# Molecular Characterization of the Haemagglutinin-Neuraminidase Gene of Human Parainfluenza Virus Type-1 Isolated from Infants Aged 6-36 Months attending Mbagathi District Hospital Nairobi Kenya 2006-2010

**Joshua Kimutai Kiptinness**

**A thesis submitted in partial fulfillment for the Degree of Master of Science in Molecular Medicine in the Jomo Kenyatta University of Agriculture and Technology**

**2013**

## DECLARATION

This is my original work and has not been presented for a degree in any other University.

Signature:…… Date:……..

**Joshua Kimutai Kiptinness**

This thesis has been submitted for examination with our approval as University Supervisors:

1. Signature:…… Date:……..

**Dr. Wallace Bulimo**

**KEMRI, Kenya**

2. Signature:…… Date:……..

**Dr. Fred Wamunyokoli**

**JKUAT, Kenya**

## DEDICATION

This work is dedicated to my Parents who believed I can do anything I put my mind to.

To my loving wife Evonne Kimutai, I wouldn’t have done this without your support each step of the way.

## ACKNOWLEDGEMENTS

I thank my supervisors, Dr. Wallace Bulimo of National Influenza Centre, Kenya Medical Research Institute, and Dr. Fred Wamunyokoli of Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology for their direction and supervision throughout the course of my project and thesis writing. I am forever grateful.

I appreciate the Walter Reed Project (WRP) for their unwavering financial and material support in form of monthly stipends and project materials that I received from the start of my project to the successful completion.

The entire Flu team who were ready to answer my endless barrage of questions and went over and beyond my expectations to ensure that I had everything I needed, many thanks. Specifically, I'd like to acknowledge the timely manner in which Rachel Achilla ordered my primers and to the rest of the team, Janet Majanja, Meshack Wadegu, Silvanus Mukunzi, Finnley Osuna, Benjamin Opot, Sam Symekher you've been a great help indeed!

To the source of all knowledge, the almighty God, for keeping me in good health and strength throughout the project.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF PLATES

## LIST OF APPENDICES

## LIST OF ABBREVIATIONS AND ACRONYMS

| | | |
|---|---|---|
| **%** | : | percent |
| **µg/ml** | : | micro grams per milliliter |
| **°C** | : | degrees celsius |
| **ARI** | : | Acute respiratory infection |
| **ATCC** | : | American Type Culture Collection |
| **bp** | : | base pairs |
| **cm** | : | centimeters |
| **$cm^2$** | : | square centimeters |
| **CPE** | : | cytopathic effect |
| $C_T$ | : | Threshold cycle |
| **DMEM** | : | Dulbecco's Modified Eagle Media |
| **DNA** | : | Deoxyribonucleic acid |
| **dNTPs** | : | deoxynucleotide triphosphates |
| **EDTA** | : | Ethylene Diamine Tetra Acetic acid |
| **ENC** | : | Effective number of codons |
| ***et al.*** | : | and others |
| **FDA** | : | Food and Drug Administration |
| **Fig.** | : | Figure |
| **HN** | : | Haemagglutinin Neuraminidase |
| **HPIV** | : | Human Parainfluenza Virus |
| **HPIV-1** | : | Human Parainfluenza Virus type 1 |

| | | |
|---|---|---|
| **HPIVs** | : | Human Parainfluenza viruses |
| **i.e.** | : | that is |
| **KEMRI** | : | Kenya Medical Research Institute |
| **LLC-MK2** | : | Monkey Kidney cells |
| **LRI** | : | Lower Respiratory Infection |
| **MAb** | : | Monoclonal Antibody |
| **MBG** | : | Mbagathi |
| **$MgCl_2$** | : | Magnesium Chloride |
| **ml** | : | milli liter |
| **mM** | : | milli Molar |
| **ng** | : | nanogram |
| **nm** | : | nanometer |
| **no.** | : | number |
| **NAAT** | : | Nucleic Acid Amplification Test |
| **ORF** | : | Open Reading Frame |
| **PBS** | : | Phosphate Buffered Saline |
| **PCR** | : | Polymerase Chain Reaction |
| **pH** | : | Acidity or alkalinity |
| **pm** | : | picomoles |
| **RNA** | : | Ribonucleic acid |
| **rpm** | : | revolutions per minute |
| **RSV** | : | Respiratory Syncitial Virus |

| | | |
|---|---|---|
| **RT-PCR** | : | Reverse Transcription Polymerase Chain Reaction |
| **qRT-PCR** | : | Quantitative Reverse Transcription Polymerase Chain reaction |
| **SeV** | : | Sendai Virus |
| **SSC** | : | Scientific Steering Committee |
| **TBE** | : | Tris Borate EDTA |
| **ti/tv** | : | transition transversion ratio |
| **ut** | : | undetermined |
| **µg** | : | micrograms |
| **µg/ml** | : | micrograms per milliliter |
| **µl** | : | microliter |
| **µM** | : | micromolar |

## ABSTRACT

Human parainfluenza virus type 1 (HPIV-1), a Paramyxovirus, is a leading cause of paediatric respiratory hospitalizations globally. Currently, there is no clinically successful vaccine against HPIV-1. Hence, there is need to characterize circulating strains of this virus in order to establish and develop a feasible and efficacious vaccine against the virus. The variable HPIV-1 Hemagglutin-Neuraminidase (HN) protein is found in the envelope of HPIV-1 where it initiates the infection process by binding to cellular receptors. HN is also the major antigen against which the human immune response is directed against. This study focused on identifying mutations in the HN gene that would be useful in understanding evolution of HPIV-1. Twenty five HPIV-1 isolates were obtained after screening nasopharyngeal samples from patients with influenza-like-illness attending Mbagathi District Hospital in Nairobi from June 2006 to December 2010. RT-PCR was carried out on the isolates using HN-specific primers to amplify 360nt in the most polymorphic region and the amplicons sequenced. Genomic data was analysed using a suite of bioinformatics software. Forty-eight polymorphic sites with a total of 55 mutations were identified at the nucleotide level and 47 mutations at 23 positions at the amino acid level. No positively selected sites were found in the HN protein. The data from the analysis of 21 isolates suggests that the HN gene which is the major antigenic target is under purifying (negative) selection.

# CHAPTER ONE

## 1.0 INTRODUCTION

### 1.1 Human parainfluenza viruses

Human parainfluenza viruses (HPIVs) are enveloped, non-segmented, single-stranded, negative-sense RNA viruses belonging to the family *Paramyxoviridae*. HPIV is genetically divided into 4 groups, namely 1 to 4. This group of viruses includes HPIV serotype 1, 2, and 3, which collectively are the second leading cause of pediatric respiratory hospitalizations following respiratory syncytial virus (RSV) infections globally (Karron and Collins, 2007). HPIV-1 and HPIV-3 are major causes of lower respiratory infections in infants, young children, the immunocompromised, the chronically ill, and the elderly (Mounts *et al.*, 2001). The hemagglutinin neuraminidase (HN) glycoprotein of human parainfluenza virus type 1 (HPIV-1) mediates attachment to the host cell and is the target of protective antibody (Anne, 2005). Since the efficacy of a potential vaccine depends on antigenic constancy (Emmalene *et al.*, 2007), determining the rate of evolution and HN homogeneity may potentially allow development of a single vaccine formulation for the prevention of disease.

### 1.2 Background information on HPIV-1

HPIV-1 is responsible for approximately 6% of pediatric respiratory tract disease (Karron and Collins 2007). HPIV-1 has worldwide distribution and probably

contributes significantly to childhood mortality in the developing world (Henrickson 1994). In Kenya, Acute Respiratory infection (ARI) accounts for 20% of hospital admissions and more than half of these are children under 5 years of age (Ochieng *et al.,* 1988).

### 1.3 Problem Statement

HPIV-1 is a significant respiratory pathogen for infants and young children, with clinical manifestations ranging from mild disease, including rhinitis and pharyngitis, to more-severe disease, including croup, bronchiolitis, and pneumonia (Karron and Collins, 2007). The contribution of HPIV-1 infections to pediatric respiratory hospitalizations varies between studies and ranges from 7% to 21% and is the second leading cause of pediatric hospitalizations for viral respiratory disease, behind RSV and ahead of influenza (Iwane *et al*., 2004). Different vaccine strategies have included the use of killed viruses (Heilman 1990), attenuated viruses (Skiadopoulos *et al.,* 2002), virus subunits (Brideau *et al.,* 1993), virus recombinants (Hu *et al.,* 1992), and Jennerian vaccines (Murphy, 1988) with clinical trials initiated with the several different vaccine candidates (Skiadopoulos *et al.,* 2002; Tang *et al.,* 2008; Durbin and Karron, 2003), each demonstrating some promise, but no vaccines have yet been clinically proven. Limited genomic data and poorly understood/studied HPIV-1 evolution has also been an impediment to the development of molecular diagnostics for HPIV-1.

## 1.4 Justification

The genetic stability of HPIV-1 HN gene in Kenya is unknown. Establishing the genetic stability of HPIV-1 HN gene will provide an increased understanding of the molecular basis for HN mutations that may eventually lead to vaccine formulation that can be designed to be both attenuated and immunogenic. This outcome would offer great benefit to the pediatric population in Kenya.

## 1.5 Objectives

### 1.5.1 General objective

To characterize the Haemagglutinin-Neuraminidase gene of HPIV-1 isolates obtained from infants aged 6-36 months attending Mbagathi District Hospital, Nairobi, Kenya, 2006-2010.

### 1.5.2 Specific objectives

i. To confirm the presence of HPIV-1 from the archived isolates.
ii. To amplify the Haemagglutinin-Neuraminidase gene from the positively identified HPIV-1 isolates.
iii. To perform sequence analysis of the Haemagglutinin-Neuraminidase amplified gene segment.
iv. To carry out phylogenetic analysis of the sequenced Haemagglutinin-Neuraminidase gene of HPIV-1 isolates.

# CHAPTER TWO

## 2.0 LITERATURE REVIEW

### 2.1 Classification and nomenclature of Human Parainfluenza Viruses

HPIV is genetically and antigenically divided into types 1 to 4. HPIV-1 is a major cause of lower respiratory infections in infants, young children, the immunocompromised, the chronically ill, and the elderly (Glezen *et al.*, 2000). These medium-sized viruses are enveloped, and their genomes are organized on a single negative-sense strand of Ribonucleic acid (RNA). The majority of their structural and biological characteristics are similar, but they each have adapted to infect humans at different ages and cause different diseases. These viruses belong to the *Paramyxoviridae* family, which is a large rapidly growing group of viruses that cause significant human and veterinary diseases (Henrickson, 2003).

### 2.2 Epidemiology

Biennial epidemics are the hallmark of HPIV-1 and occur in both hemispheres (Carballal *et al.*, 2001). The majority of infections occur in children aged 6 to 36 months, with a peak incidence in the second and third year of life. HPIV-1 can cause Lower Respiratory Infections (LRI) in young infants but is rare in those younger than 1 month. HPIV-1 and HPIV-3 have all been found to occur at low levels in most months of the year, similar to Respiratory Syncitial Virus (RSV) and influenza virus (Henrickson, 1998).

Studies in the Americas and Europe show that, on average, young children (under 5 years of age) suffer 3 to 6 episodes of Acute Respiratory Infections (ARI) per year and that one third to one half of the out-patient pediatric consultations in developing countries are due to ARI (WHO, 1981).

In Kenya, (ARI) accounts for 20% of hospital admissions and more than half of these are children under 5 years of age. ARI accounts for 25% of all deaths amongst these admissions, with 87% being in children less than 5 years old. During 1981 in the Paediatric Observation Wards of Kenyatta National Hospital, 41% of admissions and 22% of deaths were due to ARI (Ochieng *et al.*, 1988).

### 2.3 Structural organization of HPIV-1

The HPIV-1 genome contains approximately 15,000 nucleotides (Storey *et al.*, 1984; Washburne *et al.*, 1992), which encode at least six common structural proteins (3'-NP-P-M-F-HN-L-5') as shown in figure 2.1. The characteristics of the structural proteins described are shown in table 2.1.

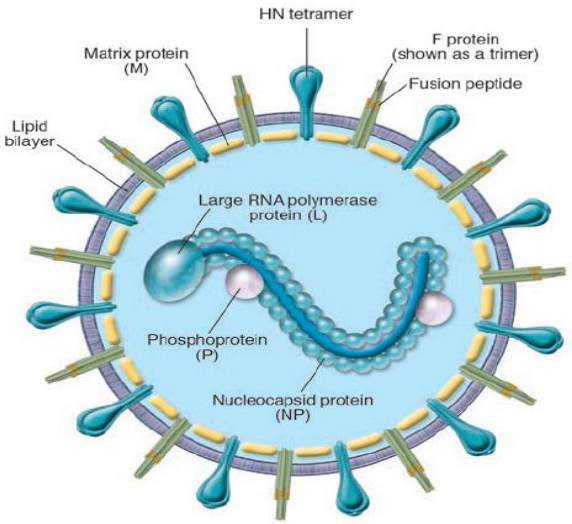



**Figure 2.1:** A schematic diagram of the HPIV-1 virion.
L, large RNA polymerase protein; M, matrix protein; NP, nucleocapsid protein; P, phosphoprotein; HN tetramer; F protein, (Anne, 2005).

**Table 2.1**: Characteristics of HPIV-1 proteins, (Anne, 2005).

| PROTEIN | MOLECULAR MASS Kilodaltons | FUNCTIONS |
|---|---|---|
| **Structural Protein** **Surface** Fusion (F) | 60 | Penetration; major protection antigen |
| Haemagglutinin Neuraminidase (HN) | 69 | Viral attachment and release; Major protection antigen |
| **Matrix** Matrix (M) | 40 | Mediates attachment of Nucleocapsid to envelope |
| **Nucleocapsid associated** Nucleoprotein (N, NP) | 58 | Major RNA-binding nucleocapsid protein |
| Phosphoprotein (P) | 60 | Major phosphorylated protein; RNA-dependent RNA polymerase activity |
| Large polymerase complex (L) | 250 | Large nucleocapsid-associated protein ; major polymerase subunit; RNA-dependent RNA polymerase activity |

## 2.4 HPIV infection cycle

The first step in infection of a cell by all HPIVs is binding to the target cell, via interaction of the viral receptor-binding molecule haemagglutinin-neuraminidase (HN) with sialic acid-containing receptor molecules on the cell surface (Figure 2.2). The viral envelope then is thought to fuse directly with the plasma membrane of the cell, mediated by the viral fusion protein (F protein), releasing the nucleocapsid into the cytoplasm (Plumper *et al.,* 2003). The nucleocapsid released into the cytoplasm after fusion contains the genome ribonucleic acid (RNA) in

tight association with the viral nucleocapsid protein, and this RNA/protein complex is the template both for transcription and for replication of the genome RNA that is packaged into progeny virions.

Virions are formed according to the prevailing model for virion assembly, when newly assembled nucleocapsids containing the full-length viral RNA genome along with the polymerase proteins bud out through areas of the plasma membrane that contain the F and HN proteins and the matrix protein. In polarized epithelial cells, the viruses bud from the apical surface of the cell. The matrix protein binds to the nucleocapsid and also interacts with the cytoplasmic tails of the HN and F proteins, in this way mediating the alignment of the nucleocapsid with the areas of the plasma membrane containing viral glycoproteins in order to set the scenario for budding (Ali and Nayak, 2000). The neuraminidase or receptor-cleaving activity of the HN molecule cleaves sialic acid–containing receptor moieties that would attach the viral HN protein to the cell surface and allows the release of newly budded particles from the cell to begin a new round of infection (Porotto *et al.*, 2001).

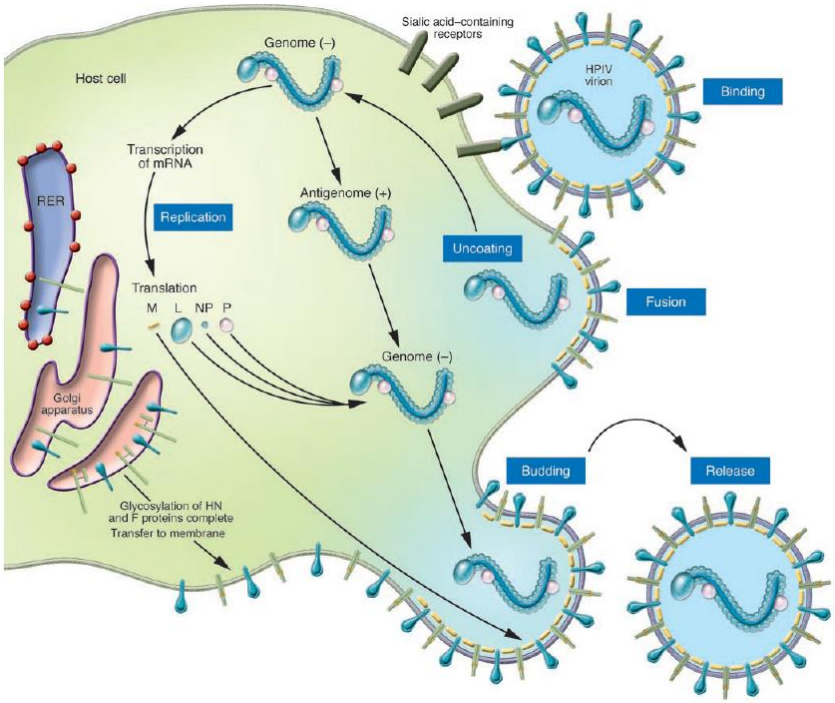



**Figure 2.2:** A schematic illustration of the HPIV infection cycle (Anne, 2005).

**2.5 HPIV recovery**

The nasopharynx and oropharynx are primary locations of initial HPIV replication. Virus is shed at high titer early in an infection and then sharply declines. Experimental infections in adults have demonstrated that illness usually starts about 3 to 4 days after inoculation and lasts from 3 to 17 days, with an average of 4 days for HPIV-1 (Tyrrell and Bynoe, 1969). Children shed virus usually from 3 to 4 days prior to the onset of clinical symptoms until approximately 10 days after infection (Frank *et al.*, 1981). Immunocompromised patients and adults (mostly with chronic lung disease) have persistently shed HPIV-1, HPIV-2, and HPIV-3 for many months due to the several-log-unit increase in viral titer (Gross *et al.*,

1973). Throat swabs, nasopharyngeal swabs, nasal washes, and nasal aspiration have all been used successfully to recover HPIV (Frayha *et al.,* 1989).

### 2.6 HPIV proteins and their Antigenic Structure

The HN protein is found on the lipid envelope of HPIV and infected cells (Henrickson and Savatski, 1996). There, it most probably exists as a tetramer and functions in virus-host cell attachment via sialic acid receptors, suggesting that it has neuraminidase activity**.**

The HN protein has hemagglutinating and neuraminidase activities. Its putative three-dimensional structure is thought to be broadly similar to that of influenza virus neuraminidase, but the framework residues of the active site are less strictly conserved (Colman *et al.,* 1993). The function of the haemagglutinin is to recognize and bind to the cell receptor, a glycoconjugate on which sialic acid is the terminal sugar (Colman *et al.,* 1983). The neuraminidase is responsible for removing sialic acid (e.g., *N*-acetylneuraminic acid) residues from the newly synthesized haemagglutinins and neuraminidases which are sialylated by the cellular enzymes (Griffin *et al.,* 1983). In the absence of functional neuraminidase, virus release is inhibited; virions are formed but remain attached to the cell surface and to each other, forming large aggregates on the surfaces of infected cells (Palese *et al.,* 1974). Since the neuraminidase (NA) is an essential gene product for virus propagation and possesses an enzyme active site whose amino acid sequence is conserved among types and subtypes of influenza viruses, it has been considered

a suitable target for antiviral design (Von Itzstein *et al.,* 1993). The NA is a receptor-destroying enzyme that catalyzes the hydrolytic cleavage of the α-ketosidic bond linking a terminal sialic acid to the adjacent carbohydrate moiety (Paulson *et al.,* 1982).

### 2.7 Similarities and differences of HPIV to other paramyxoviruses

Sendai virus (SeV), Newcastle disease virus (NDV), HPIV and several other enveloped animal and human viruses of the *Paramyxoviridae* family share a common receptor-ligand interaction and their mode of entry into host cells. The very first step for introgression of their RNA genome into host cells is membrane fusion. The fusion requires a coordinated action of two envelope glycoproteins: a receptor-binding protein, haemagglutinin-neuraminidase (HN), and a fusion protein (F), (Russel and Luque, 2006).

### 2.8 Evasion of the Immune system by HPIV

There are significant differences in the number of HN glycosylation sites between HPIV types and among strains within one type (Henrickson and Savatski, 1992; Henrickson, 1991). This may be part of the strategy used by HPIV to escape immune detection. It is the binding of the HN protein to receptors on guinea pig erythrocytes (Chanock *et al.,* 1961) that creates the well-recognized hemagglutination or hemadsorption of paramyxoviruses. However, in HPIV-1, HN protein genetic variation and evolution occur (Henrickson and Savatski 1992).

