The Genetic Diversity of Ndumu Virus Strains Isolated from Mosquitoes from Garissa, Ijara, Busia and Baringo during Routine Arbovirus Surveillance in Kenya between 2007 and 2009

Caroline Apondi Ochieng

A thesis submitted in partial fulfillment for the degree of Masters of Science in Molecular Medicine in the Jomo Kenyatta University of Agriculture and Technology

2012

DECLARATION

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

Signature……. Date ……...

Caroline Apondi Ochieng

This thesis has been submitted for examination with our approval as University supervisors.

Signature……. Date……

Dr. Rosemary Sang

KEMRI, Kenya

Signature……. Date……

Dr. Joseph Nganga

JKUAT, Kenya

Signature……. Date……

Dr. Samoel Khamadi

KEMRI, Kenya

Signature……. Date…….

Dr. Anne Fischer

ICIPE, Kenya

DEDICATION

This work is dedicated to my family, especially my loving husband Esau Otieno who unceasingly supported me both spiritually and morally in the course of this work. To my daughter Victorine Adhiambo and son Jaden Oogo who were patient with me when I needed to work extra hours to accomplish this task. To my beloved parents John and Angeline Ochieng who went out of their way to ensure that I get to where I am today. Finally, to my dear brothers and sisters who encouraged me at times when I felt like giving up.

ACKNOWLEDGEMENTS

First and foremost I appreciate God for granting me His grace and favour that enabled me to go through this entire journey. He was my strength when I felt pressed down and wearied by the challenges that I came across in the course of this work.

I feel greatly indebted to my supervisors Dr. Rosemary Sang, Dr. Joseph Nganga, Dr. Samoel Khamadi and Dr. Anne Fischer whose advice, comments, critique, professionalism and guidance have helped me to complete this task successfully. Many thanks to Dr. Rosemary Sang for allowing me to do my project in the Viral Hemorrhagic Fever laboratory in KEMRI. In addition, for availing isolates that were used for this work and also assisting me to have my samples sequenced. Without her assistance and concern, this work would have not reached where it is today. To Anne Fischer from ICIPE who assisted me to analyze my data. Her contribution in this thesis was so enormous because without data analysis then this work would not have made much sense. Despite her tight schedule, she created time and assisted me with data analysis. I wish to acknowledge Dr. Joseph Nganga for accepting to be my University supervisor and for guiding me all through from proposal development to thesis write-up.

I do appreciate the GEIS co-ordinator, Dr. David Schnabel for funding part of this project and Google project for assisting in sequencing my samples free of charge.

I wish to acknowledge George Ngondi and Ilmma from ILRI for their continued support in sequencing the isolates that were used in this project. In spite of the many challenges in trying to sequence these isolates, they did not despair until the goal was achieved. I feel greatly indebted to them.

This work would not have been successful without the assistance of the VHF laboratory staff who came in handy during the time of need. I wish to express my sincere appreciation to Dr. Marion Warigia who assisted in the primer designing and optimization process for the RT-PCR work. Her advice and guidance helped me to arrive at conditions that were optimal for the primers that I designed. To all VHF laboratory staff who assisted me in one way or another during in the course of this work. I also appreciate my classmates who gave me moral support and encouragement during hard times when I almost felt like giving up. Finally, my entire family for being patient with me especially when I was busy with this work at the time that they really needed me. Their constant love and support was like an anchor to me throughout the entire period of this work.

TABLE OF CONTENTS

TABLE

LIST OF FIGURES

LIST OF PLATES

LIST OF APPENDICES

ABBREVIATIONS AND ACRONYMS

BLAST	Basic Local Alignment Search Tool
CDC	Centre for disease control and prevention
CHIKV	Chikungunya virus
CO_2	Carbon dioxide
CPE	Cytopathic effect
CVR	Centre for Virus Research
DNA	Deoxyribonucleic acid
ds	Double stranded
EEEV	Eastern equine encephalitis virus
FBS	Foetal bovine serum
g	gravity
ICIPE	International Centre of Insect Physiology and ecology
ICTVdB	International Committee of Taxonomy of Viruses database
JKUAT	Jomo Kenyatta University of Agriculture and Technology
KEMRI	Kenya Medical Research Institute
Nduv	Ndumu virus
PBS	Phosphate buffered saline
Pfu	plaque forming unit
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase polymerase chain reaction
SISPA	Sequence-Independent Single Primer Amplification
SIRACA	Subcommittee on Interrelationships Among Catalogued Arboviruses

