioinformatic analysis of human adenoviruses. *Viruses, 2*(7), 1367-1381. doi: 10.3390/v2071367

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*(3), 991-993. doi: 10.1128/jcm.01694-09

Wickham, T. J., Mathias, P., Cheresh, D. A., & Nemerow, G. R. (1993). Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell, 73*(2), 309-319.

Wold, W. S., & Gooding, L. R. (1991). Region E3 of adenovirus: a cassette of genes involved in host immunosurveillance and virus-cell interactions. *Virology, 184*(1), 1-8.

Wu, E., Martinez, V., Alvarez, A. M., Larranaga, C., & Vela, H. (1990). [Fatal cases of adenovirus infection]. *Rev Chil Pediatr, 61*(4), 177-184.

Xu, W., McDonough, M. C., & Erdman, D. D. (2000). Species-specific identification of human adenoviruses by a multiplex PCR assay. *J Clin Microbiol, 38*(11), 4114-4120.

Yew, P. R., Liu, X., & Berk, A. J. (1994). Adenovirus E1B oncoprotein tethers a transcriptional repression domain to p53. *Genes Dev, 8*(2), 190-202.

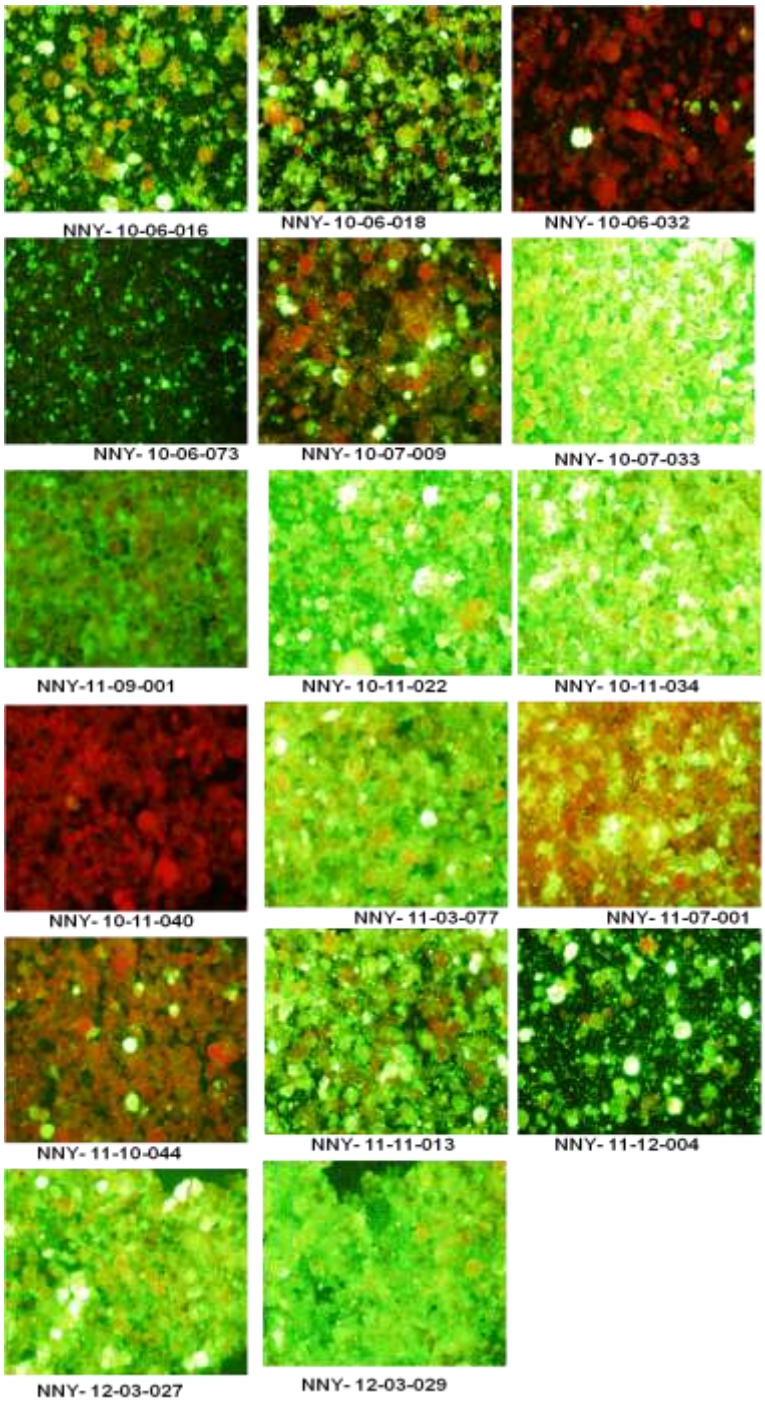
Young, C. S. (2003). The structure and function of the adenovirus major late promoter. *Curr Top Microbiol Immunol, 272*, 213-249.