## 2.9 Clinical syndromes

Clinical manifestations range from mild disease, including rhinitis, pharyngitis, and ortitis media, to more severe disease, bronchiolitis, and pneumonia (Counihan, *et al.*, 2001; Karron and Collins 2007). Also, HPIV-1 and other respiratory viruses have been shown to have procoagulant activity that may play a role in cardiovascular disease (Visseren *et al.*, 2000).

## 2.10 Pathogenesis

Upon HPIV infection of a cell, the first observable morphologic changes of the cell include focal rounding and increase in size of the cytoplasm and nucleus. Generally, HPIV decreases host cell mitotic activity 24 hours after inoculation (Konovalova *et al.*, 1967).

Paramyxoviruses are known to induce apoptosis in tissue culture cells (He *et al.*, 2001). Both the HN and F proteins have been shown to be involved with cell membrane fusion, but this appears to be virus specific, with some paramyxoviruses needing only the F protein (Hu *et al.*, 1992).

Neuraminidase activity of the HN protein may play a role in altering muscarinic receptors, leading to vagus-induced bronchospasm, and can cause HPIV-1 infected cells to resist infection when challenged with HPIV-1 (Fryer and Jacoby, 1991).

## 2.11 Molecular approaches to HPIV-1 diagnosis

Accurate detection of respiratory viruses is important to guide antiviral therapy, prevent nosocomial spread, provide surveillance, and in some cases, decrease

hospital costs and lengths of stay (Barenfanger *et al.,* 2000; Henrickson 2005). HPIV-1 infections have historically been diagnosed by virus isolation, by detection of viral antigen or RNA by immunofluoresnce assay (IFA) and nucleic acid amplification tests (NAAT), respectively, or by serological tests such as haemagglutination inhibition (HAI) tests (Leland and Ginocchio, 2007).

### 2.12 Growth in Cell Culture

HPIV-1 is isolated more easily in epithelial cell lines than in fibroblast cell lines. Many commonly used cell lines support the growth of HPIV-1, including primary monkey kidney cells , (LLC-MK2) and Buffalo green monkey kidney (BGMK (Cuevas *et al.,* 2003). LLCMK2 line was first derived from a pooled cell suspension prepared from kidneys removed from six adult Rhesus monkeys (Hull *et al.,* 1956). In culture, these cells form a contiguous monolayer. These cells are susceptible to a broad range of viruses including human parainfluenza virus type 1 (Hull *et al.,* 1962). The addition of an exogenous protease like trypsin to the cell culture medium facilitates virus recovery from nasopharyngeal aspirates for HPIV-1 (Henrickson *et al.,* 1994).

### 2.13 Immunofluorescent assays

Viral antigens are routinely detected in nasopharyngeal epithelial cells in many laboratories by Immunofluorescence assay (IFA) using a panel of monoclonal antibodies (Landry and Ferguson, 2000). The sensitivity of IFA compared to cell culture varies between laboratories depending on the reagents used and has ranged

from 70% to 83% (Krunic *et al.,* 2007; Landry, 2007). Monoclonal antibody pools that use two fluorescent dyes can be used to detect HPIV types 1, 2, and 3 (SimulFluor reagents; Chemicon International, Temecula, CA). These reagents are Food and Drug Administration (FDA) approved for direct specimen testing and culture confirmation. SimulFluor reagents have excellent sensitivity and specificity compared with individual antibodies (Landry and Ferguson, 2000).

Currently, IFA is the "gold standard" for virological diagnosis, however molecular methods result in more sensitive, specific, and rapid detection of respiratory viruses (van Elden *et al.,* 2002).

Although IFA has lower sensitivity than PCR, this may be an advantage for the detection of clinically relevant infections (Madeley and Peiris, 2002; Schindera *et al.,* 2010)

### 2.14 Polymerase Chain Reaction (PCR) based detection of HPIV

A variety of Nucleic Acid Amplification Test (NAATs) for detecting HPIV have been described, and most have shown increased sensitivity compared to that of culture. Fan and Hendrickson, 1996, described the first quantitative reverse transcription PCR (qRT-PCR), for HPIV that had an analytical sensitivity of 600 copies/ ml and was more sensitive than culture.

Several studies have shown that PCR methods appear to be more sensitive than IFA and culture for the diagnosis of acute respiratory virus infections (Templeton *et al.,* 2004; van Kraaij *et al.,* 2005 ).

A significant proportion of asymptomatic children test positive by PCR to respiratory viruses (van der Zalm *et al.,* 2009; van Gageldonk-Lafeber *et al.,* 2005).

### 2.14.1 Quantitative RT-PCR

Quantitative RT-PCR (qRT-PCR) technology, which combines nucleic acid amplification with amplicon detection, provides results more quickly than conventional PCR, has in some cases shown improved sensitivity compared to conventional PCR, and provides a uniform platform for quantifying both single and multiple pathogens in a single sample (Cockerill 2003; Dagher *et al.,* 2004; ).

In a qRT-PCR assay a positive reaction is detected by accumulation of a fluorescent signal. The cycle threshold ($C_T$)is defined as the number of cycles required for the fluorescent signal to cross the threshold (i.e. exceeds background level). $C_T$ levels are inversely proportional to the amount of target nucleic acid in the sample (i.e. the lower the $C_T$ level the greater the amount of target nucleic acid in the sample).qRT-PCR assays undergo approximately 40 cycles of amplification. $C_T < 29$ are strong positive reactions indicative of abundant target nucleic acid in the sample $C_T$ of 30-37 are positive reactions indicative of moderate amounts of target nucleic acid. $C_T$ of 38-40 are weak reactions indicative of minimal amounts of target nucleic acid which could represent an infection state or environmental contamination. (Wan *et al.,* 2011).

When compared to the specific limitations of quantification by culture, qRT-PCR offers several theoretic advantages, including the following: (i) a lower threshold of detection, (ii) the potential stability of the assay after specimen freezing and thawing, thus permitting sample batching, (iii) a less subjective assay readout, and (iv) an assay that is unaffected by therapeutic passive neutralizing antibodies or experimental antiviral agents (Stephanie *et al.,* 2005).

### 2.15 Phylogenetic analysis

Phylogenetic analysis of DNA or protein sequences has become an important tool for studying the evolutionary history of organisms from bacteria to humans. Since the rate of sequence evolution varies extensively with gene or DNA segment (Dayhoff, 1972; Nei, 1987; Wilson *et al.,* 1977), one can study the evolutionary relationships of virtually all levels of classification of organisms including kingdoms, phyla, classes, families, genera, species, and intraspecific populations.

The genetic relationships between populations can be measured by determining the genetic distance between the populations. This distance measured between two populations provides a good estimate of how divergent they are genetically. Thus, when the genetic distance is large, the genetic similarity is low and the time they diverged from each other is greater while when the genetic distance is small the genetic similarity is high and the time they diverged from each other is smaller (Cavalli-Sforza and Edwards, 1967).

### 2.16 Testing neutrality of mutations

#### 2.16.1 Nei and Gojobori method

This method computes the numbers of synonymous and non-synonymous substitutions and the numbers of potentially synonymous and potentially non-synonymous sites. The difference between synonymous and non-synonymous nucleotide substitutions (p-distance) is then tested (Nei and Gojobori, 1986).

#### 2.16.2 Kumar method

Neutrality of mutations is determined 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).

#### 2.16.3 Hughes method

This method divides non-synonymous sites into conservative (Pnc) and radical (Pnr) sites and estimates the extent to which non-synonymous differences change some qualitative amino acid property of interest (charge, polarity, etc). Non-synonymous differences are categorized as conservative or radical with respect to this property (Hughes *et al.,* 1990).

#### 2.16.4 Tajima's test

In Tajima's method of testing neutrality of mutations, the pattern of nucleotide difference is examined as measured by the difference between $\pi$ (observed average

pairwise nucleotide diversity) and θ (expected nucleotide diversity under neutrality derived from the number of segregating sites (Tajima, 1993).

### 2.17 HPIV-1 Vaccine strategies

Since the 1960s, vaccine strategies have included the use of killed viruses (Heilman 1990), attenuated viruses (Skiadopoulos *et al.,* 2002), virus subunits (Brideau *et al.,* 1993), virus recombinants (Hu *et al.,* 1992), and Jennerian vaccines (Murphy, 1988). Clinical trials have been initiated with several different vaccine candidates (Skiadopoulos *et al.,* 2002; Tang *et al.,* 2008; Durbin and Karron, 2003), each demonstrating some promise, but no vaccines have yet been clinically proven.

Sendai virus (SeV) is now being considered as a novel candidate vaccine for the pediatric respiratory viruses. Since its first discovery, SeV has provided laboratories with a rich research model for the study of HPIV-1 infections in mice (Ely *et al.,* 2007; Sealy *et al.,*2003).

During laboratory studies of SeV, the profound sequence similarities between SeV and HPIV-1 were recognized (Lyn *et al.,* 1991) Antigenic similarities were also revealed when virus specific B-cells and T-cells were shown to cross-react between HPIV-1 and SeV (Smith *et al.,* 1994; Slobod *et al.,* 1994). Each of these results highlights the possibility of SeV and HPIV-1 sequence similarities as a potential new vaccine for the protection of humans from HPIV-1.

# CHAPTER THREE

## 3.0 MATERIALS AND METHODS

### 3.1 Study

#### 3.1.1 Study design

This study was a descriptive laboratory based study on archived HPIV-1 isolates.

#### 3.1.2 Study site

This study was carried out at the Kenya Medical Research Institute (KEMRI), National Influenza Center (NIC).

### 3.2 Study population

The study utilized 25 positively (all the HPIV-1 positives from 2006 to 2010) screened archived (stored at -80°C) HPIV-1 isolates collected from Mbagathi District Hospital Nairobi from the period June 2006-December 2010, which had been previously detected by cell culture and direct immunofluorescence. These samples were obtained from children of both sexes who came with their consenting guardians or parents residing in the areas adjacent to Mbagathi district hospital and attended the outpatient clinic. A prototype strain of HPIV-1 (strain C35) was obtained from the Centers for Disease Control and Prevention collections (Atlanta, USA).

The samples had unique identification numbers showing the site where the sample was collected, the year and month of collection closely followed by the sample number.

### 3.2.1 Inclusion criteria

Only isolates that were from children aged 6 to 36 months and were positive for HPIV-1 were included in this study. This information was obtained on an influenza questionnaire (appendix I).

### 3.2.2 Exclusion criteria

Isolates from children below 6 months or above 36 months and were on treatment for other diagnosed diseases were excluded.

## 3.3 Ethical Consideration

The protocol that was used in this study was submitted and approved by the scientific steering committee (SSC), appendix II, and Ethical Review Committee (ERC), appendix III, from KEMRI. The samples were previously obtained with the patients and/or the parents'/ guardian's informed consent (appendix IV).

The specimens used in this study were already used for routine surveillance of influenza at the KEMRI-NIC and stored therein (appendix V). Therefore, patients from whom the specimens were derived were not affected by the results from this study.

### 3.4 Confirmation of the identity of HPIV-1 in the isolates

HPIV-1 viruses were amplified from the 25 isolates and isolated by inoculating the isolate in passage LLC-MK2 cells (ATCC CCL7.1). Briefly, LLC-MK2, cells were grown in T75 flasks with Dulbecco's Modified Eagle Medium (DMEM) , Sigma- Aldrich Co Ltd, UK, supplemented with 10% heat inactivated fetal bovine serum (FBS) (Hull *et al.*,1962) in a $CO_2$ incubator- Thermoscientific model BBD 62220 (Thermoscientific Inc., CA, USA) at 37°C, 5% $CO_2$, 95% humidity for 72 hours. The cells in suspension were quantified by use of a hemocytometer. Briefly, 10µl of cell suspension was added to the V-groove in the hemocytometer slowly and continuously until the silver area is just covered. It was left for 10-15 minutes before being read. The 4 squares on opposing sides of the center squares on both halves of the hemocytometer was counted and averaged. The split ratio (once weekly) was 3:50 with DMEM supplemented with 10% FBS and the cell suspension aliquoted in culture tubes (NalgeNunc, Rochester UK). At 80% (approximately $8.8x10^7$ cells/ ml and $8.8x\ 10^8$ cells/ flask) confluence, the cells were washed with Hanks balanced salt solution (Invitrogen, GIBCO) supplemented with 0.25µg trypsin per ml and inoculated with 100µl of homogenized isolate after pouring off the growth medium. The adsorption of the virus onto the LLC-MK2 cell cultures was enhanced by incubating at 37°C for 1 hour in a 5% $CO_2$ incubator, Thermoscientific model BBD 62220 (Thermoscientific Inc., CA, USA).1ml of Maintenance media was added (appendix VI). The cultures were then incubated in an incubator, at 37°C, 5% $CO_2$,

95% humidity and humidity of 80% and observed for cytopathic effects after 48 hours and up to 10 days and stored at -80°C in a HERA Freeze Ultra low freezer (Thermoscientific, CA, USA). For the positive control, the prototype strain of HPIV-1 (strain C35) obtained from the Centers for Disease Control and Prevention collections, Atlanta, Georgia was used. For the negative control, 1ml of maintenance media was added to the cells.

### 3.5 HPIV-1 identification by immunofluorescence

HPIV-1 in the supernatant was identified by the direct Immunofluorescent antibody Assay (IFA) using the SimulFluor™ Flu Para-1 Monoclonal Antibody (MAb) reagent (Chemicon International, CA, USA). Briefly, culture tubes containing the isolates were centrifuged at 3000rpm for 10 minutes and the supernatant poured into cryogenic vials and stored at -70°C. The remaining cells were washed twice with 1ml sterile Phosphate Buffered Saline (Fisher Scientific, NJ, USA), in appendix VI, by vortexing and centrifuging at 3000rpm in a 5810R eppendorf centrifuge (Hamburg, Germany) and supernatant discarded. This was to remove debris. 25µl of cell suspension was coated onto a 12-well slide (Flow Laboratories, U.K) and allowed to air-dry overnight. The wells containing the cell suspension were then fixed in cold (4°C) acetone for 10 minutes and allowed to dry. The fixed cell suspension was stained with one drop of virus SimulFluor™ Flu Para-1 MAb reagent and incubated for 30 min at 37°C in a humid chamber. After staining the slides were washed in Tween 20 Phosphate Buffered Saline

(PBS) reagent for 15 seconds to remove unbound antibody solution, a drop of mounting fluid (Millipore, Light Diagnostics, USA) was then added on the well, a cover slip applied, and then viewed by a Fluorescent microscope (Olympus Optical Co. Ltd, Model BX51Japan) at X400 magnification. An apple green fluorescence in the cell nucleus or cytoplasm was observed to indicate a HPIV-1 positive sample and recorded. Positive and negative control slides were stained and included in each run for comparison. The presence of bright green fluorescence within intact cells was considered to be a positive result. The results were confirmed by an experienced technician.

### 3.6 RNA extraction

The QIAamp Viral RNA Mini spin protocol (Qiagen, Hilden, Germany) was used for RNA extraction. All isolates were extracted according to the manufacturer's instructions. Briefly, 100 µl of sample was extracted and the nucleic acids eluted in 60 µl of PCR elution buffer and stored at -70°C until analysis and were used in PCR and qRT-PCR. In the negative control, sterile distilled water was added instead of specimen. For the positive control, the prototype strain of HPIV-1 (strain C35) obtained from the Centers for Disease Control and Prevention collections, Atlanta, Georgia was used.

### 3.7 Prevention of PCR contamination

Precautions were taken to prevent cross contamination. The preparation of reagents, processing of samples, and PCR assays were carried out in separate

rooms away from the area where amplified products were analyzed. Sterile Filtered pipette tips (Molecular BioProducts®,CA, USA) were used throughout the experiments.

### 3.8 HPIV-1 Primer design

The conventional PCR primers used in this study were designed using web-primer primer design (http://www.yeastgenome.org/cgi-bin/web-primer) from HPIV-1 HN strain Washington 1964 (Accession number AF457102, region: 6847-8740) and was selected to ensure that the size of the amplicon could be easily differentiated by agarose gel electrophoresis. The primers were ordered from Bioserve, MD, USA.

#### 3.8.1 HPIV-1 qRT-PCR

Primers and probes (Templeton *et al.,* 2004) used for qRT-PCR detection were ordered from ABI (Applied Biosystems Inc., CA USA) and are listed in table 3.1.

**Table 3.1:** HN gene primers and probe used for qRT-PCR detection of HPIV-1 isolates (Henrickson *et al.*, 2000).

| Gene | Forward primer (5'-3') | Reverse primer (5'-3') | Probe |
|---|---|---|---|
| HN | TGTTTTAAACCCGGTAA TTTCTCAT | CCTTGTTCCTGCAGCTATT ACAGA | 5'FAM-ACAACAACAGGAAATC-MGB 3' |

Reverse transcription was performed using the 7500 fast Real Time RT PCR platform (Applied Biosystems Inc., CA USA). Briefly, qRT-PCR was performed

in 25µl of reaction mixture consisting of 12.5µl of 2X one-step reverse transcription (RT)-PCR buffer (Ambion, CA, USA) one-step RT-PCR kit (containing 2.5mM dNTPs and 5.5mM MgCl),0.25µl of 5pm/µl probe, 0.25µl of 10pm/µl Forward and 10pm/µl reverse primer, 5µl of 23ng/µl RNA template, 0.625µl of enzyme, 6.125µl of nuclease free water and the volume to 25µl using double distilled water. The PCR thermal profile consisted of an initial cDNA step of 30 minutes at 50°C followed by 15 minutes at 95°C and 50 cycles of 30s at 95°C, 30s at 55°C, and 30s at 72°C. In the negative control, sterile distilled water was added instead of specimen. For the positive control, the prototype strain of HPIV-1 (strain C35) was used. qRT-PCR provided information on the viral load in each of the isolates, whereby low $C_T$ values correlated with high viral loads. Primary amplification, detection, and data analysis was performed with Applied Biosystems Fast PCR 7500 platform (Applied Biosystems Inc., CA USA).

### 3.8.2 HPIV-1 PCR

All supernatants found to be positive for HPIV-1 by qRT-PCR were further tested by conventional RT-PCR to confirm the results. The primers used (Table 3.2) targeted a different region of the same gene as the real-time PCR. This was performed in 25µl of reaction mixture consisting of 12.5µl of 2X one-step reaction mixture containing 0.4mM of each dNTPs, 3.2mM $MgCl_2$ and reverse transcriptase (Invitrogen one-step RT-PCR kit), 1µl Super script III Taq polymerase, 0.5µl of the forward (20pm/µl) and reverse (20pm/µl) primers, 3µl of

30ng/ µl RNA template, 7.5µl of nuclease free water. The PCR thermal-cycling profile consisted of an initial cDNA synthesis step of 30 min at 50°C, followed by denaturation for 2 minutes at 94°C and 45 cycles (consisting of denaturation for 45s at 94°C, primer-annealing for 45s at 57°C, and strand extension for 2 minutes at 68°C). In the negative control, sterile distilled water was added instead of template RNA. For the positive control, RNA from the prototype strain of HPIV-1 (strain C35) was used. An amplicon of 360bp (Henrickson *et al.,* 2000) was detected by agarose gel electrophoresis on 2% agarose. The genetic location of the HN open reading frame (ORF) that was amplified was 7254-7521.

**Table 3.2:** Primers used for conventional PCR detection of HPIV-1 isolates (Henrickson *et al.,* 2000).

| Target gene | Forward primer (5'-3') (HPIV-1HNf) | Reverse primer (5'-3') (HPIV-1HNr) |
|---|---|---|
| HN | GCATATATTGCATCACCAATTGAT AATGAAGGTA | CCTATGTTGTTCAAGACAAG |

### 3.9 Analysis of amplicons

Amplicons were analyzed by agarose gel electrophoresis on 2% (w/v) agarose using 1XTBE (Promega, USA) as described by Sambrook *et al.,* (1989). Briefly, 2 grams of Agarose was dissolved in 100ml of 1XTBE and poured into a gel casting tray containing combs and left to set.10µl of sample was mixed with 4µl of gel

loading buffer (Promega, USA) and loaded into the wells on the gel. Electrophoresis was done at a constant voltage of 7 volts/cm using a Bio-Rad model 200/2-power supply source. Once the electrophoresis was completed, the gel was stained with 0.5μg/ml ethidium bromide. The location of the amplicon on the gels was determined by direct examination of the gel under UV light (Sharp *et al.,* 1973) using an HP Innotech Alpha-imager and the amplicon estimated by comparing with molecular weight size marker (100bp Fermentas O Gene Ruler™) loaded alongside them and band purification using GFX band purification kit (GE Healthcare, Buckinghamshire, UK).

### 3.10 Amplicon purification

This step was important to remove substances such as primers, nucleotides, salts and non-specific PCR product, which would interfere with the subsequent sequencing reactions.