$TCID_{50}$ Tissue culture infectious dose 50% end point

VEEV Venezuelan equine encephalitis virus

VHF Viral hemorrhagic fever

WEEV Western equine encephalitis virus

ABSTRACT

Members of the genus *Alphavirus*, family *Togaviridae* are a diverse group of principally mosquito-borne RNA viruses that cause a variety of diseases worldwide. There are at least 29 species and many more subtypes of alphaviruses. Most of them infect humans for example Chikungunya and O'nyong-nyong viruses and cause public health problems whereas some are considered potential bioweapons, an example being Venezuelan equine encephalitis virus. Being RNA viruses, they are prone to genetic mutation like other RNA viruses, a good example being Chikungunya virus which has recently shown this phenomenon. During routine arbovirus surveillance study in selected study sites in Kenya, carried out in Viral hemorrhagic fever laboratory Centre for Virus research in Kenya Medical Research Institute, alphaviruses, identified as Ndumu viruses were isolated. They were from pools of mosquitoes that were collected from four different counties in Kenya, namely: Garissa, Ijara, Busia and Baringo. In this study, Ndumu virus isolates from the four districts in Kenya were sequenced to determine their genetic diversity. This was achieved by first culturing the virus isolates in Vero cells.The infected cells were harvested and the cell culture supernatant was used to isolate total RNA after which Reverse transcriptase polymerase chain reaction was performed using primers specific for the envelope gene (E1) to confirm the identity of the virus. This was followed by pyro-sequencing of the isolates using 454 sequencer. A mapping assembly of the sequence reads from the 454 sequencer was done against the only available partial genome sequence of Ndumu virus using GS Runmapper (Roche). Nucleotide and amino acid sequence alignments were done using the software

Muscle. Molecular Evolutionary Genetics Analysis software was used for phylogenetic analysis. The Ndumu virus isolates from Baringo, Busia and Garissa counties showed minimal genetic variation, only one isolate from Ijara was distinct. This study has facilitated the identification of single nucleotide polymorphisms (SNPs) within Ndumu virus genome. These SNPs however, have not changed the charge or hydrophobicity of proteins coded for. Nevertheless, this does not rule out that additional SNPs may change the protein coding sequence and affect the virus' virulence and/or host susceptibility. The information generated from this study may be used in the design of a control strategy for alphavirus infections.

CHAPTER ONE

1.0. INTRODUCTION

1.1. Background Information

The family *Togaviridae* comprises two genera, *Alphavirus* and *Rubivirus* (Weaver *et al*., 2000). As a genus, the *alphaviruses* are widely distributed throughout the world, inhabiting all of the continents except Antarctica. The geographic distributions of individual species are restricted because of specific ecological conditions for reservoir hosts and vectors (Johnson**,** 1988; Weaver *et al*., 2000). Members of the genus *Alphavirus* are typically maintained in natural cycles involving transmission by arthropod vectors among susceptible vertebrate hosts (Strauss and Strauss**,** 1994). Virus-host interactions may be highly specific, and sometimes only a single mosquito species is utilized as the principal vector of a virus, as has been reported for many Venezuelan equine encephalitis complex viruses (Weaver**,** 1998). These specific virus-vector interactions may limit the distribution of many Alphaviruses.

Possible exceptions to the presumption that all Alphaviruses have an arthropod host are the newly identified salmonid viruses, salmon pancreas disease virus (SPDV) (Weston *et al***.,** 1999) and sleeping disease virus (SDV) (Villoing *et al***.,** 2000). These viruses have been isolated only from infected Atlantic salmon and rainbow trout, respectively, and are not known to have arthropod vectors. It has been postulated that the sea louse, *Lepeophtheirus salmonis*, may play a role in the transmission of SPDV, but no evidence to support this hypothesis has been

generated. Parasitic lice have been implicated in the transmission of the newly discovered southern elephant seal alphavirus (SESV) from the coast of Australia. SESV has been grouped genetically with the Semliki Forest virus complex (La Linn *et al.*, 2001).

Originally isolated in 1959 from *Mansonia uniformis* in South Africa, Ndumu virus has been found throughout Africa, and although antibodies to the virus have been identified in humans from several African countries, no human illnesses have been associated with Ndumu Virus infection (Karabatsos 1985). Due to the fact that it belongs to the same genus as other viruses that cause human illnesses like Chikungunya virus (CHIKV), O'nyong-nyong virus (ONNV), Semliki forest virus (SFV) and Getah virus, it has the potential to raise public health concerns in humans with time. In addition, antibodies associated with Ndumu virus have been detected in human population (Kokernot *et al*., 1961) and therefore its medical importance cannot be underrated. CHIKV which belongs to *alphavirus* genus as Ndumu virus is an example that shows how these viruses can change from less virulent to more virulent strains (Schuffenecker et al., 2006).

CHIKV infection, which had been known to be self-limiting fever, had changed and led to a serious health problem around the world. In India, Eastern Africa, Western Africa and Central Africa, this viral disease is spread by mosquito bites from *Aedes* species mosquitoes, (Reinert *et al***.,** 2004; Womack, 1993; Powers and Logue, 2007). Recent research by the Pasteur Institute in Paris suggests the

virus has acquired a mutation that enables it to also be transmitted by *Aedes albopictus* (Tiger mosquito). A genetic change in position 226 of E1 in the CHIKV made it possible for the virus to no longer need cholesterol as viruses normally need cholesterol to infect the cells of their human and mosquito hosts (Afjal *et al*., 2002). Because mosquitoes often do not have enough cholesterol for viruses to efficiently infect their cells, it is possible that the more recent version of the CHIKV (the V version) could have survived and multiplied better in mosquitoes, which in turn could have contributed to its rapid spread (Afjal *et al*., 2002).