Zhang, W., & Imperiale, M. J. (2003). Requirement of the adenovirus IVa2 protein for virus assembly. *J Virol, 77*(6), 3586-3594.

APPENDICES

APPENDIX I

Immunofluorescence detection of adenovirus infected Hep-2 cells under UV light



APPENDIX II

Letter of authorization to carry out the study by KEMRI Scientific Steering Committee

KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840-00200, NAIROBI, Kenya
Tel (254) (020) 2722541, 2713349, 0722-205901, 0733-400003; Fax: (254) (020) 2720030
E-mail: director@kemri.org info@kemri.org Website:www.kemri.org

ESACIPAC/SSC/101342 16th January, 2013

Fredrick Mbui

Thro'
Director, CVR
NAIROBI

FOR DIRECTOR
CENTRE FOR VIRUS RESEARCH
P. O. Box 54628
NAIROBI

REF: SSC No. 2499 (Revised) – Genetic analysis of respiratory adenoviruses among peadiatric patients attending New Nyanza Provincial General Hospital in the period from June 2010 to June 2012

I am pleased to inform you that the above mentioned proposal, in which you are the PI, was discussed by the KEMRI Scientific Steering Committee (SSC), during its 198th meeting held on 15th January, 2013 and has since been approved for implementation by the SSC.

Kindly submit 4 copies of the revised protocol to SSC within 2 weeks from the date of this letter i.e, 30th January 2013.

We advise that work on this project can only start when ERC approval is received.

Sammy Njenga, PhD
SECRETARY, SSC

CENTRE FOR VIRUS RESEARCH
18 JAN 2013
P.O. Box 54628 - NAIROBI

APPENDIX III

Letter of authorization to carry out the study by KEMRI Ethical review Committee

KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840-00200, NAIROBI, Kenya
Tel (254) (020) 2722541, 2713349, 0722-205901, 0733-400003; Fax: (254) (020) 2720030
E-mail: director@kemri.org info@kemri.org Website:www.kemri.org

KEMRI/RES/7/3/1 **April 22, 2013**

TO: **FREDRICK MBUI (PRINCIPAL INVESTIGATOR)**

THROUGH : **DR. FRED OKOTH; DIRECTOR, CVR** forwarded April 26th 2013

RE: **SSC PROTOCOL NO. 2499 – REVISED (*RE-SUBMISSION*): GENETIC ANALYSIS OF RESPIRATORY ADENOVIRUSES AMONG PAEDIATRICS PATIENTS ATTENDING NEW NYANZA PROVINCIAL GENERAL HOSPITAL IN THE PERIOD JUNE 2010 TO JUNE 2012**

Make reference to your letter dated 4th April, 2013, Received on April 5, 2013.

We acknowledge receipt of the Revised Study Protocol version 1.2 Dated 3rd April 2013.

This is to inform you that the Ethics Review Committee (ERC) reviewed the document listed above and is satisfied that the issues raised at the 213th meeting held on 19th March, 2013 have been adequately addressed.

The study is granted approval for implementation effective this **22nd day of April 2013**. Please note that authorization to conduct this study will automatically expire on **April 21, 2014**. If you plan to continue with data collection or analysis beyond this date, please submit an application for continuing approval to the ERC Secretariat by **March 18, 2014**.

Any unanticipated problems resulting from the implementation of this protocol should be brought to the attention of the ERC. You are also required to submit any proposed changes to this protocol to the ERC to initiation and advise the ERC when the study is completed or discontinued.

You may embark on the study.

Sincerely,

Dr. Elizabeth Bukusi,
ACTING SECRETARY,
KEMRI/ETHICS REVIEW COMMITTEE

CENTRE FOR VIRUS RESEARCH 26 APR 2013 P.O. Box 54628 - NAIROBI