Briefly, the gel slice containing the 360bp amplicon was excised using a clean scalpel and placed into a pre-weighed 1.5ml tube and weighed, and the weight recorded. A 1:1 volume of Binding buffer was added to the gel slice (volume: weight). The gel mixture was incubated at 60°C until the gel slice was completely dissolved, approximately 10 minutes. The solubilized gel solution was transferred to a GeneJet purification column and centrifuged for 1 min and the flow through discarded. 100μl of Binding buffer was added to the purification column and centrifuged for 1 minute, and flow through discarded. 700μl of wash buffer was

added to the purification column and centrifuged for 1 minute and the flow through discarded. The purification column was transferred to a clean 1.5ml microcentrifuge tube and 50µl of elution buffer added to the center of the purification column membrane and centrifuged for 1 minute. The purification column was discarded and the purified DNA was stored at -20°C.

### 3.11 Sequence analysis

Direct sequencing of the PCR product was carried out using the primers HPIV-1HNf and HPIV-1HNr (see Table 3.2) on a 3500xl genetic analyzer which uses a fluorescent based DNA analysis system. The output from the DNA sequencer was a four-colour chromatogram showing peaks that represented each of the four DNA bases. The pre-sequencing PCR used 2 µl Big Dye 3.1, 2 µl Big Dye 5X buffer, 1µl distilled water, 1µl 5pm/ul HPIV-1 HN forward and reverse primer, 2µl of approximately 50ng HPIV-1 HN DNA. Thermal profile: 95 °C for 5 minutes, 95 °C for 15s, 45 °C for 30s, 68 °C for 2min 30s, and 30 cycles at 68 °C for 3 minutes.

### 3.12 Data analysis

Sequences from the amplicons were viewed and edited using Bioedit software version 7.0.0. The edited sequences were saved as FASTA format. For each individual sequence a nucleotide-nucleotide BLAST (BLASTN) search was done (http:/ www.ncbi.nlm.nih.gov/BLAST/Blast.cgi) to confirm its identity. Once their identity was confirmed, the anti-sense strand sequence was reverse complimented

and contig assembly was done to get a consensus sequence of the sense and antisense strands using the Contig Assembly Program (CAP) available in Bioedit program. The consensus sequences were then saved in FASTA format. The saved sequences were then imported into Muscle 3.6 and multiple sequence alignment was done against the reference HPIV-1 strains obtained from National Center for Biotechnology Information (NCBI). The aligned sequences were visualized using Gene Doc version 5.1.2600.2180. The aligned nucleotide sequences (7254-7521) were used in the subsequent analysis. Polymorphic sites, neutrality test and linkage disequilibrium and recombination, was determined using DnaSP version version 5.10 (Rozas *et al.,* 2003). Difference between synonymous and nonsynonymous substitutions were tested with the method of Nei and Gjobori (Nei and Gojobori, 1986) available in the program; PAL2NAL version 14 (Mikita *et al.,* 2006) and phylogenetic trees were constructed using Mr. Bayes version 3.2 software and visualization by Fig tree version 1.3.1.

# CHAPTER FOUR

## 4.0 RESULTS

### 4.1 Viral Culture

All the 25 isolates were inoculated onto LLCMK2 cell line monolayers. The cytopathic effect (CPE) exhibited consisted of rounding, shrinking and detachment of the cells from the substratum (Plate 4.1). Amongst the cultures that exhibited presence of HPIV-1, majority of the cultures (52%) showed CPE after incubation for between 7-10 days while only 8% showed CPE before day 7. 40% of the isolates inoculated with HPIV-1 virus stocks did not show any CPE (Appendix VII). However, further analyses of these HPIV-1 negative cultures using IFA showed that some of the cells in the cultures stained green with anti-HPIV-1 monoclonal antibodies, indicative of presence of HPIV-1 (Plate 4.2B; see section 4.2 below). For those that did not light up on IFA assays, a further two passages in cell cultures was carried out in order to definitively confirm the absence of HPIV-1 (Plate 4.2A).

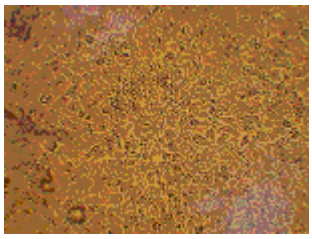
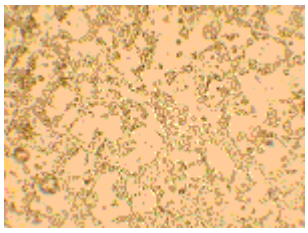
**A**  **B** 

**Plate 4.1:** Cell morphology of cultured LLCMK2 monolayers.
Plate **A** shows monolayer of uninfected cells with no CPE. Plate **B** shows a monolayer of the same cells after infection with HPIV-1 exhibiting rounding, shrinkage and detachment of the cells from the substratum, which are the morphological characteristics due to CPE. Magnification, x200.

### 4.2 Immunofluorescence assay

Presence of CPE on a monolayer of cultured LLCMK2 cells is not an indication of the presence of HPIV-1 *per se*. Thus, to determine whether the CPE observed was due to infection by HPIV-1, a specific technique, the immunofluorescence assay (IFA), using anti-HPIV-1 monoclonal antibodies, which detects antigens present on the surface of the virus within the cell, was carried out. Furthermore, since observation of CPE is subjective, cultures that may seem CPE-negative may in fact have virus in them. Thus IFA was also carried out on CPE-negative cultures to determine true and false negative HPIV-1 cultures.

Of the 25 cultures screened for HPIV-1 by IFA, 21 tested positive as shown by the apple-green fluorescence among infected cells (Plate 4.2B). Negative samples showed a red color of uninfected cells (Plate 4.2A). As expected the positive control of the proto type C35 strain showed apple green fluorescence. Of these 21 IFA-positive cultures, six did not show any CPE's in cell culture.

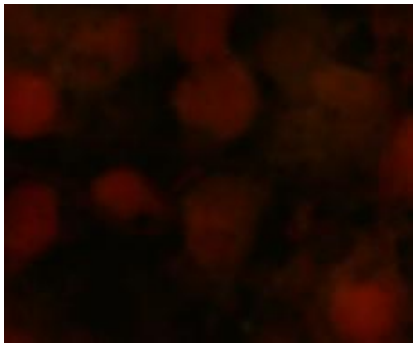
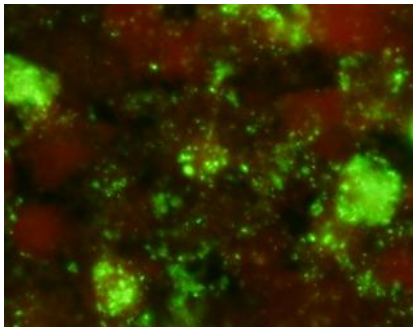
**A**  **B** 

**Plate 4.2:** HPIV-1 infection of LLCMK2 cells assayed by IFA.
Plate **A** shows uninfected cells in the presence of anti-HPIV-1 antibodies. These uninfected cells appear with a background dull red color. Plate **B** shows infected cells probed with anti-HPIV-1 antibodies. Infected cells here fluoresce with the apple green color. Magnification, x400.

## 4.3 PCR

### 4.3.1 Quantitative RT-PCR

Fifteen isolates (MBG/07/06/01, MBG/07/07/02, MBG/07/08/03, MBG/08/04/09, MBG/08/06/10, MBG/08/10/12, MBG/08/10/13, MBG/08/10/14, MBG/08/10/15, MBG/08/11/16, MBG/09/09/17, MBG/10/01/19, MBG/10/02/20, MBG/10/05/22 and MBG/10/05/23) which showed CPE upon inoculation in LLCMK2 cells and were positive on IFA displayed low ($C_T$) values ranging from 22 to 28. However, isolates that were CPE-negative but positive with in IFA and displayed higher ($C_T$) values ranging from 34 to 40 (Table 4.1).

**Table 4.1:** Usage of qRT-PCR, CPE and IFA diagnostic assays for the detection of HPIV-1 in culture.

| **Isolate name** | **$C_T$ Value** | **Presence of CPE?** | **IFA** |
|---|---|---|---|
| MBG/07/06/01 | 27 | YES | + |
| MBG/07/07/02 | 24 | YES | + |
| MBG/07/08/03 | 26 | YES | + |
| MBG/08/03/04 | UNDET. | NO | - |
| MBG/08/04/05 | 38 | NO | + |
| MBG/08/04/06 | 40 | NO | + |
| MBG/08/04/07 | UNDET. | NO | - |
| MBG/08/04/08 | 38 | NO | + |
| MBG/08/04/09 | 26 | YES | + |
| MBG/08/06/10 | 22 | YES | + |
| MBG/08/09/11 | 37 | NO | + |
| MBG/08/10/12 | 24 | YES | + |
| MBG/08/10/13 | 25 | YES | + |
| MBG/08/10/14 | 25 | YES | + |
| MBG/08/10/15 | 25 | YES | + |
| MBG/08/11/16 | 28 | YES | + |
| MBG/09/09/17 | 24 | YES | + |
| MBG/09/12/18 | 34 | NO | + |
| MBG/10/01/19 | 28 | YES | + |
| MBG/10/02/20 | 26 | YES | + |
| MBG/10/03/21 | 40 | NO | + |
| MBG/10/05/22 | 27 | YES | + |
| MBG/10/05/23 | 27 | YES | + |
| MBG/10/06/24 | UNDET | NO | - |
| MBG/10/09/25 | UNDET | NO | - |
| C35 (POSITIVE CONTROL) | 20 | YES | + |
| NEGATIVE CONTROL | UNDET | NO | - |

The $C_T$ values represent the extent of positivity using the qRT-PCR assay. "UNDET" represents a absence of virus where as the numbers reflect inverse viral titres. Presence of virus using the CPE assay is shown by a "YES" and absence by "NO". Presence of virus in culture by IFA is denoted by "+" and absence of virus is denoted by "-".

### 4.3.2 Conventional PCR

In order to sequence the positive HPIV-1 HN gene from the cell cultures, performance of conventional PCR assay was necessary. An amplicon of the 360bp was obtained by PCR with the reference C35 strain (Fig 4.1). No amplification was observed with the negative control (Fig.4.1). The 21 test cases of the HPIV-1 results were consistent with those obtained in qRT-PCR and generated the expected 360bp amplicon (Fig 4.1). Upon excision and gel purification to remove excess salt, primers, dNTPs and non-specific products, a single amplicon of 360bp was obtained (Fig. 4.2).

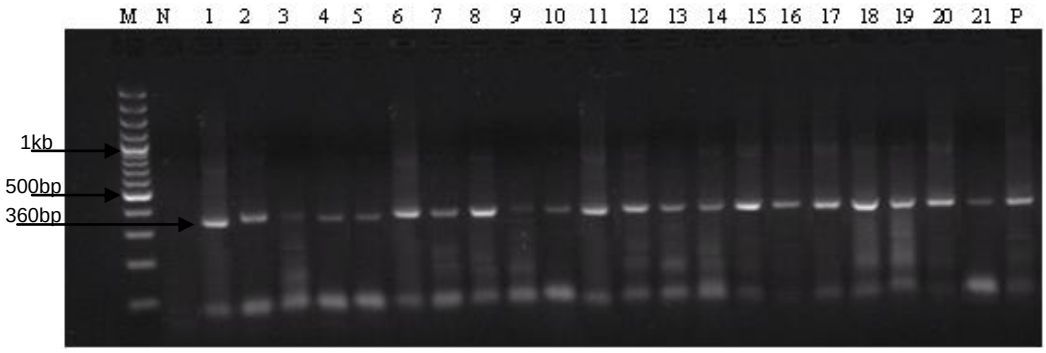



**Figure 4.1:** Gel electrophoresis of HPIV-1 HN PCR amplicons.
The lanes represent the molecular weight marker(M),negative template control (N),C35 positive control (P), DNA product for each sample 1-21;

1) MBG/07/06/01, 2) MBG/07/07/02, 3) MBG/07/08/03,
4) MBG/08/04/05, 5) MBG/08/04/06, 6) MBG/08/04/08,
7) MBG/08/04/09, 8) MBG/08/06/10, 9) MBG/08/08/11,
10) MBG/08/10/12, 11) MBG/08/10/13, 12) MBG/08/10/14,
13) MBG/08/10/15, 14) MBG/08/11/16, 15) MBG(09/09/17,
16) MBG/09/12/18, 17) MBG/10/01/19, 18) MBG/10/02/20,
19) MBG/10/03/21, 20) MBG/10/03/22, 21) MBG/10/05/23.

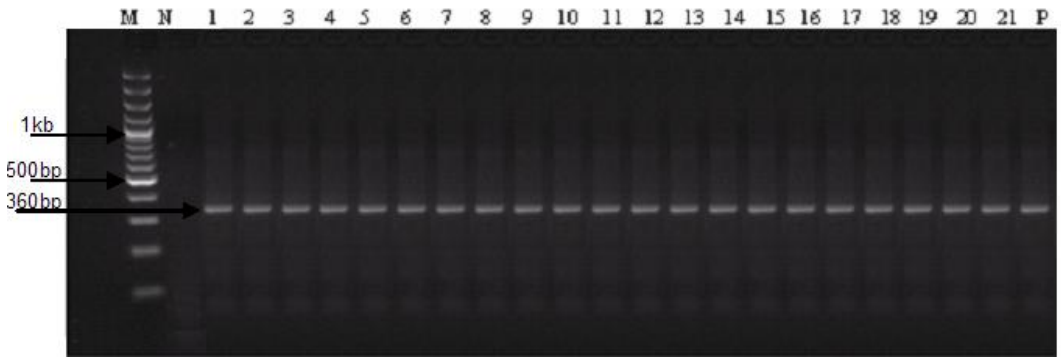



**Figure 4.2:** Gel electrophoresis of purified HPIV-1 HN PCR amplicons.
The lanes represent the molecular weight marker(M),negative template control (N),C35 positive control (P), DNA product for each sample 1-21;

1) MBG/07/06/01, 2) MBG/07/07/02, 3) MBG/07/08/03 (3), MBG/08/04/05 (4), MBG/08/04/06 (5),MBG/08/04/08 (6), MBG/08/04/09 (7), MBG/08/06/10 (8),MBG/08/08/11 (9),MBG/08/10/12 (10), MBG/08/10/13(11),MBG/08/10/14(12), MBG/08/10/15(13), MBG/08/11/16(14), MBG(09/09/17 (15), MBG/09/12/18 (16), MBG/10/01/19 (17), MBG/10/02/20(18), MBG/10/03/21(19),MBG/10/03/22 (20), MBG/10/05/23(21).

### 4.4 Nucleotide sequences of the purified 360bp HN amplicons

Upon sequencing, a four-colour chromatogram showing peaks that represent each of the four DNA bases was obtained. The qualities of the chromatogram at both ends were not good and therefore were trimmed. After trimming, a 267bp long sequence starting from nucleotide position 7254-7521 of the genome of each sample was analyzed. The FASTA nucleotide sequence for each of these isolates is presented in appendix VIII.

### 4.5 Nucleotide sequence alignment of the HN gene fragments of the HPIV-1 isolates

The HN genes of the 21 isolates obtained were 267bp long with an open reading frame (ORF) coding for 89 amino acids (Appendix IX). This 267bp portion of the HN gene segments corresponded to nucleotides 7254-7521 of the HN gene in the HPIV-1 genome using the Washington 1964 HPIV-1 reference strain numbering. Multiple sequence alignment for the 21 nucleotide sequences of the isolates and representative strains (Acc. Nos: M86786.1, U70938.1, U86785.1, U70936.1, U70937.1, U70947.1 and the positive prototype C35) revealed conserved and variable regions within these nucleotide sequences. Overall, all Mbagathi sequences were similar to corresponding haemagglutinin-neuraminidase sequence segments of disparate HPIV-1 isolates found in the GenBank. Among themselves, the highest nucleotide sequence similarities were observed between isolates MBG 08/04/08, and MBG 10/03/22 (98%), while the lowest similarities were between isolates MBG 07/06/01 and MBG 10/03/21 (94%) (Appendix X). Compared to the HN gene sequence of the Washington 1964 HPIV-1 reference strain, similarities ranged from 94% to 99%: the highest similarity was with isolate MBG 10/3/21 (99%) and the lowest with isolate MBG 07/06/01 (94%).

## 4.6 Nucleotide sequence polymorphism in the HN gene amongst the HPIV-1 isolates

Forty-eight variable (polymorphic) sites with a total of 55 mutations were identified when the Mbagathi sequences were analyzed together. In addition, eleven synonymous changes were observed at positions 7305, 7356, 7362, 7422, 7437, 7479, 7740, 7506, 7509, 7512 and 7521 (using the Washington 1964 HPIV-1 reference strain numbering). Finally, thirty-six replacement changes (Single nucleotide polymorphism) were observed at positions 7256, 7259, 7260, 7261, 7265,7265, 7279, 7291,7295, 7298, 7299, 7311, 7315, 7316, 7316, 7320, 7328, 7328, 7345, 7349, 7384, 7390, 7391, 7393, 7418, 7420, 7454, 7459, 7470, 7486, 7493, 7499, 7509, 7510, 7511 and 7513 (see appendix XI).

Amongst the isolates analyzed, nucleotide sequences at regions 7363-7383 and 7394-7417 (Fig. 4.3) showed the highest conservation in all the mbagathi isolates.

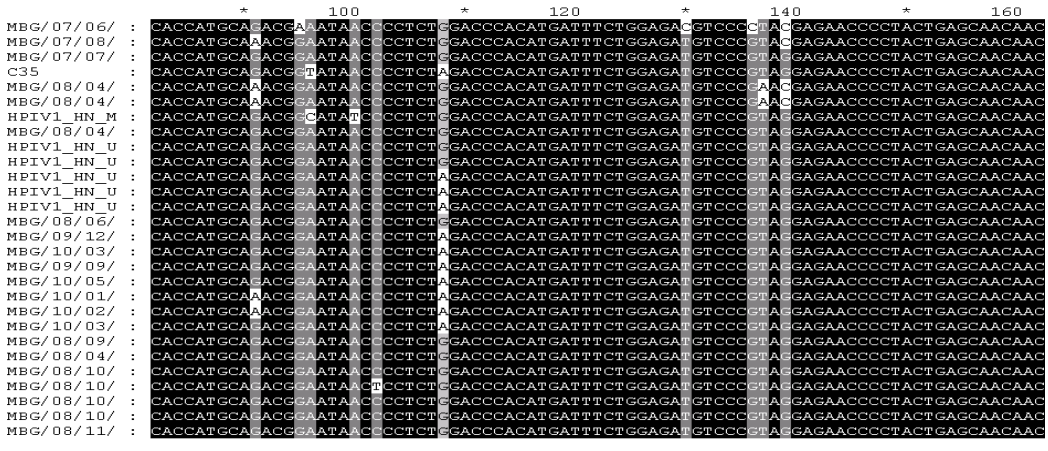



**Figure 4.3:** Nucleotide alignment of selected HN fragment showing conserved regions and polymorphic regions.

A fragment of HN HPIV-1 ranging from nucleotide 7336-7416 from 21 isolates including the reference strains and HPIV-1 positive control were aligned. The name followed by a number represents the source of the sample. Polymorphic sites have a grey background while the conserved regions have a dark black background. The numbers at the end of the sequence represent the position of the last nucleotide. The alignment was done with Muscle3.6 (Edgar, 2004). The aligned sequence was imported into GeneDoc Version 5.1.2600.2180 for visualization and viewing.

There were two loci with increased frequency of mutation along the HN gene. These included regions covering nucleotides 7291 to 7320 and 7345 to 7362 (using the Washington 1964 HPIV-1 reference strain numbering). All the HN gene segments of the Mbagathi isolates differed from those of the prototype C35 strain at 8 positions. These included position 7266, 7295, 7320, 7350, 7362, 7461, 7470 and 7511. Of the six reference strains, all except the sequence with accession number M86786.1 had the same nucleotide substitution at positions 7326, 7461 and 7521 which were T, C and G respectively. In contrast, all the Mbagathi isolates had different nucleotide substitutions at these positions which were C, T and A respectively. However, similar to the other reference strains included in this study, M86786.1 had T and C substitutions at position 7470 and 7511 (Appendix XI) .

### 4.7 Amino acid sequence substitutions amongst the HPIV-1 isolates compared to reference strains

The portion of the HN protein analyzed consisted of 89 amino acids from position 118 to 205 (according to the Washington 1964 HPIV-1 reference strain numbering). This protein region corresponds to the DNA region whose nucleotide sequences had been determined (section 4.5). There were 23 positions at which 47 mutations were observed among the amino acids.