In 2004-2005, a Chikungunya outbreak occurred along Kenyan Coast and Comoros Island. This was associated with the presence of *Aedes aegypti* mosquitoes (Sang *et al*., 2008). Following reports of febrile-arthralgic disease outbreaks, epidemiologic and entomologic investigations were conducted in Lamu, Mombasa and Comoros in the period between June 2004 and March 2005. Epidemiologic and laboratory studies confirmed extensive CHIKV outbreaks in Lamu (beginning June 04; attack rate > 50%), Mombasa (beginning November 04), and the Comoros Islands (beginning January 05; attack rate > 50% on Grand Comoros). The attack rate >50% means that out of the entire population that was exposed to the CHIKV, more than 50% of the population got infected. Entomologic investigations identified *Aedes aegypti* as the likely primary vector in Comoros.

By October 2006 on Réunion Island alone, which has a population of 760,000, at least 266,000 cases had been reported (Flauhaut, 2007). The epidemic swept eastward into the Indian subcontinent, where by the end of the year it had caused >1.3 million cases; attack rates were 45% in some regions (WHO, 2006; Mudur, 2006). By the beginning of 2007, the epidemic was on the decline on La Réunion and the Seychelles (Barret, 2007a), but it seemed to be continuing in areas of India. New outbreaks had been reported from early 2007 in Malaysia and mid 2007 in Indonesia (Barret; 2007b; Banks, 2007). These events demonstrated the public health impact alphavirus infections can have. The impact of genetic change therefore, emphasizes the importance of monitoring and detecting the genetic diversity of Ndumu virus which belongs to the *alphavirus* genus.

1.2. General overview of Alphaviruses

The family *Togaviridae*, genus *Alphavirus* consists of a group of enveloped, single-stranded, positive-sense RNA viruses. These viruses are principally mosquito-borne and have a nearly worldwide distribution (Griffin, 2001). The genus contains at least 29 distinct, known species (some with multiple subtypes or varieties), subdivided into seven antigenic complexes. New species have been described recently from mosquitoes namely Trocara virus (Travassos da Rosa *et al*., 2001), from fish, the Salmon pancreas disease virus and Sleeping disease virus (Weston *et al*., 1999;Villoing *et al*., 2000), and from a louse infesting the southern elephant seal referred to as Southern elephant seal virus (La Linn *et al*., 2001). Many of the alphaviruses are significant pathogens of humans and domestic animals. Clinically,

disease manifests either as a mild, self-limiting febrile illness, central nervous system infection (i.e., encephalitis), or as a febrile illness with rash and polyarthropathy. New World viruses such as Venezuelan equine encephalitis virus (VEEV), Eastern equine encephalitis virus (EEEV), and Western equine encephalitis virus (WEEV) mainly cause encephalitis, whereas the Old World viruses such as Ross River, Barmah Forest, O'nyong-nyong, Chikungunya, and Sindbis viruses cause an arthralgia syndrome (Tsai *et al.*, 2002). Getah virus has been reported to induce abortion or stillbirth in pregnant swine (Kiyomasu *et al*., 1991) and Highlands J virus causes dramatic decreases in egg production and mortality in domestic birds (Wages *et al*., 1993). Disease of salmon and rainbow trout, caused by two recently described fish alphaviruses (Weston *et al*., 1999; Villoing *et al*., 2000; Jewhurst *et al*., 2004), has the potential for significant economic damage to the commercial fish industry. Some alphaviruses such as EEEV have the potential to be used as bioweapons (Vogel *et al*., 1997; Sidwell and Smee, 2003), and therefore the ability to detect any and all alphaviruses has applications for biodefense.

1.3. The genome of Alphaviruses

Alphaviruses are small sized, spherical, enveloped viruses with a genome consisting of a single strand of positive-sense RNA (Strauss and Strauss, 1994; Johnston and Peters, 1996; Schlesinger and Schlesinger, 1996). The total genome length ranges between 11,000 and 12,000 nucleotides, and has a 5’ methylated cap, and 3’ poly-A tail. There are two open reading frames (ORF’s) in the genome encoding non-structural and structural proteins. The nonstructural protein genes are coded in the 5'

two-thirds of the genome and encode proteins for transcription and replication of viral RNA. The structural genes encode four structural proteins: Capsid protein C, Envelope glycoprotein E1, Envelope glycoprotein E2, and Envelope glycoprotein E3 and are translated from a subgenomic mRNA colinear with the 3' one-third of the genome. Replication occurs within the cytoplasm, and virions mature by budding through the plasma membrane, where virus-encoded surface glycoproteins E2 and E1 are assimilated.

Schematic diagram of the alphavirus genome organization using VEEV

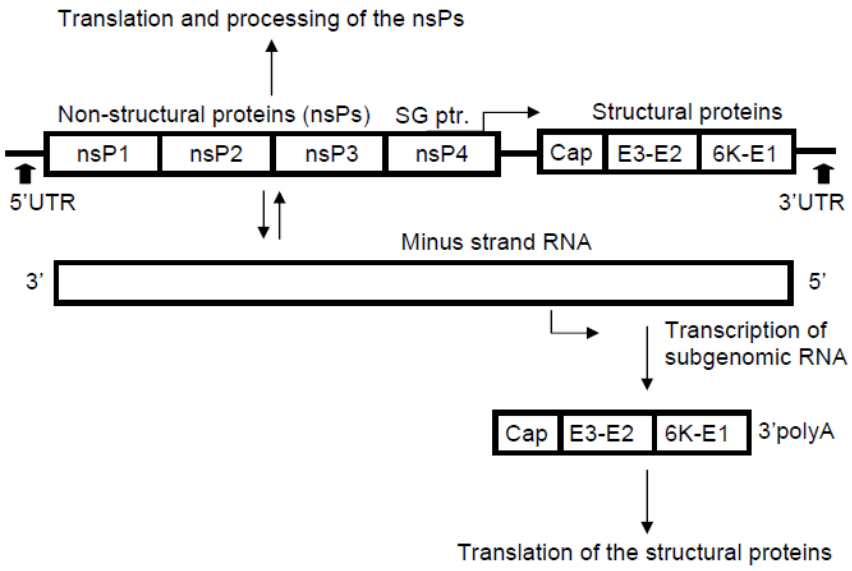


Figure 1.3.1 Organization of the VEEV genome with replication strategy

The VEEV genome is organized into two regions: the four non-structural proteins (NSPs) are located immediately after the 5' untranslated region (UTR) and the the three structural proteins. (Adopted from Colpitts, 2007).

1.4. Serological characteristics of alphaviruses

The glycoproteins E2 and E1 are the targets of serologic reactions and tests such as neutralization, complement fixation, hemagglutination inhibition and antibody production assays (Calisher et al., 1980). The alphaviruses show various degrees of antigenic cross-reactivity in these reactions, forming the basis for the seven antigenic complexes, 29 species, many subtypes and varieties of alphaviruses defined previously (Karabatsos, 1975; Tesh, 1982; Calisher and Karabatsos 1988). The E2 protein is the site of most neutralizing epitopes, while the E1 protein contains more conserved, cross-reactive epitopes.

Currently, the International Committee of Taxonomy of Viruses (ICTV) defines a virus species as a "polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche" (Van Regenmortel, 1990; Van Regenmortel, 2000). This definition included additional criteria in comparison to the previous classification, but this lead to more subjective interpretation in some cases. For example, EVEV was considered a species distinct from VEEV (Weaver *et al*., 2000), although initially it was considered a subtype of VEEV (Calisher and Karabatsos, 1988). Phylogenetic studies examining VEEV subtype I viruses in greater detail had shown clearly that EVEV fell within the VEEV subtype IAB/C/D clade (Powers *et al*., 1997; Salas *et al*., 2001). However, EVEV clearly constituted a replicating lineage (occurring only in Florida and was genetically distinct based on this distribution) and occupied a particular ecological niche (for example, it uses a mosquito vector different from those of all other VEE complex viruses). Also, EVEV had not been associated with the emergence of epidemics and epizootics like

the subtype ID and IE viruses (Weaver, 1998). Synonymizing EVEV with VEEV had been previously proposed (Kinney *et al*, 1998; Meissner *et al*, 1999); although justified in many theoretical respects, this would have important practical implications due to biological safety recommendations (Centre for Disease Control, 1999).

An additional example of the difficulties in virus classification and taxonomy was the original classification of Barmah Forest virus in the family *Bunyaviridae* based on antigenic criteria (Marshall *et al*., 1982; Dalgarno *et al*., 1984,). However, subsequent genetic characterization revealed it to be a member of the *Alphavirus* genus based on nucleic acid and protein sequences.

Despite the fundamental differences between the antigenic and polythetic species definitions, the systematics of the alphaviruses developed on antigenic grounds alone (Calisher and Karabatsos, 1988) agrees remarkably well with those of the ICTV (Weaver *et al*., 2000). The more detailed nature of the Subcommittee on Interrelationships Among Catalogued Arboviruses (SIRACA) classification of antigenic subtypes can lead to minor genetic changes that have a dramatic effect on antigenicity and thus the rapid appearance of new taxa. An example is an antigenic subtype of EEEV isolated from a human in Mississippi in 1983 (Calisher *et al*, 1990). Although this strain met antigenic criteria as a subtype, genetic analyses demonstrated that minor genetic changes resulted in the addition of an N-linked glycosylation site in the E2 protein (Weaver *et al*., 1992). Although there was no

evidence that this genotype persisted beyond 1983, these kinds of antigenic changes could be epidemiologically important.

Another example is VEEV, where only one or two amino acid substitutions in the E2 envelope glycoprotein can result in the generation of subtype IC equine-virulent strains from enzootic, equine-avirulent subtype ID progenitors (Wang *et al*., 1999). These changes may have dramatic effects on pathogenicity and host range, leading to epizootics. A completely natural classification would not distinguish these subtypes because they are paraphyletic and the epizootic viruses do not appear to constitute ongoing lineages. However, subtyping of VEEV is extremely important for public health purposes as this helps to identify the genetic differences of the viruses.