MBG/08/04/09 was the only isolate that had one amino acid substitution at codon 140 which was a conservative change i.e. from Glutamate (E) to Aspartate (D) since these two amino acids have similar lengths and charges/polarities. The rest of the isolates had amino acid substitutions with different charges or changes from polar to non-polar side groups. Furthermore, 9 out of 21 (43%) of the isolates had a change at codon 195 from an arginine (R) to a serine (S). MBG/07/07/02 also had mutations at four codons at positions 121, 131, 137 and 174 where the amino acid common to the rest of the isolates ( D, N, Q, N) was replaced with amino acids of different charges (V, H, H, Y) respectively. MBG/08/04/09, had two mutations at codons 120 and 140 where Arginine (R) was replaced with Serine (S) and Glutamate (E) replaced with Aspartate(D).

Amongst the isolates, nine instances of a substitution involving replacement of Arginine (R) with Serine (S) were observed at codon 195. Similarly five instances of substitutions involving replacement of Glutamine (E) with Lysine (K) were

observed at codon 127 and four instances where Glycine (G) replaced Arginine (R) observed at codon 165 (Fig. 4.4).

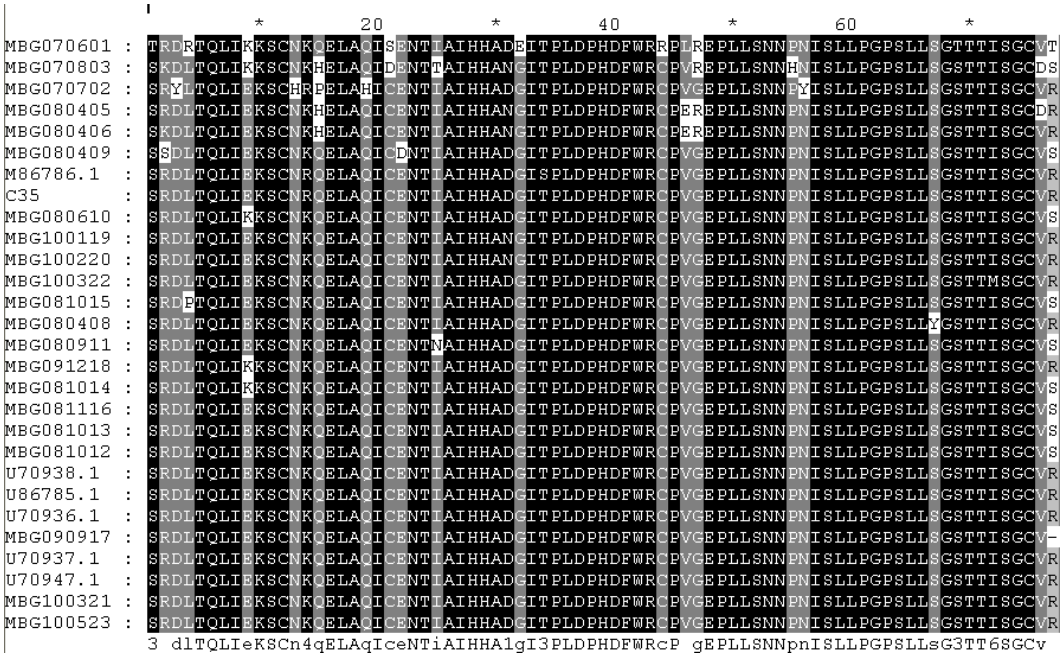



**Figure 4.4:** Multiple sequence alignment of all isolates showing amino acid substitutions within the region 118-195.

### 4.8 Statistical analysis of neutrality

Several sequences statistical methods of testing neutrality of mutations were applied to the from region HN.

#### 4.8.1 Measuring neutrality using the synonymous and non-synonymous substitutions

For the above tests, the aligned sequences were fed into Mega 5 and codon based Z-test was applied.In this test negative selection was examined by testing the null hypothesis of dN<dS. It was observed that the probability of dN<dS was less than

one (dN<dS<1) (see appendix XII). This implied that mutations occurring with the 267bp region were under negative selection.

### 4.8.2 Kumar method

Using, Kumar's method (Kumar et al., 2004) it was observed that dS=0.1230 and dN=0.0890 and hence φ=dN/dS=0. 0.0890/0.1230 = 0.7235.

### 4.8.3 Hughes method

Radical synonymous substitutions were observed at seven codons at positions 127 (E-K), 131 (N-H), 120 (R-S), 137 (Q-H), 162 (C-R), 194 (R-S) and 195 (R-T) (Table 4.2). A single conservative non-synonymous substitution was observed at codon 140 (E-D). The substitution E-K at codon 127 changes an amino acid from negatively charged to one that is positively charged while R-S and R-T changes convert positively charged residue to neutral residue. N-H, Q-H, C-R, is a substitution of a neutral amino acid with a positively charged amino acid.

**Table 4.2:** Non-synonymous radical substitutions at 7 codon positions.

| Codon position | Non-synonymous Radical substitutions |
|---|---|
| 127 | E-K |
| 131 | N-H |
| 120 | R-S |
| 137 | Q-H |
| 162 | C-R |
| 194 | R-S |
| 195 | R-T |

### 4.8.4 Tajima's test

Tajima and Fu and Li's test (Tajima, 1993) was used to test for excess or lack of singleton nucleotides by comparing estimates of θ based on the number of singletons vs that derived from S (the D* index) or π (the F* index). The table 4.3 summarizes the output.

**Table 4.3:** Tajima's test of neutrality.

| Gene | Size (bp) | N | S | Si | K | π | θ | Tajima D | Fu&Li's D* | Fu&Li's F* |
|---|---|---|---|---|---|---|---|---|---|---|
| HN (7254-7521) | 267 | 19 | 48 | 0 | 19 | 0.031150 | 0.05806 | -1.87251 | -2.61989 | -2.79312 |

The average N, is the number of alleles sequenced; S, number of polymorphic sites; Si, number of singleton nucleotide alleles; K, number of haplotypes; π, observed average pairwise nucleotide diversity; θ, expected nucleotide diversity under neutrality (derived from S).

The nucleotide diversity index (π) for HN gene starting from nucleotide 7254-7521 was 0.03115. This means that there were 3.1% nucleotide differences between pairs of alleles on average. The nucleotide frequency distribution was tested for statistical parameters from neutral expectations. The overall value of Tajima's D for the HN gene was negative (-1.87251).

### 4.9 Testing potential confounders in the method of Nei and Gojobori

The method of Nei and Gojobori assumes a uniform codon usage and an equal rate for transition (T-C and A-G) and transversions (A-T, A-C, G-C, G-T). These two parameters were tested as discussed below since they could confound the outcome of the test.

### 4.9.1 Codon usage bias

To compute for this, the aligned sequences were fed into DNAsp and the codon usage bias at the 1st, 2nd and 3rd positions determined. Effective number of Codons (ENC) and Codon Bias Index (CBI) (Appendix XIII) were used as general measurement of codon usage bias. ENC versus GC3 plot revealed the influence of base composition constraints imposed on codon usage preferences (Fig. 4.5).

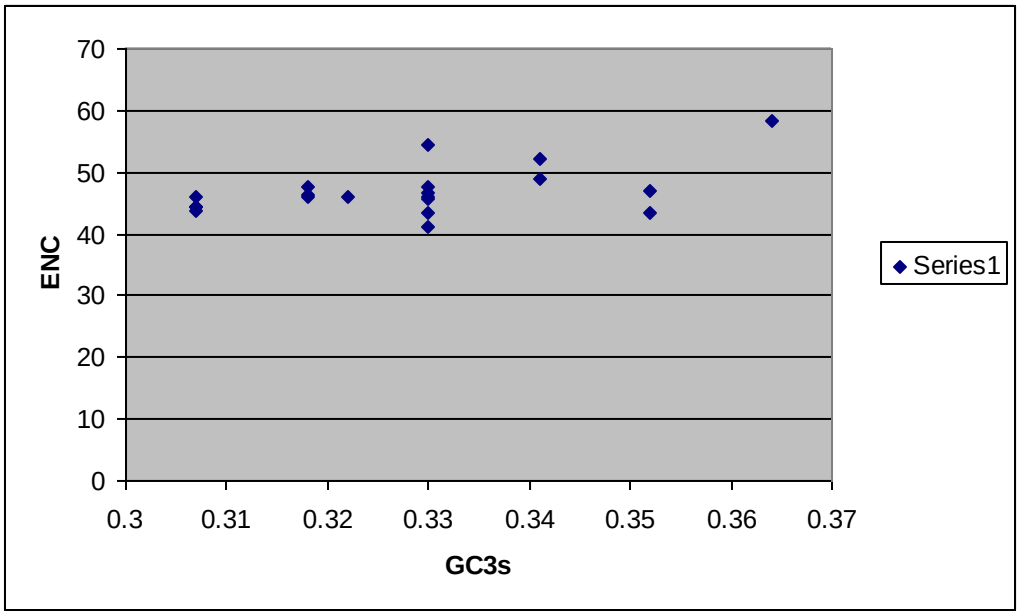



**Figure 4.5:** Distribution of the effective number of codons (ENC).
ENC in relation to the GC-content at the synonymous third codon position (GC3s) of the 21 HPIV-1 isolates.

### 4.9.2 Transition/ Transversion ratio

When the aligned sequence were analysed using Mega 5 (Kumar *et al.*, 2004), it was found that the nucleotide frequencies in HPIV-1 HN from M86786.1 (the reference sequence) was T (25.1%), C (25.5%), A (31.4%) and G (18%). Looking

at the field isolates, the average nucleotide frequencies were: T (24.9%), C (25%), A (32.5%) and G (17.6%).

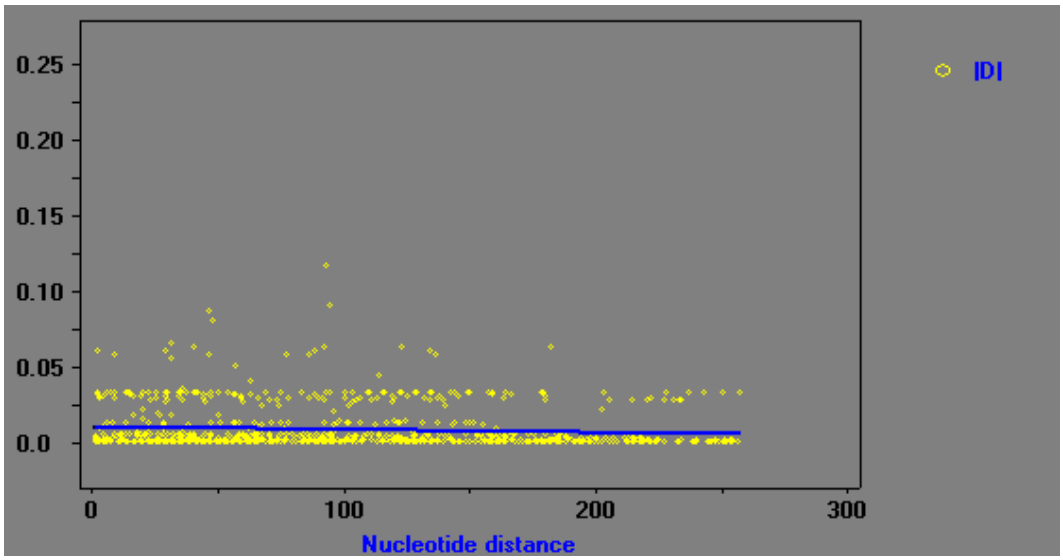
It was observed that out of the 55 mutations 25 (45%) led to A-G (11), C-T (14) transitions versus 30 (55%) A-C (9), A-T (10), G-C (4), G-T (7) transversions (Table 4.4). Using Kimura-2 parameters, the ratio of transition/ transversion (ti/tv) was computed. Transition was found to be similar to transversion (ti/tv= 0.818).

**Table 4.4:** Transitions and Transversions of the HN gene, HPIV-1.

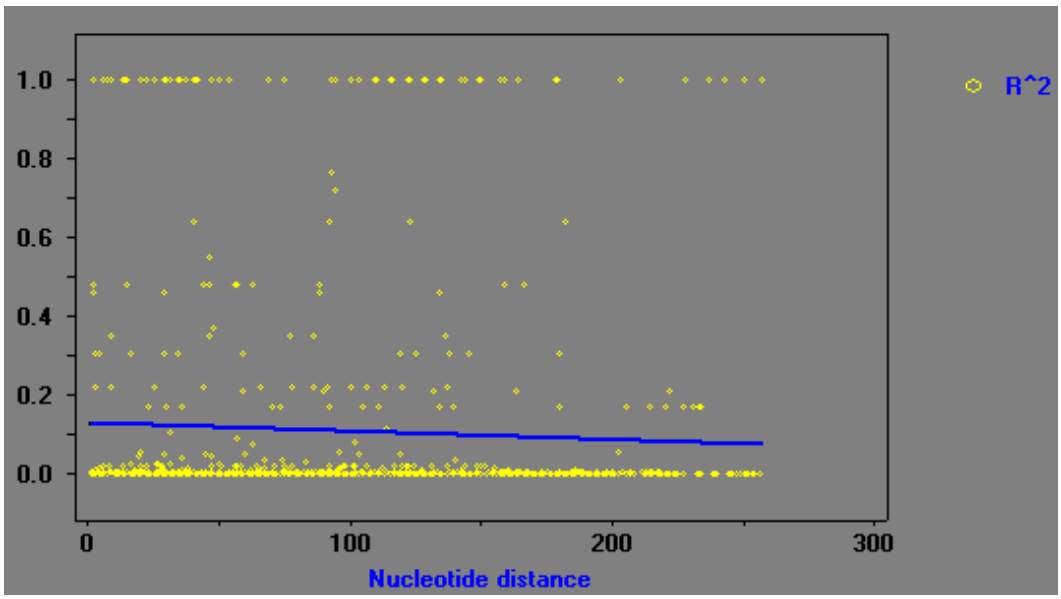
| **Transitions** | | **Transversions** | | | |
|---|---|---|---|---|---|
| A-G | C-T | A-C | A-T | G-C | G-T |
| 11 | 14 | 9 | 10 | 4 | 7 |

### 4.9.3 Recombination and Linkage disequilibrium

The presence of recombination influences the ability to detect selection since it breaks up the associations between sites under selection and linked variation (Charlesworth *et al.,* 1997). Therefore, measures of recombination and linkage disequlibrium were investigated across the region of HPIV-1 HN studied by passing aligned sequences through DNAsp.4.0. Linkage disequilibrium (LD) as measured by D and recombination measured by R^2 (Fig.4.6), decline with nucleotide distance (Lewontin 1964), showing negative correlations that are significant ($P<0.05$) with high significant values visibly clustered in the top left-hand corner of each plot.




**|D| values: Y = 0.0117 - 0.0203X (861 points)**




**R^2 values: Y = 0.1316 - 0.2080X (861 points)**

**Figure 4.6:**Linkage disequilibrium (D) and recombination (R^2).
A decline with nucleotide distance is observed

All Polymorphic sites were considered; Sites segregating for three or four nucleotides were not considered; Number of polymorphic sites analysed was 48; Number of pairwise comparisons was 861. The blue indicate significant point while yellow non-significant. |D| and R^2 decline with distance (inverse relationship) as evident in A and B respectively, implying high recombination rate.

### 4.10 Evolutionary tree of HPIV-1 isolates

#### 4.10.1 Using nucleic acid sequences to generate a phylogenetic tree

Phylogenetic relationships were inferred from the aligned nucleic acid sequences by the Bayesian method was implemented in the Fig tree program. 4 distinct clusters (branches) in a ladder like fashion form. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Support for specific tree topologies was estimated by posterior probability analysis. All the reference strains clustered together under 1 main branch with 67% posterior probability, and further divided into 2 distinct branches. MBG 07/06/01 was on a separate branch from the rest of the isolates though all shared a common ancestor. In addition, MBG07/06/01 had 11 unique nucleotide substitutions that were specific to it only while the rest of the isolates and reference strains had the same nucleotide at the respective positions (Table 4.5).

**Table 4.5:** Unique codon substitution for MBG/07/06/01.

| **Codon position (MBG/07/06/01)** | **Substitution** |
|---|---|
| 7256 | G to C |
| 7265 | G to T |
| 7349 | G to C |
| 7384 | G to A |
| 7390 | T to C |
| 7459 | T to A |
| 7484 | G to C |
| 7493 | C to T |
| 7499 | C to T |
| 7506 | T to G |
| 7513 | A to T |

Most of the 2006-2008 isolates clustered together on a separate branch. Similarly the 2009-2010 isolates clustered together on another separate branch as shown on Fig. 4.7.

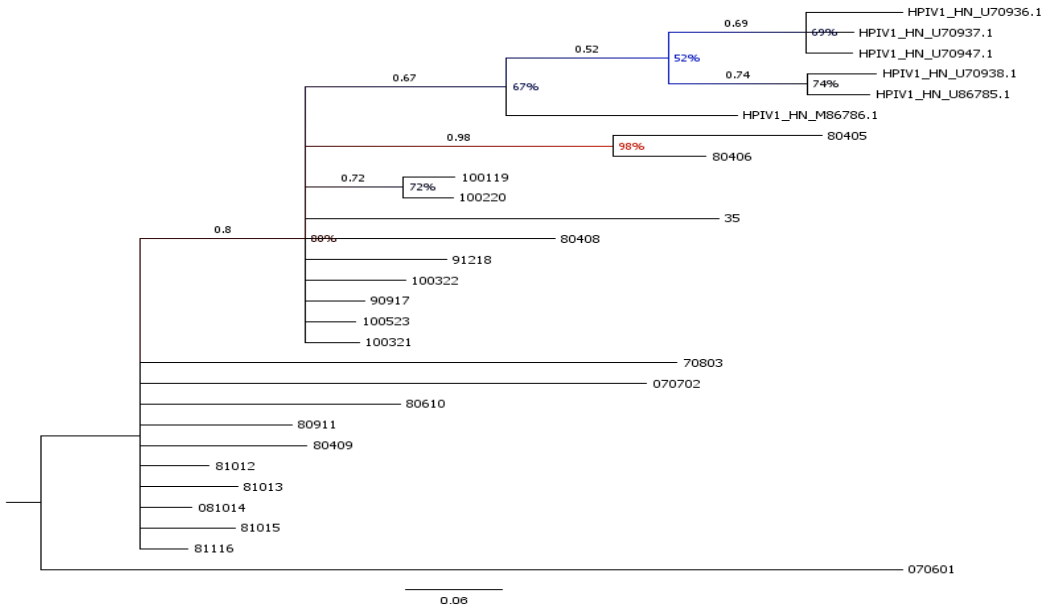



**Figure 4.7:** Phylogenetic topology of MBG HPIV-1 HN gene.
A Topology tree prepared in MEGA5 and compared to type strains obtained from GenBank. The length of the horizontal bar indicates the estimated substitution per nucleotide position.

### 4.10.2 Generation of a phylogenetic tree using amino acid sequences

There were 4 distinct branches that were formed. As in the nucleotide clustering, the reference strains clustered on the same branch. However, M86786.1 clustered with C35 strain and MBG 07/07/02. MBG07/06/01 which was also on a separate branch in the nucleic acid tree clustered with MBG081014, MBG 080610 and MBG 070601 as shown in Fig.4.8.

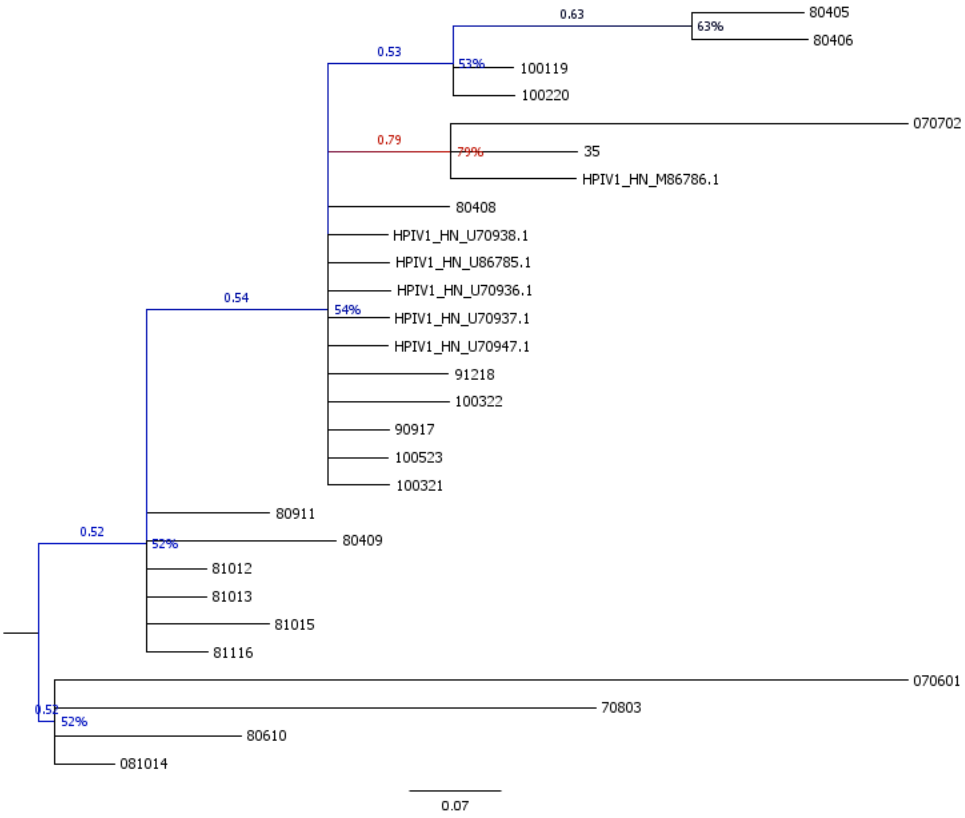



**Figure 4.8:** Phylogenetic topology for MBG HPIV-1 HN amino acid sequences. The length of the horizontal bar indicates the estimated substitution per amino acid position.