1.5. Alphavirus evolution

Previous studies of the evolutionary relationships among alphaviruses have relied on phylogenetic analyses of either partial or complete sequences from one or more of the seven protein genes (Levinson, 1990; Weaver; 1995; Weaver *et al*., 1997). Overall, these studies have produced relationships in agreement with the antigenically based approaches used traditionally for alphavirus classification (Calisher *et al*., 1980; Calisher and Karabatsos, 1988; Weaver *et al*., 2000). For example, viruses in the VEE (Weaver *et al*., 1992; Powers *et a.l*, 1997), EEE (Weaver *et al*., 1994; Brault *et al*., 1999), and WEE antigenic complexes (Weaver *et al*., 1997) have each been shown to be monophyletic (WEE complex for the envelope glycoproteins only). Additionally, phylogenetic studies have shown that

most of the New World viruses in the WEE antigenic complex (WEEV, Highlands J virus, Fort Morgan virus, and Buggy Creek virus [a variant of Fort Morgan virus]) are descendants of an ancestral alphavirus that resulted from a recombination event (Weaver *et al*., 1997). The recombination combined the E2 and E1 envelope protein genes from a Sindbis-like virus and the remaining genes from an EEEV-like ancestor (Hahn *et al*., 1988). The Old World serogroups have been studied in less detail; the Chikungunya, O'nyong-nyong, Semliki Forest, and Ross River viruses, belonging to the Semliki Forest virus complex, are monophyletic in some analyses and paraphyletic in others, with Middelburg virus falling into this group in some phylogenetic trees (Weaver *et al*., 1993;
Weaver, 1995).

1.6. Justification of the study

Ribonucleic acid (RNA) viruses are known to be prone to genetic mutations due to the error-prone nature of RNA-dependent RNA polymerase and new variants could raise public health concerns. Ndumu virus is an RNA virus and is a member of the alphavirus genus, whose members are of medical importance like Chikungunya and Onyong'-nyong viruses. The results would determine if there was genetic diversity in the Ndumu viruses as they occurred in four different districts of diverse ecologies. This diversity might have an impact on the infectivity and/or host range of the virus. Therefore, it is of importance to understand viral diversity to enable better surveillance and prediction of pandemic threats. Being an RNA virus, Ndumu virus has the potential to undergo evolutionary changes which eventually might make it

infective to humans. Protective test surveys in South Africa indicated that human beings residing in widely scattered areas had been exposed to Ndumu virus (Kokernot *et al*., 1961). Novel viruses have caused outbreaks in human population and therefore it is important to be able to document the genetic changes before and after evidence of human transmission to understand the mechanisms of disease emergence in human populations. This can also serve as early warning information so that preventive measures can be put in place before an outbreak occurs.

1.7. Null hypothesis

There is no genetic diversity among Ndumu virus isolates from Garissa, Ijara, Busia and Baringo counties in Kenya.

1.8. General Objective

To assess the genetic diversity of Ndumu virus isolates from Garissa, Ijara, Busia and Baringo counties in Kenya.

1.8.1. Specific Objectives

1. To isolate Ndumu virus in Vero cells.
2. To assess the genetic diversity of Ndumu virus isolates from Garissa, Ijara, Busia and Baringo counties in Kenya.
3. To determine the phylogenetic relatedness of the Ndumu virus isolates to other known alphaviruses.

CHAPTER TWO

2.0. LITERATURE REVIEW

2.1. Ndumu Virus

Ndumu virus derived its name from the region from which it was first isolated and identified i.e. a station located near Ndumu in Northern KwaZulu Natal, Republic of South Africa. The first two strains of this virus were isolated from mosquito pools: *Mansonia uniformis* and *Aedes circumluteolus* respectively, which were collected from this station (Kokernot *et al*., 1961). This station had been the base for periodic field activities concerned with the study of arthropod-borne viruses in South Africa. A major objective of the field program conducted there was to search for new agents. In the process an unknown virus was discovered. Serological studies utilizing Complement-Fixation and Neutralization test techniques had shown the two virus strains to be immunologically indistinguishable. Similar studies including the Hemagglutination-Inhibition technique indicated that the virus was not related to a wide spectrum of other arthropod-borne viruses. The study of physical and chemical properties showed that Ndumu virus was sensitive to the action of sodium desoxycholate. A vervet monkey was inoculated with Ndumu virus and this virus was detected in the serum on the 8th and 10^{th} postinoculation days. Protection test surveys indicated that human beings residing in widely scattered areas in southern Africa had been exposed to Ndumu virus (Kokernot *et al*., 1961). However, the epidemiology and epizootiology of these virus isolates remain obscure. Ndumu virus has been found throughout Africa, and although antibodies to the virus have been identified in humans from several African countries, no human illnesses have been

attributed to Ndumu virus infection (Kokernot *et al*. 1961, Karabatsos 1985). This may be due to misdiagnosis and/or lack of proper diagnostic tools that are being used in health facilities. At the moment, there is very little information about this virus because it has not been studied as much as other alphaviruses.

2.2. Morphology of Ndumu virus

Virions consist of an envelope and a nucleocapsid. During their life cycle, virions have not been observed outside a cellular environment and have a cell-associated cycle. Virus capsid is enveloped with a tight fitting membrane. Virions are spherical and measure about 70 nm in diameter. Surface projections are distinctive glycoprotein spikes composed of two virus proteins forming heterodimers that cover the surface. Capsid/nucleocapsid is round and exhibits icosahedral symmetry. The nucleocapsid is isometric and has a diameter of about 40 nm (Büchen-Osmond, 2006).