### 4.11 Molecular Evolution rate

The evolutionary rate was determined by use of an Evolutionary Calculator in DNAsp and it was found to be $11.2 \times 10^{-4}$substitutions/site/year.

# CHAPTER FIVE

## 5.0 DISCUSSION

Knowledge of the molecular characteristics of HPIV-1 is important in aiding vaccine development and in identifying therapeutic targets of this ubiquitous pathogen. Molecular characterization and phylogenetic analyses of HPIV-1 has mainly been based on studies carried out in developed countries. Isolation and molecular characterization of HPIV-1 has not been documented in Kenya. Yet, there is a real potential that HPIV-1 strains found in Kenya may be genetically distinct or may evolve differently leading to allelic variants that may be unique. Allelic polymorphism of HPIV-1 HN presents a dilemma for vaccine development. To begin to unravel these molecular variations amongst these viruses circulating in Kenya, twenty-five HPIV-1 clinical isolates were analyzed to determine the extent of antigenic and genetic variation. The isolates analyzed here were obtained from Mbagathi District Hospital in Nairobi. These isolates were obtained from samples collected over a span of approximately 4 years beginning in 2006.

This study confirmed Niesters' observations that presence of HPIV-1 in a patient specimen (NP swab) can be detected by viral isolation in cultured LLC-MK2 cell line with a time lag of 7–10 days and molecular (PCR) tests (Niesters, 2002). Whereas majority of patient specimens analyzed showed CPE's between days 7-10 and were later confirmed to have the virus using serology (IFA) and molecular (PCR) assays, there were some specimens that did not show any CPE

upon inoculation in cell culture but indicated presence of virus using the PCR assay. This result may be explained by the fact that the different patients had different viral loads in their systems at the time of sampling due to either disparate immune status of the patients or due to variation in the time of sampling since was infection was established (Noyola *et al.,* 2000). Viral load of a specimen is directly proportional to how quickly CPE develops in inoculated cultures of a susceptible cell line (Tian *et al.,* 2012). Furthermore, scoring of the presence/absence of CPE is subjective and is influenced by the experience of the individual making the readings (Gupta *et al.,* 1996). Thus, some flasks may have indeed had CPE but due to the subjective nature of this assay, they were scored as negative. Finally, the fact that the molecular assay detected virus in all samples that were inoculated and showed CPE as well as others that did not show CPE, may be attributed to the high sensitivity of the molecular (PCR) assay in detecting presence of virus in a specimen (Ellis and Zambon, 2002). Overall, these results confirm that the molecular assay (PCR) is the most sensitive when detecting presence of virus in a nasopharyngeal (NP) patient specimen (Ellis and Zambon, 2002).

The traditional rapid diagnostic test for HPIV-1, Immunofluorescence assay (IFA), has levels of sensitivity and specificity that come close to those of RT-PCR (Rahman *et al.,* 2008). This assay relies on antigen-antibody reactions to detect the presence of the virus in clinical samples. The PCR assay detected virus in all the twenty-five patient samples. However, using IFA, twenty one of these

were reported as containing HPIV-1. The failure of IFA to detect HPIV-1 virus in the four out of the twenty-five samples might be due to a low viral load or the presence of non-infectious virions in the specimen, which may have resulted from insufficient specimen or inappropriate transport or storage conditions (i.e. specimens were kept for extended period at room temperature or higher) consistent with previous reports (Eugene-Ruellan *et al.,* 1998). The immunofluorescence assay to detect antigens has been associated with sensitivity problems when the viral load is low (Casiano-Colon *et al.,* 2003).

RT-PCR was employed to investigate the presence of HPIV-1 virus in the twenty five putative isolates for HPIV-1. Similar to the IFA results discussed above, twenty one of these putative isolates generated the expected 360bp bands using specific primers. Even though RT-PCR is the gold-standard diagnostic method for HPIV-1 due to its high sensitivity and specificity (Ellis and Zambon, 2002), these results taken together indicate that both methods are sensitive and specific enough for HPIV-1 detection. Furthermore, the absolute concordance in the positivity of both methods indicated that there were no strain sequence variation amongst the Mbagathi HPIV-1s in the primer annealing regions of the HN gene. Although primers may be designed to react with temporally and geographically diverse HPIV-1 isolates, unanticipated strain sequence variation in the primer regions could result in occasional false-negative results (Echevarria *et al.,* 1998).

Similar to the RT-PCR and IFA results, the 21 isolates were also positive using the real time quantitative RT-PCR (qRT-PCR) assay. These observations validated the sensitivity and specificity of the IFA and qRT-PCR assays since qRT-PCR technique is the most sensitive technique for identifying target pathogens (Terlizzi *et al.*, 2009). However, disparate isolates gave different $C_T$ values. The lowest $C_T$ value was 24 while the highest was 39. The high $C_T$ values were indicative of low viral load of HPIV-1 in the sample isolates. This is supported by observations by Wan *et al.*, 2011 who found a linear relationship between the copy number/viral particles present in the sample and the $C_T$ values on the chromatogram. In fact, $C_T$ values have been used to estimate the viral load in a sample (Alma *et al.*, 2007) with an inverse relationship whereby high $C_T$ values represent a low viral load, while a low $C_T$ value show a high viral load.

When qRT-PCR was performed on the nine samples that yielded virus when inoculated into culture, the mean $C_T$ value was 25.6 whereas amongst the twelve that did not yield virus in culture, all had un-determined (ut) $C_T$ value. These findings mirrored those of Bredius *et al.*, 2004 who observed $C_T$ values ranging from 27 to 42 amongst NP samples considered positive only by qRT-PCR versus $C_T$ values ranging from 18 to 22 for culture positive samples. The un-determined results may have been due to degradation of viral RNA in the samples arising from multiple freeze-thaw cycles between the initial isolation of the HPIV-1 and the RT-PCR procedures (Echevarria *et al.*, 1998).

In this study, the consensus nucleotide sequence for a 267bp region of the HPIV-1 HN genomic RNA was determined and compared to the Washington 1964 wild-type HPIV-1 strain found to be virulent (Murphy *et al.,* 1975). The present HPIV-1 isolates showed an overall high level of nucleotide sequence identity (94-98%) of this HN amplified region. This result suggests that several lineages of highly conserved HN HPIV-1 were prevalent in the Mbagathi isolates. This is in line with the study findings of Katsumi *et al.,* 2011, who found that HN gene is highly conserved in HPIV-1.The aligned nucleotide positions of HPIV-1 HN, and ranging from nucleotide position 7254-7521 (Washington 1964 isolate numbering system) exhibited low polymorphism. The 2 largest conserved areas on the HN when converted to protein code were between amino acid 151-161 and 175-184. Data from previous studies have shown that areas that are highly conserved are probably in the globular head (Henrickson and Savatski, 1997). Therefore, both the 151-161 and 175-184 regions are probably in the globular head of the HN protein. Earlier studies with various paramyxoviruses have reported that both the stalk and the head region of the HN protein are involved in fusion promotion (Gravel and Morrison, 2003).

Forty eight (48) segregating (polymorphic) sites with a total of 55 mutations were identified when the sequences were analyzed together. There are two things that affect the number of segregating sites (S); i) if the mutation rate per site is assumed to be constant along the sequence, the number of segregating sites should on average be proportional to length of the sequences; ii)S should increase

with sample size (Pluzhnikov and Donnelly, 1996). In the current study, the evident low polymorphism in HN gene can be attributed to both factors, i.e. the short length (267bp) of the region of the gene analyze and small sample size of 25.

Estimation of the transition-transversion (ti/tv) rate bias is important not only in understanding of the patterns of DNA sequence evolution, but also as a reliable estimation of sequence distance and phylogeny reconstruction (Yang *et al.,* 1999). The pattern of nucleotide substitutions yielded an apparent ti/tv of 0.8. This means that there was a bias towards transition substitutions in this region of HN gene. Transition versus transversion bias has been shown to result from the relatively high rate of mutation of methylated cytosines to thymine (Keller *et al.,* 2007). In general, it is assumed that there is a universal bias in favor of transitions over transversions due to the underlying chemistry of mutation. Saturation of transitions at high levels of genetic divergence is commonly believed to explain the bias towards transition (Moritz *et al.,* 1987).

In terms of physico-chemical properties, polymorphisms at the amino acid level in the region analyzed affected charge, size, and aromacity or were conserved. There were 23 substitutions involving a change in charge. The most prevalent involving 9 sites was the change from arginine (positively charged) to serine (small uncharged) followed by 5 sites involving change from glutamate (negatively charged) to lysine (positively charged). Others were asparagines (small polar) to histidine and glutamine to histidine. In terms of size, the observed substitutions were asparagine-tyrosine. Other substitutions like glutamate-

aspartate, arginine- serine, aspartate-valine and glycine- arginine affected size only and in some cases polarity. There was a tendency to move from non-polar to negative charge observed. In conclusion, there was more radical non-synonymous amino acid changes (7 positions) observed than conservative non-synonymous changes (1 position) on the HN gene segment. Since shape defines the complementarity of an epitope and paratope (Frank 2002), any alteration in the size of the residues that constitutes an epitope could reduce the strength of an immune response.

Several tests of neutrality have been developed and applied to real data. The neutral theory maintains that most observed molecular variation-both polymorphism within species and divergence between species is due to random fixation of selectively neutral mutations (Kimura, 1983). In this study several of these tests were applied to ascertain each other. Analysis of selection pressure in the isolates showed that synonymous substitutions (dS) predominated over non-synonymous substitutions (dN), and no positively selected sites were found in HN protein. This implied that the amino acid change was deleterious, purifying selection thus reducing its fixation rate, therefore $\varphi<1$ (Yang *et al.,* 2000). Similar results were obtained by Mizuta *et al.,* 2011, who did a selection pressure analysis of HN gene HPIV-1 in Japan, and obtained a low dN\dS ratio of 0.17. It is also a possibility this could result in stabilizing selection through the purging of deleterious variations that arise (Hudson *et al.,* 1995).

Tajima's test of neutrality confirmed the above observation, where D was found to be negative for HN region of HPIV-1, which was significant and consistent with the hypothesis of selective sweeps. When an advantageous mutation appears in a population and sweeps through to fixation, it leaves behind a trail in patterns of linked neutral diversity (Maynard *et al.*,1974). A feature of genealogies under hitch-hiking is that if there are mutations, they are likely to be recent, hence are likely to be at a low frequency in the sample. Consequently, hitchhiking is associated with negative Tajima D and Fu and Li D statistics (Braverman *et al.*, 1995; Fu, 1996).

Codon usage and genome GC content are highly correlated with the GC compositions on the three codon position and is preferred as an indicator of evolution (Mooers and Holmes, 2000). It is important to identify the determinants of codon choices in order to obtain a better understanding of viral evolution. Based on the effective number of codons (ENC) in relation to the GC content at the synonymous third codon position (GC3) plot (Fig. 4.5), the closely grouped and sparsely distributed ENC values and GC3s for the HPIV-1 suggest that there was a similar codon usage preference shared among the Mbagathi isolates. Previous studies from human RNA viruses (Jenkins and Holmes, 2003) reported that mutational pressure is an important determinant of the codon bias observed. However, their results showed that weak translational selection may also have some influence in shaping codon usage bias. Given the evidence that both mutational pressures and selection are involved in the phenomenon of codon bias,

the current accepted model is the major codon preference model, also known as the mutation-selection-drift balance model of codon bias (Bulmer, 1991; Chen *et al.,* 2004). This model proposes that selection favors the major (or preferred) codons over minor codons. However, mutational pressure and genetic drift allow the minor codons to persist. Under this mutation selection- drift balance model, codon bias is the result of positive selection for mutations that increase the frequency of major codons (preferred mutations) and purifying selection against mutations that decrease the frequency of major codons (unpreferred mutations). This model postulates that the selection on codon bias is generally weak (Bulmer, 1991; Duret, 2002).

Evidence of recombination was observed in region 7254-7521 of the HN gene of HPIV-1 as demonstrated by negative correlation between linkage disequilibrium and nucleotide distance. Linkage disequilibrium (D) values were higher than those estimated for recombination (R^2). In this study, alleles that are close to one another were found to have significant linkage disequilibrium, as compared to those that were far apart, though the whole region analysed was just 267 bp long. This can be explained by the fact that recombination rates are known to vary by more than an order of magnitude across the genome. Because breakdown of linkage disequilibrium is primarily driven by recombination, the extent of D is expected to vary in inverse relation to the local recombination rate (Nachman, 2002). Another possible explanation is the hitchhiking effect, in which an entire haplotype that flanks a favoured variant can be rapidly swept to high

frequency or even fixation (Wang *et al.,* 2002; Parsch *et al.,* 2001; Verrelli *et al.,* 2001).

The HPIV-1 isolates showed an overall high level of nucleotide sequence identity (94-99%) of the HN coding region. Pairwise distance values based on the nucleotide sequences among the isolates were relatively low. In addition, there were no positively selected sites found. These results suggest that several quasispecies of HPIV-1 albeit highly conserved in the HN gene were prevalent in Mbagathi isolates.

It was evident that all the 21 samples shared a common ancestor since they all were monophyletic. The nucleotide sequence tree suggests that the immediate ancestor for each of the two main branches further evolved into the variants within each main branch. The 4 distinct branches can be attributed to the likely nucleotide difference. This finding suggests that geography may not be the current evolutionary force leading to the continuation of these genotypes (Hetherington *et al.,* 1994). This is also in line with findings from a study done that found unique genotypes may circulate throughout the country and then reappear when they experience immunologic pressure to form multiple quasispecies around the evolving parent genotypic strain (Ambrose *et al.,* 1995).The ladder-like phylogeny demonstrated by this tree suggests that evolution of HPIV-1 is occurring with HN maintaining its function and may be useful in predicting the emergence of antigenic variants of HPIV-1 circulating in the region.

Phylogenetic analysis can give an estimation of the viral evolution rate and cluster classification. Furthermore, an evolutionary calculator can enable analysis of the time scale of the evolution of viral genes. In the present study, the viral evolution rate of HPIV-1, cluster classification, and the evolutionary time scale of the present isolates was estimated by applying an evolutionary calculator to the detailed phylogenetic analysis of the HN coding region in HPIV-1. The rate of molecular evolution ($11.2 \times 10^{-4}$substitutions/site/year) was comparable to that obtained in another HPIV-1 genome study carried out in Japan (Mizuta *et al.*, 2011) that found the rate to be $7.68 \times 10^{-4}$ substitutions/site/year. MBG 07/06/01 seemed to evolve at a different rate than the rest of the isolates demonstrated by the distinct branching pattern and the unique substitutions shown in both the nucleotide and amino acid phylogenetic trees. It is possible that genome properties other than size, such as polarity or structure, may be associated with substitutions of the viral genome (Sanjuán *et al.,* 2010). The results implied that the antigenic and genetic subgroups are very stable.

In summary, these results showed that HN protein is highly conserved. In addition, no positively selected sites were detected. To my best knowledge, this is the first report of these findings of HPIV-1 HN gene in the region.

# CHAPTER SIX

## 6.0 CONCLUSIONS AND RECOMMENDATIONS

### 6.1 Conclusions

i. The presence of HPIV-1 was confirmed in the archived isolates upon culturing in LLCMK2 cell line.

ii. The HN gene of the positively identified isolates was amplified using RT-PCR.

iii. The sequences of the HN gene segment of 267bp was successfully analyzed.

iv. The phylogenetic analysis of the sequenced HN gene segment of HPIV-1 isolates was performed which revealed that it was under negative selection.

### 6.2 Recommendations

i. Vaccine manufacturers should consider developing a subunit vaccine against HPIV-1 based on the HN protein because the current study has revealed that this protein is antigenically stable therefore amenable to easy vaccine targeting.

ii. Comprehensive studies to document virus strain characteristics of HPIV-1 should be carried out since these will be important for the selection of appropriate vaccine strains (in terms of culture growth characteristics) and may thus contribute to vaccine development.

## REFERENCES


**Ali, A. and Nayak, D. P. (2000).** Assembly of Sendai virus: M protein interacts with F and HN proteins and with the cytoplasmic tail and transmembrane domain of F protein. *Virology*, **276:** 289–303.

**Alma van de Pol, C., Anton van Loon M., Tom, F. W., Nicolaas, J. G., Monique N., Els, K. B., Rob, S. and John, W.A. (2007). .**Increased Detection of Respiratory Syncytial Virus, Influenza Viruses, Parainfluenza Viruses, and Adenoviruses with Real-Time PCR in Samples from Patients with Respiratory Symptoms. *Journal of Clinical Microbiology*, **45:** 2260-2262.

**Ambrose, M. R., Hetherington, S. V., Watson, A. S., Scroggs, R. A. and Portner, A. (1995).** Molecular evolution of the F glycoprotein of human parainfluenza virus type 1. *J. Infect. Dis*. **171:** 851–856

**Anne M., (2005).** Entry of parainfluenza virus into cells as a target for interrupting childhood respiratory disease. *J Clin Invest*. **115:** 1688–1698**.**

**Barenfanger, J., Drake, C., Leon, N.., Mueller, T. and Troutt, T. (2000).** Clinical and financial benefits of rapid detection of respiratory viruses: an outcomes study. *J. Clin. Microbiol*. **38:** 2824–2828.

**Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langley, C.H. and Stephan, W. (1995).** The hitchhiking effect on the site-frequency spectrum of DNA polymorphisms. *Genetics*, **140:**783–796.

**Bredius, R. G., Templeton, K. E., Scheltinga, S. A., Claas, E. C., Kroes, A. C. and Vossen, J. M. (2004).** Prospective study of respiratory viral infections in pediatric hemopoietic stem cell transplantation patients. *Pediatr. Infect. Dis. J.* **23:** 518–522.

**Brideau, R. J., Oien, N. L., Lehman, D. J., Homa, F. L. and Wathen, M. W. (1993).** Protection of cotton rats against human parainfluenza virus type 3 by vaccination with a chimeric FHN subunit glycoprotein. *J. Gen. Virol*. **74:** 471-477.

**Bulmer, M. (1991).** The selection-mutation-drift theory of synonymous codon usage. *Genetics,* **129:** 897–907

**Carballal, G., Videla, C. M., Espinosa, A., Savy, V., Uez, O., Sequera, M. D., Knez, V., Requeijo, P. V., Posse, C. R. and Miceli, I. (2001).** Multicentered study of viral acute lower respiratory infections in children from four cities of Argentina, 1993– 1994. *J. Med. Virol*. **64:** 167–174.

**Casiano-Colon, A.E., Hulbert, B.B., Mayer, T.K., Walsh, E.E. and Falsey, A.R. (2003).** Lack of sensitivity of rapid antigen tests for the diagnosis of respiratory syncytial virus infection in adults. *J. Clin. Virol.* 28, 169–174.

**Cavalli-Sforza, L.L. and Edwards, A.W.F. (1967).** Phylogenetic analysis: models and estimation procedures. *Am. J. Hum. Gene.,* **19:** 122–257.

**Chanock, R. M., Johnson, K. M., Cook, M. K., Wong, D. C. and Vargosko, A. (1961).** The hemadsorption technique, with special reference to the

problem of naturally occurring simian parainfluenza virus. *Am. Rev. Respir. Dis*. **83:** 125-129.

**Charlesworth, B., Nordborg, M. and Charlesworth, D. (1997).** The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res. Camb.* **70:** 155–174.

**Chen S.L., Lee W., Hottes A.K., Shapiro L. and McAdams H.H. (2004).** Codon usage between genomes is constrained by genome-wide mutational processes. *Proc. Natl. Acad. Sci. USA,* **101:**3480–85.

**Cockerill, F. R. (2003).** Application of rapid-cycle real-time polymerase chain reaction for diagnostic testing in the clinical microbiology laboratory. *Arch. Pathol. Lab. Med*. **127:** 1112–1120.

**Colman, P. M., Varghese, J. N. and Laver, W. G. (1983).** Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature,* **303:** 41–44.