2.3. Nucleic Acid and proteins of Ndumu virus

The genome contains a single molecule of linear positive-sense, single-stranded RNA. The complete genome is about 11700 nucleotides long (Büchen-Osmond, 2006). The virus genome has been only partially sequenced. The longest sequence (4886bp) contains partial sequence of NSP4 and complete structural proteins sequence. The 5'-end of the genome has a methylated nucleotide cap and the 3'-terminus has a poly (A) tract. The viral genome encodes two glycoproteins (an

envelope and a nucleocapsid) and non-structural proteins (NSP1 to NSP4) (Büchen-Osmond, 2006).

2.4. Viral genome sequencing by random priming methods

The emergence of highly pathogenic viral agents from zoonotic reservoirs has energized a wave of research into viral ecology, viral discovery (Venter *et al*., 2004; Breitbart and Rohwer, 2005; Angly *et al*.,2006; Culley *et al*., 2006) and a parallel drive to develop large datasets of complete viral genomes for the study of viral evolution and pandemic prediction (Holmes *et al*.,2005; Ghedin *et al*., 2005). Viral discovery has been aided by the development of sequence independent methodologies for the generation of genomic data (Ambrose and Clewley, 2006). The most prominent of these methodologies include representational difference analysis (RDA) and sequence independent single primer amplification (SISPA) with several variations. Several modifications of the SISPA method have so far been implemented including random-PCR (rPCR) (Froussard, 1992). The rPCR method combines reverse transcription primed with an oligonucleotide made up of random hexamers tagged with a known sequence which is subsequently used as a primer-binding extension sequence. This initial modification was first used to construct a whole cDNA library from low amounts of viral RNA. A more recent modification, the DNAse- SISPA technique (Allander *et al*., 2001, Allander *et al*., 2005, Breitbart and Rohwer, 2005), includes steps to detect both RNA and DNA sequences. Combining sample filtration through a 0.22 micrometer column and a DNAse I digestion step led to the identification of viruses from clinical samples. The DNase-

SISPA technique has been used for the detection of novel bovine and human viruses from screens of clinical samples (Allander *et al*., 2001; Allander *et al*., 2005, Allander *et al*., 2007). Other groups have used the protocol for the characterization of common epitopes in enterovirus [Shin *et al*, 2003], for the identification of a novel human coronavirus (van der Hoek *et al*, 2004) and for viral discovery in the plasma of HIV infected patients (Jones *et al*., 2005). In addition to its utility for viral discovery and viral surveillance, the DNase-SISPA method has utility in obtaining full genome sequence from uncharacterized viral isolates or viral isolates from highly divergent families. The sequencing principle is the same as in the direct sequencing (Sanger) method although it does not rely on the PCR primers.

CHAPTER THREE

3.0. MATERIALS AND METHODS

This study was approved by KEMRI and given the protocol number SSC 1621 and all Scientific and Ethics of Research were adhered to in the study. It was laboratory based and archived samples in the VHF laboratory were used.

3.1. Virus isolates

The samples that were used in this study were archived samples in the VHF laboratory. They were isolated during routine arbovirus surveillance work in the VHF laboratory. During the surveillance work, the entomology team goes out to selected sites to collect mosquitoes using CDC light traps (http://johnwhock.com/download/manuals/instr_512_CDCMiniature.pdf). The captured mosquitoes are then identified/ classified using taxonomic keys (Jupp, 1986) by species and sex. After classification, they are pooled up with a maximum of 25 mosquitoes per pool. During all these stages, the cold chain is maintained at every stage. This is followed by homogenization of the whole mosquitoes using one, 4.5-mm-diameter, copper-clad steel beads (BB-caliber airgun shot) placed in the tube with the mosquitoes and diluent and shaken vigorously. The diluents contains Eagle's Minimum Essential Media (MEM) (Sigma-Aldrich, St. Louis, MO) with Earle's salts and reduced $NaHCO_3$, supplemented with 10% heat-inactivated fetal bovine serum (FBS), (Sigma-Aldrich), 2% L-Glutamine (Sigma-Aldrich), and 2% antibiotic/ antimycotic solution with 10,000 units penicillin, 10mg streptomycin and 25µg amphotericin B per ml (Sigma-Aldrich). The supernatant is harvested by

spinning the homogenate in Eppendorf tubes at 12,000 rpm for 15 min (bench top centrifuge eppendorf 5417R) to remove the suspended solids, without removing the beads. The supernatant is then used for inoculation in cell culture to isolate the virus (es) from the mosquito pools. The mosquito pools that cause CPE in cell culture are harvested and tested by RT-PCR after isolating RNA from the cell culture supernatants. Detection and identification by RT-PCR is done using a panel of primers. After identification, they are stored as virus stocks for future use. The virus isolates that were used in this study had been grown in vero cells passage 10 and preserved as virus stocks in arbovirus/Virus hemorrhagic fever (VHF) laboratory in KEMRI in the liquid nitrogen (-196°C) tank. They were identified as Ndumu virus by RT-PCR as explained in section 3.3.1 and 3.3.2 and by sequencing of the diagnostic fragment (E1 protein). A total of eight isolates, two mosquito pools from each region were used. Each mosquito pool had a maximum of 25 mosquitoes. They included: Garissa- GSA/S1/936 and GSA/S5/4278; Busia- BSA/S4/2265 and BSA/S4/2268; Baringo- BAR/S2/3527 and BAR/S2/3526 and Ijara- AMH 001346 and AMH 001056.