**Colman, P. M., Hoyne, P. A. and Lawrence, M.C. (1993).** Sequence and structure alignment of paramyxovirus haemagglutinin-neuraminidase with influenza virus neuraminidase. *J. Virol*. **67:** 2972-2980.

**Counihan, M. E., Shay, D. K., Holman, R. C., Lowther, S. A. and Anderson, L. J. (2001).** Human parainfluenza virus-associated hospitalizations among children less than five years of age in the United States. *Pediatr. Infect. Dis. J*. **20:** 646–653.

**Cuevas, L. E., Nasser, A. M., Dove, W., Gurgel, R. Q., Greensill, J. and Hart, C. A. (2003).** Human metapneumovirus and respiratory syncytial virus, *Brazil. Emerg. Infect. Dis*. **9:** 1626–1628.

**Dagher, H., Donniger, H., Hutchinson, P., Ghildyal, R. and Bardin, P. (2004).** Rhinovirus detection: comparison of real-time and conventional PCR. *J. Virol. Methods,* **117:** 113–121.

**Dayhoff, M. O. (1972**). Atlas of Protein Sequence and Structure. *MD:Natl. Biomed. Res.*

**Durbin Anna, P. and Karron, A. R. A. (2003).** Progress in the Development of Respiratory Syncytial Virus and Parainfluenza Virus Vaccines. *Clinical Infectious Diseases,* **37:** 1668–77.

**Duret, L. (2002).** Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* **12:**640–49

**Echevarria, J. E., Erdman, D. D., Swierkosz, E. M., Holloway, B. P. and Anderson, L. J. (1998).** Simultaneous detection and identification of human parainfluenza viruses 1, 2, and 3 from clinical samples by multiplex PCR. *J. Clin. Microbiol*. **36:** 1388–1391.

**Edgar Robert, C. (2004).** MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Research,* **5:**1792-97.

**Ellis, J.S. and Zambon, M.C. (2002).** Molecular diagnosis of influenza. Rev *Med Virol.* **12:**375–89.

**Emmalene, J. B., Adam, C., Sonja, R. S., Peter, L. C., Mario, H, S. and Brian R. M. (2007).** Attenuation and efficacy of human parainfluenza virus type 1 (HPIV1) vaccine candidates containing stabilized mutations in the P/C and L genes. *Virology Journal*, **4**:67.

**Eugene-Ruellan, G., Freymuth, F., Bahloul, C., Badrane, H., Vabret, A. and Tordo, N. (1998).** Detection of respiratory syncytial virus A and B and parainfluenza virus 3 sequences in respiratory tracts of infants by a single PCR with primers targeted to the L-polymerase gene and differential hybridization. *J. Clin Microbiol*. **36:** 796–801.

**Ely, K.H., Ahmed, M., Kohlmeier, J.E., Roberts, A.D., Wittmer, S.T. and Blackman, M.A. (2007).** Antigen-specific CD8+ T cell clonal expansions develop from memory T cell pools established by acute respiratory virus infections. *J Immunol.* **179:** 3535–42.

**Fan, J. and Hendrickson, K. J. (1996).** Rapid diagnosis of human parainfluenza virus type 1 infection by quantitative reverse transcription-PCR-enzyme hybridization assay. *J. Clin. Microbiol*. **34:** 1914–1917.

**Frank A. L., Taber, L. H., Wells, C. R., Wells, J. M., Glezen, W. P. and Pareues, A. (1981).** Patterns of shedding myxoviruses and paramyxoviruses in children. *J. Infect. Dis.* **144:** 433–441.

**Frank, S. A. (2002).** Immunology and Evolution of Infectious Disease, Princeton University Press.

**Frayha, J., Castriciano, S., Mahony, J. and Chenesky, M. (1989).** Nasopharyogeal swabs and nasopharyngeal aspirates equally effective for the diagnosis of viral respiratory disease in hospitalized children. *J. Clin. Microbiol*. **27:** 387–1389.

**Fryer, A. D. and Jacoby, O. B. (1991).** Parainfluenza virus infection damages inhibitory Mz muscarinic receptors on pulmonary parasympathetic nerves in the guinea-pig. *B. J. Pharmacol*. **102:** 267–271.

**Fu, Y.X. (1996).** New statistical tests of neutrality for DNA samples from a population. *Genetics*, **143:**557–570.

**Glezen, W. P., Greenberg, S. B., Atmar, R. L., Piedra, P. A. and Cough, R. B. (2000).** Impact of respiratory virus infections on persons with chronic underlying conditions. *JAM,* **283:** 499–505

**Griffin, J. A., Basak, S. and Compans, R. W. (1983).** Effect of hexose starvation and the role of sialic acid in influenza virus release. *Virology,* **125:** 324–334.

**Gross, P. A., Green, R. H. and Curnen, M. G. (1973).** Persistent infection with parainfluenza type 3 virus in man. *Am. Rev. Respir. Dis*. **108:** 894–898.

**Gravel, K. A. and Morrison, T. G. (2003).** Interacting domains of the HN and F proteins of Newcastle disease virus. *J. Virol*. **77:** 11040–11049.

**Gupta C.K., Leszczynski J., Gupta R.K. and Siber G.R. (1996).** An enzyme immunoassay based micro-neutralization test for titration of antibodies to

human cytomegalovirus (CMV) and its correlation with direct ELISA measuring CMV IgG antibodies. *Biologicals*, **2:** 41-9.

**He, B., Lin, G. Y., Durbin, J. E., Durbin, R. K. and Lamb, R. A. (2001).** The SH integral membrane protein of the paramyxovirus simian virus 5 is required to block apoptosis in MDBK cells. *J. Virol*. **75:** 4068–4079.

**Heilman, C.A. (1990).** From the National Institute of Allergy and Infectious Diseases and the World Health Organization. Respiratory syncytial and parainfluenza viruses. *J Infect Dis*. **161:** 402–6.

**Henrickson, K. J. (1991).** Monoclonal antibodies to human parainfluenza virus type 1 detect major antigenic changes in clinical isolates. *J. Infect. Dis*. **164:** 1128-1134.

**Henrickson, K. J. (1994).** Lower respiratory viral infections in immunocompetent children. *Adv. Pediatr. Infect. Dis*. **9:** 59–96.

**Henrickson, K. J. (1998).** Viral pneumonia in children. Semin. *Pediatr. Infect. Dis*. **9:** 227–233.

**Henrickson, K. J. and Savatski, L. L. (1992).** Genetic variation and evolution of human parainfluenza virus type 1 haemagglutinin neuraminidase: analysis of 12 clinical isolates. *J. Infect. Dis*. **166:** 995-1005.

**Henrickson, K. J., Kuhn, S. M., Savatski, L. L. and Sedmak, J. (1994).** Recovery of human parainfluenza virus types one and two. *J. Virol. Methods*, **46:** 189–206.

**Henrickson, K. J. and Savatski, L. L. (1996).** Two distinct human parainfluenza virus type one genotypes detected during the 1991 epidemic. *J. Clin. Microbiol*. **34:** 695–700.

**Henrickson, K. J. and Savatski, L. L. (1997).** Antigenic structure, function, and evolution of the haemagglutinin-neuraminidase protein of human parainfluenza virus type 1. *J. Infect. Dis.* **176:** 867–875.

**Henrickson, K. J., Kuhn, S. M. and Savatski, L. L. (1994).** Epidemiology and cost of human parainfluenza virus types one and two infections in young children. *Clin. Infect. Dis.* **18:** 770–779.

**Henrickson, K.J. and Fan, J. (2000).** PCR assay using unequal primer concentrations to detect HPIV 1,2,3 and RSV A,B and influenza virus A, B. *MCW research foundation.*

**Henrickson, K.J. (2003).** Parainfluenza Viruses. *CMR*.**16:** 242–264.

**Henrickson, K. J. (2005).** Cost-effective use of rapid diagnostic techniques in the treatment and prevention of viral respiratory infections. *Pediatr. Ann*. **34:** 24–31.

**Hetherington, S. V., Watson, A. S., Scroggs, R. A. and Portner, A. (1994).** Human parainfluenza virus type 1 evolution combines cocirculation of strains and development of geographically restricted lineages. *J. Infect. Dis.* **169:** 248–252.

**Hu, X. L., Ray, R. and Compans, R. W. (1992).** Functional interactions between the fusion protein and haemagglutinin-neuraminidase of human parainfluenza viruses. *J. Virol*. **66:** 1528–1534.

**Hudson R.R. and Kaplan N.L. (1995).** Deleterious background selection with recombination. *Genetics,* **141:** 1605–17.

**Hughes A.L., Ota T. and Nei M. (1990).** Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility complex molecules. *Mol Biol Evol*.**7:** 515–524.

**Hull, R. N., Cherry, W. R. and Johnson, I. S. (1956).** The adaption and maintenance of mammalian cells to continuous growth in tissue culture. *Anat. Rec.* **124:** 490.

**Hull, R. N., Cherry, W. R. and Tritch, O. J. (1962).** Growth characteristics of monkey kidney cell strains LLC-MK1, LLC-MK2, and LLC-MK2(NCTC-3196) and their utility in virus research. *J Exp Med.* **115:** 903-18.

**Iwane, M. K., Edwards K. M., Szilagyi P. G., Walker F. J., Griffin M. R.,. Weinberg G. A, Coulen C., Poehling K. A., Shone L. P., Balter S., Hall C. B., Erdman D. D., Wooten K. and Schwartz B. (2004).** Population-based surveillance for hospitalizations associated with respiratory syncytial virus, influenza virus, and parainfluenza viruses among young children. *Pediatrics,* **113**:1758-1764.

**Jenkins, G.M. and Holmes, E.C. (2003).** The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res*. **92:** 1–7.

**Karron, R.A. and Collins P.L. (2007).** Parainfluenza Viruses. In: Knipe, DM.; Howley, PM., editors. *Fields Virology*, **5:** 1497-1526.

**Katsumi M., Mika S., Miho K., Hiroyuki T., Yoko A., Tatsuya I., Chieko A., Noriko K., Tsutomu I., Masahiro N., Kunihisa K., Tadayuki A. and Hirokazu K. (2011**). Detailed genetic analysis of haemagglutinin-neuraminidase glycoprotein gene in human parainfluenza virus type 1 isolates from patients with acute respiratory infection between 2002 and 2009 in Yamagata prefecture, Japan. *Virol J.* **8:** 533.

**Keller I., Bensasson D. and Nichols R. A. (2007).** Transition-transversion bias is not universal: A counter example from grasshopper pseudogenes. *PLoS Genet*. **3:** 22.

**Kimura, M. (1983).** The Neutral Theory of Molecular Evolution, Cambridge University Press10.1371. *Journal.pgen. /0030022*

**Konovalova, N. G., Blyumkin, V. N. and Zakstelskaya, L. Y. (1967).** Action of parainfluenza viruses on the mitotic activity of primary trysinized human embryo kidney cell cultures. *Acta Virol*. **11:** 564.

**Krunic, N., Yager, T. D., Himsworth, D., Merante, F., Yaghoubian, S. and Janeczko, R. (2007).** xTAG RVP assay: analytical and clinical performance.*J. Clin. Virol*. **40:** 39–S46.

**Kumar, S., Tamura, K. and Nei, M. (2004).** Mega3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform*. **5:** 150–163.

**Landry, M.L. (2007).** Rhinoviruses,. *In* P. R. Murray, E. J. Baron, J. H. Jorgensen, M. L. Landry, and M. A. Pfaller (ed.), *Manual of clinical microbiology*, **9:** 1405–1413.

**Landry, M. L. and Ferguson, D. (2000).** SimulFluor respiratory screen for rapid detection of multiple respiratory viruses in clinical specimens by immunofluorescence staining. *J. Clin. Microbiol.* **38:** 708–711.

**Leland, D. S. and Ginocchio, C. C. (2007).** Role of cell culture for virus detection in the age of technology. *Clin. Microbiol. Rev*. **20:** 49–78.

**Lewontin, R. C. (1964).** The interaction of selection and linkage general considerations; heterotic models. *Genetics*, **49:** 49–67.

**Lyn, D., Gill, D.S, Scroggs, R.A. and Portner, A. (1991).** The nucleoproteins of human parainfluenza virus type 1 and Sendai virus share amino acid sequences and antigenic and structural determinants. *J Gen Virol.***72:** 983–7.

**Madeley, C.R. and Peiris, J.S. (2002).** Methods in virus diagnosis: immunofluorescence revisited. *J Clin Virol*. **25:** 121-134.

**Maynard, J. and Haigh, J. (1974).** The hitch-hiking effect of a favorable gene. *Genet. Res.* **23:** 23–35.

**Mikita S., David T. and Peer B. (2006).** PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. **34:** 609-612.

**Mizuta K., Saitoh M., Kobayashi M., Tsukagoshi H., Aoki Y., Ikeda T., Abiko C., Katsushima N., Itagaki T., Noda M., Kozawa K., Ahiko T. and Kimura H. (2011)**. Detailed genetic analysis of haemagglutinin neuraminidase glycoprotein gene in human parainfluenza virus type 1 isolates from patients with acute respiratory infection between 2002 and 2009 in Yamagata prefecture, Japan. *Virology Journal,* **8:**533.

**Mooers, A.O. and Holmes, E.C. (2000).** The evolution of base composition and phylogenetic inference. *Trends Ecol. Evol.* **15:** 365–369.

**Moritz C., Dowling T.E. and Brown W.M. (1987).** Evolution of animal mitochondrial DNA: Relevance for population biology and systematics.Annu Rev Ecol Syst **18:**269.

**Mounts, A., Kaur, H., Parshar, U.D., Ksiazek, T.G., Cannon, D., Arokiasamy, J.T., Anderson, L.J. and Lye, M.S. (2001).** A cohort study of health care workers to assess nosocomial transmissibility of Nipah Virus, Malaysia,1999. *J.Infect. Dis.* **183:** 810-813.

**Murphy, B.R. (1988).** Current approaches to the development of vaccines effective against parainfluenza viruses [review]. *Bull. World Health Organ.* **66:** 391–397.

**Murphy B.R., Richman D.D., Chalhub E.G., Uhlendorf C.P., Baron S. and Chanock R.M. (1975).** *Infect. Immun.* **12:** 62-68.

**Nachman, M.W. (2002)**. Variation in recombination rate across the genome: evidence and implications. *Curr Opin Genet Dev.* **12:** 657-63

**Nei, M. (1987).** *Molecular Evolutionary Genetics.* New York: Columbia Univ. Press.

**Nei M. and Gojobori T. (1986).** Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* **3:** 418-26.

**Niesters, H.G.M. (2002).** Clinical virology in real time. *J. Clin. Virol.* **25:** 3–12.

**Noyola, D. E., Clark B., O'donnell F., T., Atmar, R. L., Greer, J. and Demmler, G. J. (2000).** Comparison of a New Neuraminidase Detection Assay with an Enzyme Immunoassay, Immunofluorescence, and Culture for Rapid Detection of Influenza A and B Viruses in Nasal Wash Specimens. *J. Clin Micro*. **38:** 1161–1165.

**Ochieng, W. O., Hazlett, D. T. G., Bell, T. M., Tukei, P. M., Ademba, G. R., Magana, J. M., Gathara, G. W., Wafula, E. M., Pamba, A., Ndinya-Achola, J. O. and Arap Siongoki, T.K. (1988).**Viral etiology and epidemiology of acute respirstory unfections in children in Naiobi, `Kenya *Am. J. Trop. Med. Hyg.* **39:** 632-640

**Palese, P., Tobita, K., Ueda, M. and Compans, R. W. (1974).** Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology,* **61:** 397–410.

**Parsch, J., Meiklejohn, C.D. and Hartl, D.L.** (**2001).** Patterns of DNA sequence variation suggest the recent action of positive selection in the janus-ocnus region of *Drosophila* simulans.*Genetics.***159:** 647-57.

**Paulson, J. C., Weinstein, J., Dorland, L., Van Halbeek, H. and Vliegenthart, J. F. G. (1982).** Newcastle disease virus contains a linkage-specific glycoprotein sialidase. *J. Biol. Chem*. **257:** 12734–12738.

**Plumper, R.K., Lakdawala, A.S., Gernert, K.M., Snyder, J.P. and Compans, R.W. (2003).** Structural features of paramyxovirus F protein required for fusion initiation. *Biochemistry,* **42:** 6645–6655.

**Pluzhnikov, A. and Donnelly. P. (1996).** Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics,* **144:**1247–1262.

**Porotto, M., Greengard, O., Poltoratskaia, N. Horga, M.A. and Moscona, A. (2001).** Human parainfluenza virus type 3 HN-receptor interaction: the effect of 4-GU-DANA on a neuraminidase-deficient variant. *J. Virology.* **76:** 7481–7488.

**Rahman, M., Vandermause, M.F. and Kieke, B.A. (2008).** Performance of Binax NOW Flu A and B and direct fluorescent assay in comparison with a composite of viral culture or reverse transcription polymerase chain

reaction for detection of influenza infection during the 2006 to 2007 season. *Diagn Microbiol Infect Dis*. **6:**162–6.

**Rozas J., Sánchez-DelBarrio J.C., Messeguer X. and Rozas R.(2003).** DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics.***19:** 2496-7.

**Russell C.J. and Luque L.E. (2006).** The structural basis of paramyxovirus invasion. *Trends Microbiol.* **14:** 243-6.

**Sambrook, J., Frisch, E.F. and Maniatis, T. (1989).** Molecular cloning: A Laboratory Manual, 2nd Ed. Oxford University Press. London.

**Sanjuán R., Nebot M.R., Chirico N., Mansky L.M. and Belshaw R. (2010)**.Viral mutation rates. *J Virol*. **84:** 9733-9748.

**Schindera, C., Kraemer, A.L., Regamey, N., Aebi, C., Gorgievski-Hrisoho, M. and Barbani, M.T. (2010).** Immunofluorescence versus xTAG multiplex PCR for the detection of respiratory picornavirus infections in children. *J Clin Virol*. **48:** 223-225.

**Sealy, R., Surman, S., Hurwitz, J.L. and Coleclough, C. (2003).** Antibody response to influenza infection of mice: different patterns for glycoprotein and nucleocapsid antigens. *Immunology,* **108:**431–9.

**Sharp, P.A., Sugden, B. and Sambrook, J. (1973).** Detection of two restriction Endonuclease activities in Haemophilus parainfluenza using analytical agarose. *Biochemistry*, **12:** 3055-3057.

**Skiadopoulos, M.H., Tatem, J.M., Surman, S.R, Mitcho, Y.,Wu, S. L. and Elkins, W.R. (2002)**. The recombinant chimeric human parainfluenza virus type 1 vaccine candidate, rHPIV3-1cp45, is attenuated, immunogenic, and protective in African green monkeys. *Vaccine*, **20:**1846–52.

**Slobod, K.S., Leggiadro, R.J., Presbury, G., Smith, F.S. and Hurwitz, J.L. (1994).** Peripheral T cell receptor repertoire among CD4+ and CD8+ subsets during acute infectious mononucleosis. *Viral Immunol*. **7:**151–3.

**Smith, F.S., Portner A., Leggiadro, R.J., Turner, E.V. and Hurwitz, J.L. (1994).** Age-related development of human memory T-helper and B-cell responses toward parainfluenza virus type-1. *Virology*, **205:** 453–61.

**Storey, D. G., Dimock, K. and Kang, C. Y. (1984).** Structural characterization of virion proteins and genomic RNA of human parainfluenza virus 3. *J. Virol*. **52:** 761–766.

**Stephanie, M.P., David, L.W., Shauna, A.T., Chadi, E.S., Lisa, M.H., Jody, A.A., Anami, P. and John, P.V. (2005).** Comparison of a Real-Time Reverse Transcriptase PCR Assay and a Culture Technique for Quantitative Assessment of Viral Load in Children Naturally Infected with Respiratory Syncytial Virus. *Journal of Clinical Microbiology*, **43:** 2356-2362.

**Tang, L.F., Wang, T.L., Tang, H.F. and Chen, Z.M. (2008).** Viral pathogens of acute lower respiratory tract infection in China. *Indian Pediatr*. **45:** 971-975.

**Tang, R.S., Spaete, R.R., Thompson, M.W., MacPhail, M., Guzzetta, J.M. and Ryan, P.C. (2008).** Development of a PIV-vectored RSV vaccine: preclinical evaluation of safety, toxicity, and enhanced disease and initial clinical testing in healthy adults. *Vaccine*, **26:** 6373–82.

**Tajima F. (1993).** Simple Methods for Testing the Molecular Evolutionary Clock Hypothesis. *Genetics*, **135:** 599–607.

**Templeton, K. E., Scheltinga, S. A., Beersma, M. R. C., Kroes, A. C. M. and Claas, E. C. J. (2004).** Rapid and sensitive method using multiplex real-time PCR for diagnosis of infections by influenza A and influenza B viruses, respiratory syncytial virus, and parainfluenza viruses 1, 2, 3, and 4. *J. Clin. Microbiol*. **42:** 1564–1569.

**Terlizzi, M.E., Massimiliano**, **B., Sidoti F., Sinesi, F., Rosangela, V., Stefano**, **G., Costa, C. and Rossana, C. (2009).** Quantitative RT Real Time PCR and indirect immunofluorescence for the detection of human parainfluenza virus 1, 2, 3. *Journal of Virological Methods*,**160:** 172–177.

**Tian D., Zhang W., He J., Liu Y. and Song Z. (2012).** Novel, Real-Time Cell Analysis for Measuring Viral Cytopathogenesis and the Efficacy of Neutralizing Antibodies to the 2009 Influenza A (H1N1) *Virus. PLoS ONE*, **7:** 31965.

**Tyrrell, D. A. and Bynoe, M. L. (1969).** Studies on parainfluenza type 2 and 4 viruses obtained from patients with common colds. *Br. Med. J*. **1:** 471–474.

**Van der Zalm, M. M., Van Ewijk, B.E., Wilbrink, B., Uiterwaal, C.S., Wolfs, T.F. and Van der Ent C.K. (2009)**. Respiratory pathogens in children with and without respiratory symptoms. *J Pediatr*. **154:** 396-400.

**Van Elden, L.J., Van Kraaij, M.G., Nijhuis, M., Hendriksen, K.A., Dekker, A.W., Rozenberg-Arska, M. and Van Loon, A.M. (2002).** Polymerase chain reaction is more sensitive than viral culture and antigen testing for the detection of respiratory viruses in adults with hematological cancer and pneumonia. *Clin. Infect. Dis*. **34:** 177–183.

**Van Gageldonk-Lafeber A.B., Heijnen M.L., Bartelds A.I., Peters M.F., Van der Plas S.M. and Wilbrink B. (2005).** A case-control study of acute respiratory tract infection in general practice patients in The Netherlands. *Clin Infect Dis*. **41:** 490-497.

**Van Kraaij, M. G. J., Van Elden, L. J. R., Van Loon A. M., Hendricksen, K. A. W., Laterveer, L., Dekker, A. W. and Nijhuis, M. (2005).** Frequent detection of respiratory viruses in adult recipients of stem cell transplants with the use of real-time polymerase chain reaction, compared with viral culture. *Clin. Infect. Dis*. **40:** 662–669.

**Verrelli, B.C. and Eanes, W.F. (2001).** Clinal variation for amino acid polymorphisms at the Pgm locus in *Drosophila melanogaster*. *Genetics,* **157:** 1649-63.

**Visseren, F. L. J., Bouwman, J. J. M., Bouter, K. P., Diepersloot, R. J. A., Groot, G. D. and Erkelens, D. W. (2000).** Procoagulant activity of

endothelial cells after infection with respiratory viruses. *Thromb. Haemostasis*, **84:** 319–324.

**Von Itzstein, M., Wu, W. Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Van Phan, T., Smythe, M. L., White, H. F., Oliver, S. W., Colman, P. M., Varghese, J. N., Ryan, D. M., Woods, J. M., Bethell, R. C., Hotham, V. J., Cameron, J. M. and Penn, C. R. (1993).** Rational design of potent sialidase-based inhibitors of influenza virus protection. *Nature*, **363:** 418–423.

**Wan, C., Huang Y., Cheng, L., Fu, G., Shi S.H., Chen, H., Deng C., Lin F. and Lin, J. (2011).** The development of a rapid SYBR Green I-based quantitative PCR for detection of Duck Circovirus. *Virol. J.* **8:** 465.

**Wang, W., Thornton, K., Berry, A. and Long, M. (2002).** Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science*, **295:** 134-7.

**World Health Organization (1981).** Clinical management of acute respiratory infections in children: a WHO memorandum. *Bull WHO*, **59:** 707-767.

**Washburne, J. F., Bocchini, J. A. and Jamison, R. M. (1992).** Summertime respiratory syncytial virus infection: Epidemiology and clinical manifestations. *South. Med. J.* **85:** 579–583.

**Wilson, A. C., Carlson, S. S. and White, T. J. (1977).** Biochemical evolution. *Annu. Rev. Biochem.* **46:** 573–639.

**Yang, Z. and Anne, Y.D. (1999).** Estimation of the Transition/ Transversion Rate Bias and species sampling. *J. Mol.* **48:** 274-283.

**Yang, Z. and Bielawski J. P. (2000).** Statistical methods for detecting molecular Adaptation. *Trends in Ecology and Evolution,* **12:** 496-503.

## APPENDICES

## APPENDIX I

**INFLUENZA QUESTIONNAIRE**

Volunteer ID number (cryovial number) _____

Date of collection (DD/MM/YY) ____/____/____ Hospital _____

Patient information: Date of birth (if known) (DD/MM/YY) ____/____/____

Otherwise: Age: Years ____ Months____ (if less than 1 year)

Gender: ☐ Male ☐ Female

Residence (city or village): _____District:_____ Province:_____

Specify neighbourhood of residence: _____

During the past five days, has the volunteer been mostly in (check one)?

☐ City or village of Residence

☐ In the country, but not in city of residence, Where? _____

☐ Out of the country, Where? _____

Temperature at presentation of patient to the outpatient clinic: _____ °C

Method: ☐ Rectal ☐ Axillary ☐ Oral

Type of swab obtained: ☐ nasopharyngeal ☐ oropharyngeal

For the symptoms, check "Y" if the patient has had the symptom in association with the present illness, "N" if they have not, or "U" if uncertain.

| | Y | N | U | | Y | N | U |
|---|---|---|---|---|---|---|---|
| | ☐ | ☐ | ☐ | | ☐ | ☐ | ☐ |

| | Y | N | U | | Y | N | U |
|---|---|---|---|---|---|---|---|
| Cough | ☐ | ☐ | ☐ | Abdominal pain | | | |
| Difficulty breathing | ☐ | ☐ | ☐ | Nasal stuffiness | ☐ | ☐ | ☐ |
| Chills | ☐ | ☐ | ☐ | Runny nose | ☐ | ☐ | ☐ |
| U | ☐ | ☐ | ☐ | | ☐ | ☐ | ☐ |
| Sore throat | | | | Sputum production | | | |
| Muscle aches | ☐ | ☐ | ☐ | Headache | ☐ | ☐ | ☐ |
| Retroorbital pain | ☐ | ☐ | ☐ | Joint pain | ☐ | ☐ | ☐ |
| Malaise | ☐ | ☐ | ☐ | Fatigue | ☐ | ☐ | ☐ |
| Vomiting | ☐ | ☐ | ☐ | Diarrhea | ☐ | ☐ | ☐ |
| Neurological | ☐ | ☐ | ☐ | Bleeding | ☐ | ☐ | ☐ |

Manifestations: Specify:

Other significant clinical findings or comments:

## APPENDIX II

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: kemri-hq@nairobi.mimcom.net; director @ kemri. org; website: www.kemri.org

ESACIPAC/SSC/4772

20th July, 2009

Kiptinness Joshua Kimutai

Thro'
Director, CVR
NAIROBI

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

REF: **SSC No. 1654 (Revised) – Isolation and molecular characterization of human parainfluenza virus type 1 in infants attending Mbagathi District Hospital, Nairobi, Kenya**

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 **158th meeting** held on **30th June, 2009** and has since been approved for implementation by the SSC.

The SSC however, advises that work on this project can only start when ERC approval is received.

Sammy M. Njenga PhD
Ag. SSC SECRETARY

CENTRE FOR VIRUS RESEARCH
23 JUL 2009
P. O. Box 54628 - NAIROBI

*In Search of Better Health*

## APPENDIX III


**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 November 11, 2010,

TO: KIPTINNESS JOSHUA KIMUTAI, PRINCIPAL INVESTIGATOR

THRO': DR. FREDERICK OKOTH, THE DIRECTOR, CVR NAIROBI

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

RE: SSC 1654 (*REQUEST FOR ANNUAL RENEWAL*): ISOLATION AND MOLECULAR CHARACTERIZATION OF HUMAN PARA-INFLUENZA VIRUS TYPE-1 IN INFANTS ATTENDING MBAGATHI DISTRICT HOSPITAL, NAIROBI, KENYA.

This is to inform you that during the 184th meeting of the KEMRI/ERC meeting held on 9th November 2010, the above study was reviewed.

Thank you for your Continuing Review Report for the period October 2009 to October 2010.

The Committee is satisfied that sufficient progress has been made in the review period, and therefore grants the study **approval** to determine the nucleotide sequences and phylogenetic relationship of the HPIV-1 strains.

Please note that authorization to conduct this study will automatically expire on **10th November 2011.** 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 **15th September 2011**.

You are required to submit any amendments to this protocol and other information pertinent to human participation in this study to the SSC and ERC for review prior to initiation.

Yours sincerely,

R.C.Kithinji

R. C. KITHINJI,
FOR: SECRETARY,
KEMRI/NATIONAL ETHICS REVIEW COMMITTEE

CENTRE FOR VIRUS RESEARCH
15 NOV 2010
P.O. Box 54628 - NAIROBI

## APPENDIX IV

**INFORMED CONSENT AGREEMENT**

**TITLE OF STUDY:**
Influenza Surveillance in Kenya

**INSTITUTIONS:** United States Army Medical Research Unit-Kenya (USAMRU-K), Kenya Medical Research Institute (KEMRI) and Centers for Disease Control- International Emerging Infections Program- Kenya (CDC-IEIP).

**PRINCIPAL INVESTIGATORS:** David Schnabel, MD MPH; Walter Ochieng, MD MPH; Wallace Bulimo PhD;Rober Breiman, MD; Kariuki Njenga, PhD; Evans Amukoye MBCHB M.Med.

**INTRODUCTION:**This is a research study to find out what viruses, if any, cause influenza like illness in Kenya. Influenza-like illnesses can cause you to have fever, sore throat, runny or stuffy nose, cough, headache, and muscle ache. We would also like to get our consent to store the specimen for future research about respiratory viruses.

**PARTICIPATION:** Your participation and sample donation in this study is voluntary. If you do not want to donate a specimen, there will be no penalty. You (your child) will receive the same health care if you do not choose to participate. You (your child) may stop your (your child's) participation at any time without penalty or loss of benefits. A copy of the consnent form will be provided to you/ your child.

**PROCEDURES TO BE FOLLOWED:**

**QUESTIONNAIRE:** If you agree to participate, or act as a legal representative for an adult or child to participate in the survey, we will ask you (your child) some questions about your (your child's) age, sex, job, where you live, history of illness with fever and symptoms.

**LAB TEST:** Two nasal (throat) swabs will be taken from you (your child) and tested for germs that cause the flu. The results of all testing performed will be shared with the medical or clinical officer caring for you (your child).

**RISKS:** The nasal (throat) swabs may be uncomfortable. The swabs will be taken by trained medical personnel who will try to minimize this discomfort.

**BENEFITS:** There is no direct benefit to you (your child) aside from the satisfaction that your (your child's) participation may help to better understand this type of disease in Kenya in the future.

**COMPENSATION:** You will not be paid for your (your child's) participation.

**DURATION OF PARTICIPATION:** This study only requires two nasal (or throat) swabs and the questionnaire. There is no follow-up or further information needed.

**WHO CAN PARTICIPATE IN THIS STUDY:** Any person ages 2 months and above who has been sick for no more than 3 days with a fever at least 38°C and influenza-like illness and has been in an outpatient clinic can participate if they agree. A maximum of 5 volunteers per day will be enrolled per study site. If the doctor finds an obvious source of
infection causing the fever, like pneumonia, then you (your child) should not participate.

**ASSURANCE OF CONFIDENTIALITY OF VOLUNTEER'S IDENTITY:** Records relating to you (your child's) participation in the study will remain private. You (your child's) name will not be used in any report resulting from this study. All files and laboratory specimens will have only a unique study number, not your (your child's) name.

**USE OF SAMPLES:** Your (your child's) samples may be kept when testing is completed. You will not be contacted about future use of your donated specimen, but USAMRU-K and KEMRI will obtain approval from their ethical review committees before using your specimen. This is so that they can be tested for germs we may discover in the future.

**REVIEW OF RECORDS:**Interview and consent forms will be kept in a locked file at KEMRI or a designated storage facility for 10 years following completion of the study. These data sheets will be made available only to the Principal Investigators, clinical personnel who require this information for legal reasons or to study an outbreak.

**MEDICAL CARE FOR RELATED INJURY:** Should you (your child) be injured as a direct result of participating in this project, you (your child) will be provided medical care, at no cost to you (your child), for that injury. You (your child) will not receive injury compensation, only medical care. You (your child) should also understand that this is not a waiver or release of your (your child's) legal rights. You should discuss this issue thoroughly with the principal investigator before you (your child) enroll in this study.

**PERSONS AND PLACES FOR ANSWERS IN THE EVENT OF RESEARCH RELATED INJURY:** If you think you (your child) has a medical problem related to this study, please report to Dr. David Schnabel, Medical Research Unit, Box 606 Village Market, Nairobi, Phone: 254-20-2729303.

**PERSONS AND PLACES FOR ANSWERS REGARDING YOUR RIGHTS AS A RESEARCH SUBJECT:** If during the course of this study, you have questions about the nature of the research or you believe you (your child) have/ has sustained a research-related injury, you should contact Dr. Schnabel at P.O Box 606 Village Market, Nairobi, Kenya, tel 254-20-2729303, or the doctor on duty at the hospital. The Secretary of the Kenya National Ethical Review Committee, c/ o Kenya Medical Research Institute, P.O. Box 54840, Nairobi, Kenya, tel.254-20-2722541 of Deputy for Regulatory Compliance and Quality, Human Subject Protection, MRMC-RCQ-HR, 504 Scott St., Fort Detrick, MD 21702, USA, Phone: 0001-301-619-2165 can also be contacted concerning you (your child's) rights as a research volunteer.

I, _____ (Name) having full capacity to consent for myself, an adult, or a child named: _____ (Subject's name)and having attained my birthday, do hereby consent to my/ his/ her participation in the epidemiologic study:

"Influenza Surveillance in Kenya" under the direction of Dr. David Schnabel, Dr. Walter Ochieng, Dr. Wallace Bulimo, Dr. Robert Breiman, Dr. Kariuki Njenga, and Dr. Evans Amukoye.

The methods and means by which the study will be conducted and the risks which may have been reasonable expected have been explained to me by _____.

I have been given the opportunity to ask questions concerning this investigational study, and any such questions have been fully answered to my full and complete satisfaction.

You may at any time during the course of the study withdraw this consent and remove yourself/ your child from the study without prejudice.

Subject's or Guardian's Signature: _____ Date:_____

Permanent Address: _____

Witness's Name: _____

Witness's Signature:_____ _____Date:

Study Number:

## APPENDIX V

**CONSENT FORM AGREEMENT FOR STORAGE AND SHIPPING OF SAMPLES**

The tests that are part of this study may require shipping your (your child's) sample and in some cases the viruses isolated from them, outside of Kenya. The purpose of sending some samples and/ or isolates outside of Kenya is to confirm the initial results with a laboratory that specializes in influenza and other respiratory diseases. If the sample is positive for influenza more information can be learned about what kind of influenza it is and this information may contribute to the annual vaccine. Once testing is complete, your (your child's) samples will be stored indefinitely at the KEMRI laboratories. A code will be used to identify the stored samples, not your name. Also, your (your child's) samples may be used in future research studies on respiratory diseases. You (your child) will not receive payment for any future value your samples may be found to have. You will not receive any notice of future uses of your (child's) samples. Any future research involving your (your child's) samples will be reviewed by an ethical review committee to ensure the use is confidential, ethical and meets all government guidelines. You (your child) do/ does not have to donate your (your child's) samples for future research. You (your child's) samples or agree to have your samples shipped outside of Kenya. Please check one of the two options and sign below:

_____ I do not want my samples/ virus isolates shipped outside Kenya or stored for future use.

_____ I agree to give up ownership and rights to the samples/ virus isolates that I am donating. I understand that my samples may be shipped outside of Kenya and stored for future use.

_____
Name of subject / guardian

_____ _____
Signature of subject/ guardian Date

_____
Name of Witness

_____ _____
Signature of witness Date

## APPENDIX VI

**Preparation of media and reagents**

**a**) **DMEM maintenance media + trypsin (for MDCK cells)**

1. To a 500ml bottle of DMEM add:

(i) 5.4 ml of L-glutamine (2mM final conc.)

(ii) 5.4 ml of P/S (1X " )

(iii) 13.5 ml of BSA (0.2% " )

(iv) 13.5 ml of HEPES (25mM " )

(vi) 0.44 ml gentamicin (40μg/ml)

2. Label as "+ BSA + P/S/G + HEPES + Gln", with your name, today's date, and an expiration date of 4 weeks from the date prepared. Store at 4°C.

3. To 100 ml of the above solution add:

(i)0.2 ml of trypsin (4μg/ml final conc.)

4. Label as "LLCMK2 maintenance media + trypsin", with your name and today's date.

Make this solution fresh, daily.

**b) PBS reagent**

| | |
|---|---|
| Sodium chloride | - 81% |
| Sodium phosphate dibasic | - 14% |
| Potassium phosphate monobasic | - 3% |
| Potassium chloride | - 2% |

Dissolve in 1000ml of sterile distilled water

## APPENDIX VII

### HPIV-1 CPE onset

| **Sample ID** | 1 | 2 | D 3 | C A 4 | P Y 5 | E S 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| MBG/07/06/01 | - | - | - | - | - | - | - | + | | |
| MBG/07/07/02 | - | - | - | - | - | - | + | | | |
| MBG/07/08/03 | - | - | - | - | - | - | + | | | |
| MBG/08/03/04 | - | - | - | - | - | - | - | - | - | - |
| MBG/08/03/05 | - | - | - | - | - | - | - | - | - | - |
| MBG/08/04/06 | - | - | - | - | - | - | - | - | - | - |
| MBG/08/04/07 | - | - | - | - | - | - | - | - | - | - |
| MBG/08/04/08 | - | - | - | - | - | - | - | - | - | - |
| MBG/08/04/09 | - | - | - | - | - | - | - | - | + | |
| MBG/08/04/10 | - | - | - | - | - | - | - | + | | |
| MBG/08/04/11 | - | - | - | - | - | - | - | - | - | - |
| MBG/08/06/12 | - | - | - | - | - | - | + | | | |
| MBG/08/09/13 | - | - | - | - | - | - | - | + | | |
| MBG/08/10/14 | - | - | - | - | - | - | - | - | - | + |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| MBG/08/10/15 | - | - | - | - | - | - | - | - | + | |
| MBG/08/10/16 | - | - | - | - | - | - | + | | | |
| MBG/08/10/17 | - | - | - | - | - | - | + | | | |
| MBG/08/11/18 | - | - | - | - | - | - | - | - | - | - |
| MBG/09/09/19 | - | - | - | + | | | | | | |
| MBG/09/12/20 | - | - | - | - | - | + | | | | |
| MBG/10/01/21 | - | - | - | - | - | - | - | - | - | - |
| MBG/10/01/22 | - | - | - | - | - | - | - | - | + | |
| MBG/10/02/23 | - | - | - | - | - | - | + | | | |
| MBG/10/03/24 | - | - | - | - | - | - | - | - | - | - |
| MBG/10/03/25 | - | - | - | - | - | - | - | - | - | - |
| C35 | - | - | - | - | - | - | + | | | |
| Maintenance media | - | - | - | - | - | - | - | - | - | - |

## APPENDIX VIII

## FASTA Nucleotide sequences of the HPIV-1 isolates

HPIV-1_HN_M86786.1

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA GACAGGAATT
GGCTCAGATA TGCGAAAACA CCATTGCTAT
81 TCACCATGCA GACGGCATAT CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCCACC
ACAATTTCAG GATGTGTTAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCG

MBG/07/06/01

1 ACCAGAGATC GCACACAATT AATAAAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TCCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGAAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAC
GTCCCCTACG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGAACTACC
ACAATTTCAG GATGTGTCAC GATACCTTTA
241 TTATTAATTG GGGACGCATT ATATGCA

MBG/07/07/02

1 AGCAGATATC TCACACAATT AATAGAGAAG TCATGCCACA GACCGGAATT
GGCTCATATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCTATAT CTCATTATTA CCTGGGCCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTAAG GATACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

MBG/07/08/03

1 AGCAAAGATC TCACACAATT AATAAAGAAG TCATGCAACA AACACGAATT
GGCTCAGATA GACGAGAACA CCACTGCTAT
81 TCACCATGCA AACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTACG AGAACCCCTA CTGAGCAACA

161 ACCACAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGATAG CATACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

C35

1 AGCAGAGATC TTACACAATT AATAGAGAAG TCATGCAACA GACAGGAATT
GGCTCAGATA TGCGAAAACA CCATTGCTAT
81 TCACCATGCA GACGGTATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCAACC
ACAATCTCAG GATGTGTTAG ACTACCTTCA
241 TTATCAATTG GTGATGAAAT ATATGCA