SITE	Sample Identity	Invertebrate vector species	Virus identity	Date of collection
Garissa	GSA/S1/936	*Ae. mcintoshi*	Ndumu	5.09.2007
	GSA/S5/4278	*Ae. ochraceus*	Ndumu	10.05.2008
Ijara	AMH 001056	*Ae. ochraceus*	Ndumu?	17.12.2009
	AMH 001346	*Ae. mcintoshi*	Ndumu	17.12.2009
Baringo	BAR/S2/3526	*Cx.rubinotus*	Ndumu	22.10.2007
	BAR/S2/3527	*Cx.rubinotus*	Ndumu	22.10.2007
Busia	BSA/S4/2265	*Coq.fraseri*	Ndumu	2.12.2007
	BSA/S4/2268	*Coq.fraseri*	Ndumu	2.12.2007

Table 3.1 Virus isolates from four different counties in Kenya

3.2. Cell culture

3.2.1. Cell lines

Vero cells refers to a cell line derived from the kidney of the African green monkey and is used primarily in virus replication studies and plaque assays. The Vero cells that were used to culture all the virus strains were from ATCC (ATTC® Number: CCL-81™) and had been passaged ten times. They were cultured in T-25 cm^2 culture flask overnight using Eagle's Minimum Essential Media (MEM) (Sigma-Aldrich, St. Louis, MO) with Earle's salts and reduced $NaHCO_3$, supplemented with 10% heat-inactivated fetal bovine serum (FBS), (Sigma-Aldrich), 2% L-Glutamine (Sigma-Aldrich), and 2% antibiotic/ antimycotic solution with 10,000 units penicillin, 10mg streptomycin and 25µg amphotericin B per ml (Sigma-Aldrich) incubated at 37°C in a humidified incubator with 5.0 % CO_2, to form confluent monolayers.

3.2.2. Cell culture inoculation

The virus isolates were retrieved from the liquid nitrogen tank and thawed at room temperature. The cell cultures which had been grown in 25cm^2 tissue culture flasks were washed with sterile PBS pre-warmed in the water bath at 37°C. Two hundred microlitre of thawed sample was inoculated in the monolayer followed by incubation in 5% CO_2 incubator at 37°C for 1hr with frequent rocking at an interval of 15 minutes in order to allow for virus adsorption. After incubation, the infected cells were maintained in 5ml MEM with 2% FBS, 2% L-Glutamine and 2% antibiotic/ antimycotic, incubated at 37°C in 5% CO_2. The cells were monitored twice a day for cytopathic effects (CPE). The photographs showing the type of CPE formed by each virus isolate was taken using a camera. The infected cells were then harvested by spinning down and the supernatant used for molecular assay while the remainder was kept as virus stocks in the -80°C freezer in the VHF laboratory.

3.2.3. Harvesting of the virus

When about 50% CPE was observed, the virus was harvested by freezing down the infected cells in the -80 ºC freezer overnight to lyse the cells thus releasing the viral particles. Then the frozen cells were thawed on ice and the suspension was transferred into a 15 ml centrifuge tube and centrifuged (KUBOTA KS 5000 centrifuge) at 277xg for 5min to sediment the cells. The supernatant containing the virus was placed in 1ml cryovials tubes in 0.5ml aliquots and stored at -80ºC freezer.

3.3. Genetic analysis

3.3.1. RNA Extraction

Viral RNA was extracted in triplicates from 250µl each of cell culture suspension that was harvested after 24 hours post inoculation by using the TRIZOL® LS reagent (Invitrogen)- Chloroform (Sigma) method (Chomczynski and Sacchi, 1987) and the RNA pellet re-suspended in 11 µl of nuclease free water.

3.3.2. RT-PCR Assays

A two-step reverse transcription PCR (RT-PCR) was carried out. First strand cDNA was synthesized by combining 5ng of random Hexamer primer (Invitrogen) and 10µl of RNA and the mixture was incubated at 70°C for 10 minutes to denature the RNA and also to allow the primer to anneal to the RNA. The mixture was cooled down at 4°C for 5 minutes and then the following were added to the tubes: 4 µl of 5x first strand buffer (Invitrogen), 0.01 µmoles of dNTPs (Invitrogen), 0.02 µmoles of DTT (Invitrogen), 10U of RNase Out inhibitor (Invitrogen) and 100U of SuperScript III reverse transcriptase (Invitrogen) and incubated at the following conditions: 25°C for 15 min, 42°C for 50 min, 70°C for 15 min and 4°C hold temperature. The final volume for this reaction was 20µl. Then, PCR was carried out using the Amplitaq Gold PCR mastermix (Applied Biosystems). The following components were combined together: 12.5 µl of Amplitaq Gold PCR master mix, 25 picomoles each of primer ND 124F (5'-CAC CCT AAA AGT GAC GTT-3') and ND 615R (5'- ATT GCA GAT GGG ATA CCG-3') (Kinney and Pfeffer, 2001; Bryant *et al*. 2005), 2 µl of the cDNA and 9.5 µl of Water to top up to 25 µl. They