MBG/08/04/05

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACACGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA AACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGAACG AGAACCCCTA CTGAGCAACA
161 ACCCCAACAT CTCATTATTA CCGGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGCGATAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

MBG/08/04/06

1 AGCAAAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACACGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA AACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGAACG AGAACCCCTA CTGAGCAACA
161 ACCCCAACAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTTAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

HPIV-1_HN_U70938.1

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGTGAAAACA CTATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCTAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCCACC
ACAATTTCAG GATGTGTTAG ACTACCTTCA

241 TTATCAATTG GTGATGCAAT ATATGCG

MBG/08/04/08

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
AGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTA TGGATCTACC
ACAATTTCAG GATGTGTACG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

HPIV-1_HN_U86785.1

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGTGAAAACA CTATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCCACC
ACAATTTCAG GATGTGTTAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCG

MBG/08/04/09

1 AGCAGCGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGACAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTAAG CATACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

MBG/08/06/10

1 AGCAGAGATC TCACACAATT AATAAAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTTAG CATACCTTCA
241 TTATCAATTG GTGAGTGCAT ATATGCA

MBG/08/09/11

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCAATGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTTAG CATACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCC

MBG/08/10/12

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTAAG CATACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

MBG/08/10/13

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CTCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTTAG CATACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

MBG/08/10/14

1 AGCAGAGATC TCACACAATT AATAAAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTTAG CATACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

MBG/08/10/15

1 AGCAGAGATC CCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTTAG CATACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

MBG/08/11/16

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTGGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTTAG CATACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

HPIV-1_HN_U70936.1

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CTATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCTAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCCACC
ACAATTTCAG GATGTGTTAG ATTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCG

MBG/09/09/17

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTCAG ATTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCG

MBG/09/12/18

1 AGCAGAGATC TCACACAATT AATAAAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTGAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

HPIV-1_HN_U70937.1

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CTATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCTAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCCACC
ACAATTTCAG GATGTGTTAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCG

HPIV-1_HN_U70947.1

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CTATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCTAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCCACC
ACAATTTCAG GATGTGTTAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCG

MBG/10/01/19

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA AACGGAATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTTAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

MBG/10/02/20

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA AACGGAATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTTAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

MBG/10/03/21

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTTAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

MBG/10/03/22

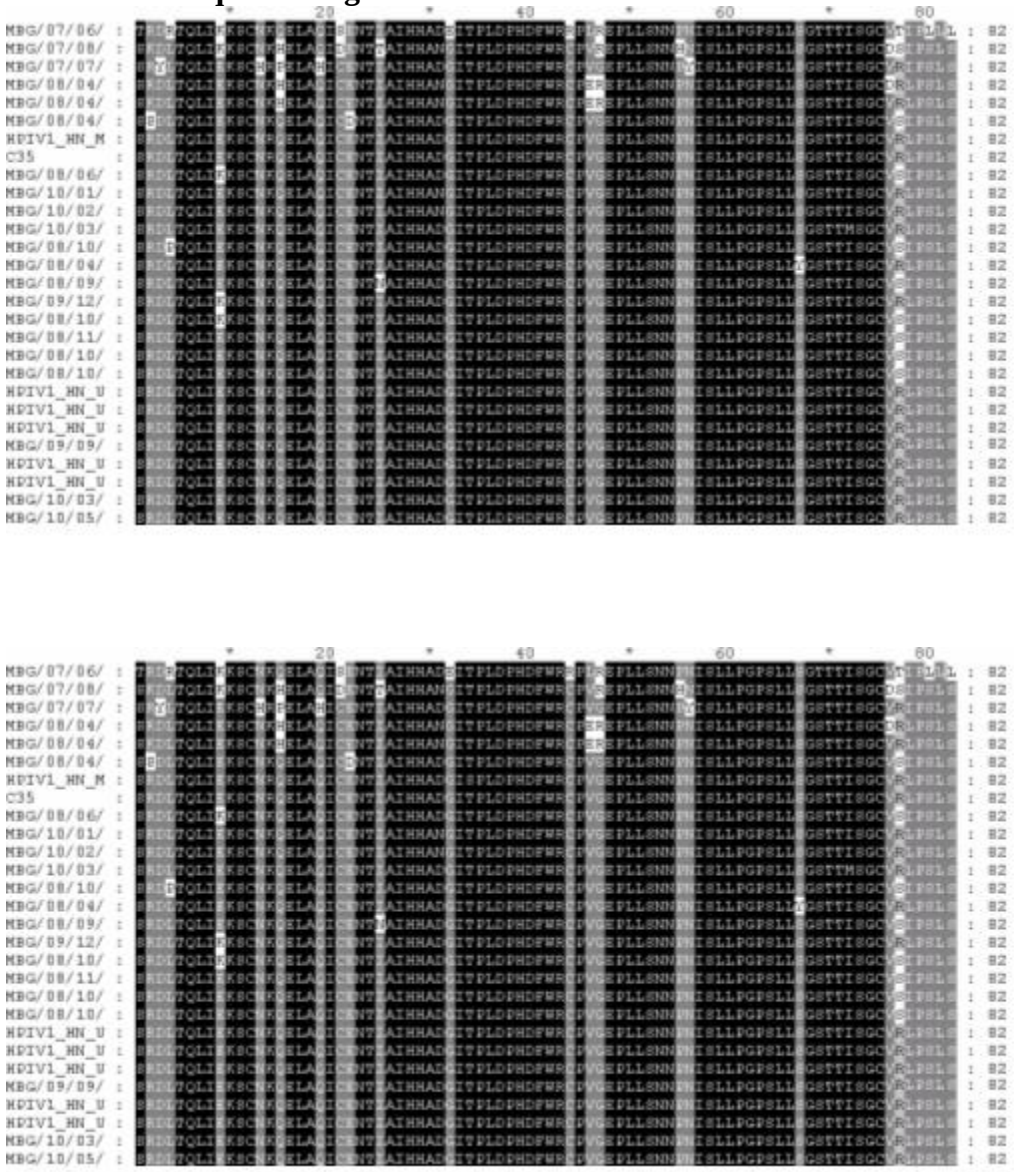
1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATGTCAG GATGTGTCAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

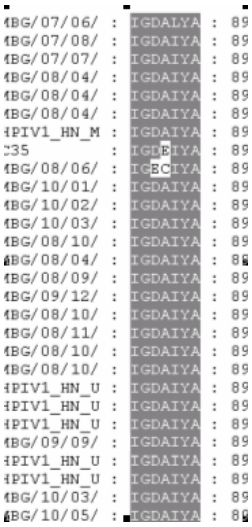
MBG/10/05/23

1 AGCAGAGATC TCACACAATT AATAGAGAAG TCATGCAACA AACAGGAATT
GGCTCAGATA TGCGAGAACA CCATTGCTAT
81 TCACCATGCA GACGGAATAA CCCCTCTAGA CCCACATGAT TTCTGGAGAT
GTCCCGTAGG AGAACCCCTA CTGAGCAACA
161 ACCCCAATAT CTCATTATTA CCTGGACCAA GTCTACTTTC TGGATCTACC
ACAATTTCAG GATGTGTCAG ACTACCTTCA
241 TTATCAATTG GTGATGCAAT ATATGCA

## APPENDIX IX

### Amino acids sequence alignment




IGDALYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGEEIYA : 89
IGECIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89
IGDAIYA : 89

## APPENDIX X

### Highest and lowest nucleotide sequence similarities

**Similarities between MBG 08/04/08 and MBG 10/03/22**

>lcl|7315
Length=267

Score = 460 bits (249), Expect = 1e-134
Identities = 261/267 (98%), Gaps = 0/267 (0%)
Strand=Plus/Plus

```
Query  1
AGCAGAGATCTCACACAATTAATAGAGAAGTCATGCAACAAACAGGA
ATTAGCTCAGATA  60
       |||||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct  1
AGCAGAGATCTCACACAATTAATAGAGAAGTCATGCAACAAACAGGA
ATTGGCTCAGATA  60

Query  61
TGCGAGAACACCATTGCTATTCACCATGCAGACGGAATAACCCCTCTG
GACCCACATGAT  120
       |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  61
TGCGAGAACACCATTGCTATTCACCATGCAGACGGAATAACCCCTCTA
GACCCACATGAT  120

Query  121
TTCTGGAGATGTCCCGTAGGAGAACCCCTACTGAGCAACAACCCCAAT
ATCTCATTATTA  180
       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121
TTCTGGAGATGTCCCGTAGGAGAACCCCTACTGAGCAACAACCCCAAT
ATCTCATTATTA  180

Query  181
CCTGGACCAAGTCTACTTTATGGATCTACCACAATTTCAGGATGTGTAC
GACTACCTTCA  240
       |||||||||||||||| |||||||||||| ||||||||||  ||||||||||
```

Sbjct 181
CCTGGACCAAGTCTACTTTCTGGATCTACCACAATGTCAGGATGTGTCA
GACTACCTTCA 240

Query 241 TTATCAATTGGTGATGCAATATATGCA 267
||||||||||||||||||||||||||
Sbjct 241 TTATCAATTGGTGATGCAATATATGCA 267

**Similarities between MBG 07/06/01 and MBG 10/03/21**
>lcl|55485
Length=267

Score = 396 bits (214), Expect = 3e-115
Identities = 249/266 (94%), Gaps = 2/266 (1%)
Strand=Plus/Plus

Query 3
CAGAGATCTCACACAATTAATAGAGAAGTCATGCAACAAACAGGAATT
GGCTCAGATATG 62
||||||| |||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct 3
CAGAGATCGCACACAATTAATAAAGAAGTCATGCAACAAACAGGAATT
GGCTCAGATATC 62

Query 63
CGAGAACACCATTGCTATTCACCATGCAGACGGAATAACCCCTCTAGA
CCCACATGATTT 122
|||||||||||||||||||||||||||| |||||||||||| ||||||||||||
Sbjct 63
CGAGAACACCATTGCTATTCACCATGCAGACGAAATAACCCCTCTGGA
CCCACATGATTT 122

Query 123
CTGGAGATGTCCCGTAGGAGAACCCCTACTGAGCAACAACCCCAATAT
CTCATTATTACC 182
||||||| ||||| || ||||||||||||||||||||||||||||||||||||||||
Sbjct 123
CTGGAGACGTCCCCTACGAGAACCCCTACTGAGCAACAACCCCAATAT
CTCATTATTACC 182

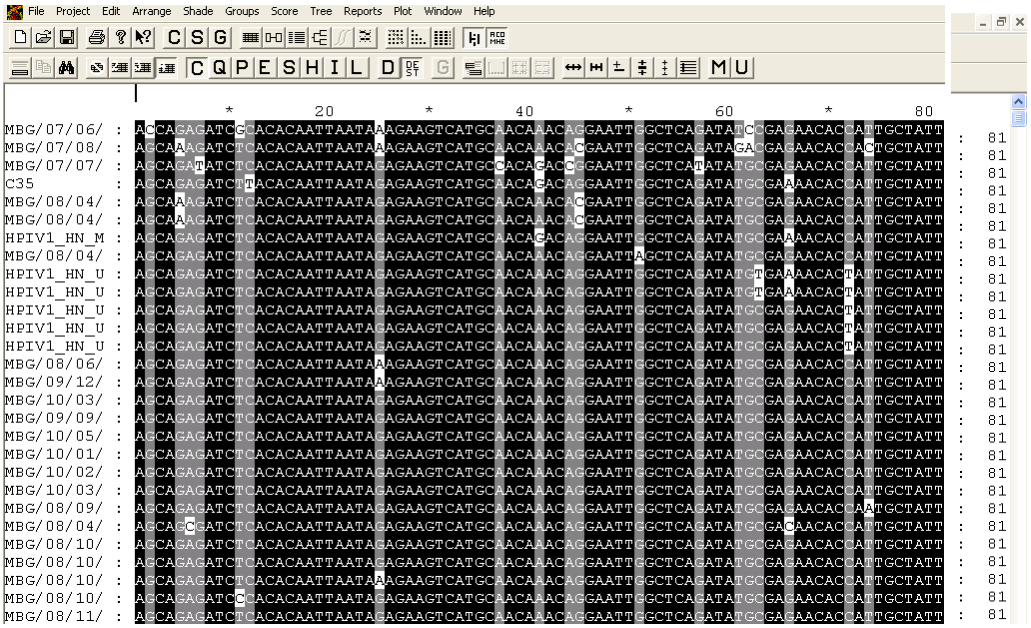
Query 183
TGGACCAAGTCTACTTTCTGGATCTACCACAATTTCAGGATGTGTTA-
GACTACCTTCAT 241
|||||||||||||||||||| |||||||||||||||||||| | || |||||| ||

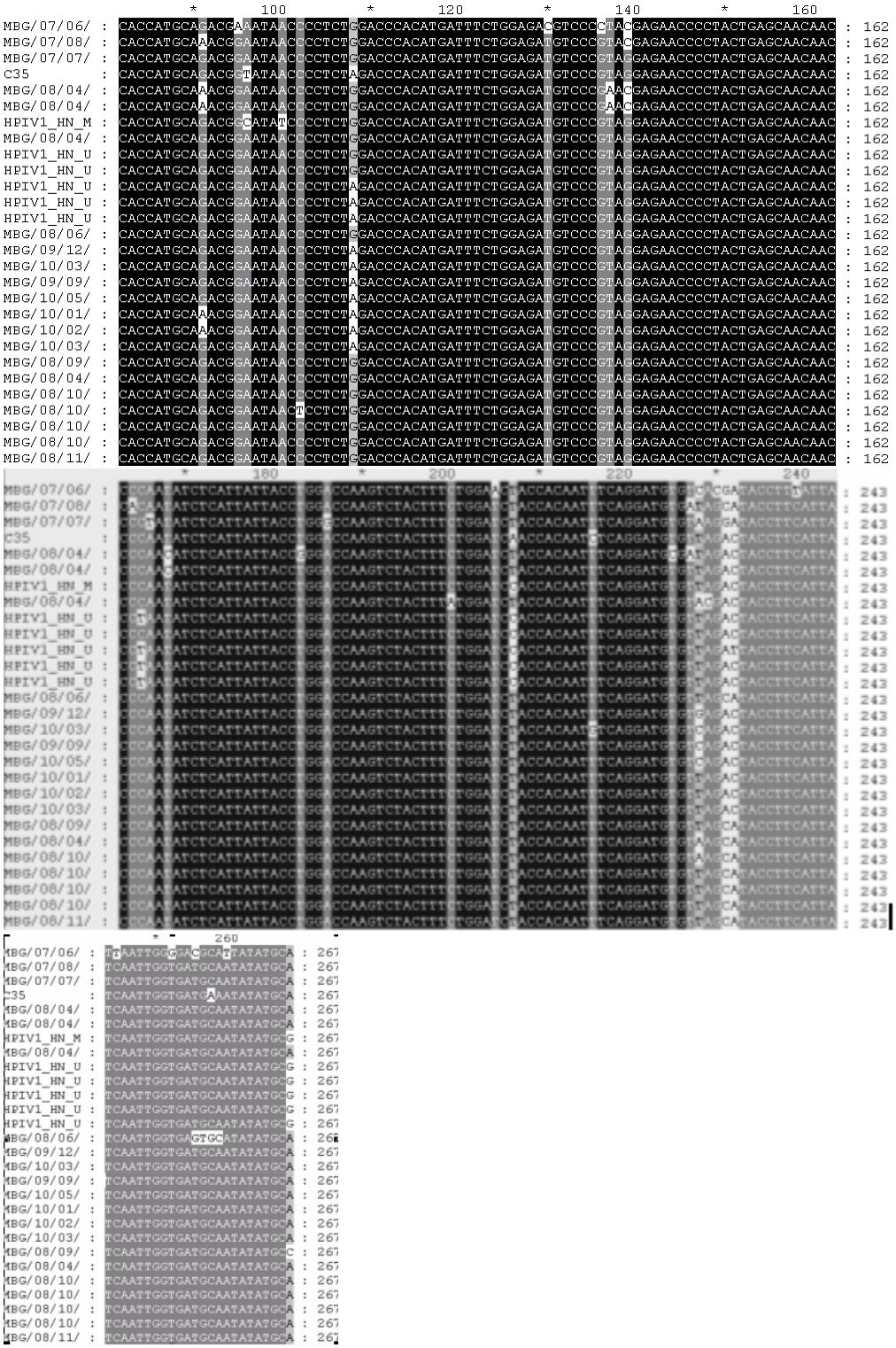
Sbjct 183
TGGACCAAGTCTACTTTCTGGAACTACCACAATTTCAGGATGTGTCACG
A-TACCTTTAT 241

Query 242 TATCAATTGGTGATGCAATATATGCA 267
||| |||||| || ||| ||||||||
Sbjct 242 TATTAATTGGGGACGCATTATATGCA 267

# APPENDIX XI

## Nucleotide polymorphisms





## APPENDIX XII

### Synonymous versus non-syonymous substitution rates

| Sample ID | No. of synonymous sites<br>*S* | No. of non-synonymous sites<br>*N* | Synonymous substitution rate<br>*dS* | Non-synonymous substitution rate<br>*dN* | *dN/dS* |
|---|---|---|---|---|---|
| MBG/07/06/01 | 67.2 | 199.8 | 0.1231 | 0.0890 | 0.7229 |
| MBG/07/07/02 | 63.3 | 203.7 | 0.1266 | 0.0352 | 0.2782 |
| MBG/07/08/03 | 64.5 | 202.5 | 0.0670 | 0.0729 | 1.0886 |
| MBG/08/04/05 | 64.3 | 202.7 | 0.1208 | 0.0356 | 0.2943 |
| MBG/08/04/06 | 66.3 | 200.7 | 0.0808 | 0.0360 | 0.4457 |
| MBG/08/04/08 | 63.1 | 203.9 | 0.1212 | 0.0150 | 0.1236 |
| MBG/08/04/09 | 60.7 | 206.3 | 0.0751 | 0.0296 | 0.3939 |
| MBG/08/06/10 | 63.0 | 204.0 | 0.0884 | 0.0405 | 0.4578 |
| MBG/08/09/11 | 61.3 | 205.7 | 0.0726 | 0.0248 | 0.3413 |
| MBG/08/10/12 | 63.4 | 203.6 | 0.0875 | 0.0199 | 0.2276 |
| MBG/08/10/13 | 65.9 | 201.1 | 0.0826 | 0.0202 | 0.2446 |
| MBG/08/10/14 | 65.9 | 201.1 | 0.0826 | 0.0202 | 0.2446 |
| MBG/08/10/15 | 66.5 | 200.5 | 0.0649 | 0.0254 | 0.3916 |
| MBG/08/1/16 | 65.0 | 202.0 | 0.667 | 0.0201 | 0.3012 |
| MBG/09/12/18 | 67.7 | 199.3 | 0.0950 | 0.0153 | 0.1612 |
| MBG/10/01/19 | 68.8 | 198.2 | 0.0760 | 0.0103 | 0.1349 |

| | | | | | |
|---|---|---|---|---|---|
| MBG/10/02/20 | 69.1 | 197.9 | 0.0754 | 0.0155 | 0.2050 |
| MBG/10/03/21 | 68.1 | 198.9 | 0.0796 | 0.0102 | 0.1326 |
| MBG/10/03/22 | 67.8 | 199.2 | 0.0964 | 0.0153 | 0.1581 |
| MBG/10/05/23 | 68.9 | 198.1 | 0.0919 | 0.0103 | 0.1116 |
| C35 | 65.1 | 201.9 | 0.0990 | 0.0101 | 0.1016 |
| HNU70938.1 | 67.7 | 199.3 | 0.0618 | 0.0102 | 0.1647 |
| HNU86785.1 | 66.6 | 200.4 | 0.0470 | 0.0101 | 0.2151 |
| HNU70936.1 | 69.4 | 197.6 | 0.0911 | 0.0103 | 0.1131 |
| HNU70937.1 | 68.8 | 198.2 | 0.0760 | 0.0103 | 0.1349 |

## APPENDIX XIII

### Codon bias

========== Codon Bias ==========
Scaled Chi-square, SChi2: 0.684
Scaled Chi-square computed without Yates' correction
G+C percentage: 50.00 %
Effective number of codons, ENC: 46.967
Codon Bias Index, CBI: 0.474

========== G+C content ==========
G+C content at noncoding positions, G+Cn: Not applicable
Average number of sites: 0.00
G+C content at second codon positions, G+C2: 0.470
Average number of sites: 89.00
G+C content at (synonymous) third codon positions, G+C3s: 0.328
Average number of sites: 87.95
G+C content at coding positions, G+Cc: 0.426
Average number of sites: 267.00

G+C content (selected region), G+C: 0.426
Average number of sites: 267.00

Sites/Codons with alignment gaps were excluded in single DNA sequences