were then placed on a thermocycler (Applied Biosystems GeneAmp®PCR System 9700) set at the following conditions: 95 °C for 10 min followed by 35 cycles of 95 °C for 30 sec, 50 °C for 30 sec , 72 °C for 45 sec and a final extension of 72 °C for 7 min and 4 °C hold temperature. The primers targeted the envelope (E1) gene and the expected PCR product was 508bp. This was done to confirm the identity of the virus stocks before proceeding with sequencing.

3.3.3. Nucleotide Sequencing

Sequence-independent single primer amplification (SISPA) method was used to convert viral RNA to cDNA using random-tagged and poly-T tagged primers (FR26RV-N and FR40RV-T). cDNA synthesis was done by combining 40pmoles of primer FR26RV-N (5'GCCGGAGCTCTGCAGATATCNNNNNN3'**)** and 10 µl of the RNA template and incubated at 65ºC for 5 minutes then immediately placed on ice. The following components were added to each tube: 4µl of First Strand Buffer (5X; Invitrogen), 0.2 µmoles of DTT (Invitrogen), 0.01 µmoles of dNTPs, 8U of RNase Out inhibitor (Invitrogen) and 100U of SuperScriptIII reverse transcriptase (Invitrogen)and incubated at the following conditions: 25°C for 10 min, 50°C for 50 min, 85°C for 10 min and 4°C hold temperature.

A Klenow reaction which converts the cDNA into dsDNA on one DNA strand was then carried out. Second strand DNA was synthesized using Klenow exo-DNA polymerase, in the presence of random tagged and virus specific 5' end oligo primers. To the cDNA reaction (19.7µl) the following components were added: 0.5ul (2.5 units) of the 3'-5'exo- Klenow DNA Polymerase and incubated at the

following conditions: 37°C for 60 minutes, 75°C for 10 minutes and 4°C hold temperature.

This was followed by Shrimp Alkaline Phosphatase/Exonuclease Treatment (SAP-EXO treatment). The SAP/Exo treatment degrades nucleotides and any single-stranded DNA (primers) left over after the PCR. To the Klenow reaction the following components were added: 1 µl SAP buffer (10X), 1 µl of SAP, 0.2µl of Exo and water to top up to a total volume of 40 µl. This was followed by an incubation of 37°C for 60 minutes and at 72°C for 15 minutes.

PCR was then carried out using the Invitrogen Accuprime Kit for PCR amplification of Klenow products. The following components were combined together: 5 µl of Buffer I (10x), 1U of taq polymerase, 4µl Klenow Product, 40picomoles of primer FR26RV (5'GCCGGAGCTCTGCAGATATC 3') and 38.8 µl of Water. They were then placed on a thermocycler (Applied Biosystems GeneAmp®PCR System 9700) set at the following conditions: 94 °C for 2 min followed by 35 cycles of 94 °C for 30 sec, 55 °C for 30 sec , 68 °C for 15 sec and a final extension of 68 °C for 10 min and 4 °C hold temperature. Finally, the PCR products were sequenced using the GS FLX Titanium Sequencing Kit XLR70 (Roche) in combination with the matching GS FLX Titanium PicoTiterPlate Kit 70 × 75 (Roche) in a 454 sequencer according to the sequencing method manual (Roche) which is available in the following website: (http://www.high-throughput-sequencing.com/manuals_roche/jan2010/GSFLXTitanium_Sequencing_Method_Manual_RevJan2010.pdf)

This sequencer was able to sequence the complete genomes for the virus isolates in short fragments (contigs) ranging from 300bp to 800bp.

The PCR products for confirmation of identity were analyzed on a 2% agarose gel (AGTC Bioproducts LTD). A mapping assembly of the nucleotide sequence data from the 454 sequencer was done using GS Runmapper (Roche 454). The sequence reads were aligned to the longest available Ndumu virus sequence (Accession No. AF339487.1). The nucleotide sequences for the isolates under study were also translated into amino acid sequences using the translate tool in (http://www.dsimb.inserm.fr/~fuchs/M2BI/AnalSeq/Annexes/Serveurs_db/ExPASy%20Proteomics%20Server.htm) Expasy proteomics server. Nucleotide and amino acid sequence alignments were done using Muscle v 3.7 software (Edgar, 2004; Dereeper *et al*., 2008). Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 was used for phylogenetic analysis (Tamura *et al*., 2007) on both the nucleotide and translated amino acid sequences to determine the diversity of Ndumu virus from different districts and the phylogenetic relatedness of the Ndumu virus strains and other known alphaviruses by use of neighbour joining tree method based on the number of nucleotide or amino–acid differences. Bootstrap resampling to determine confidence values on the groupings within trees was performed with one thousand replicates (Felsenstein, 1985).

CHAPTER FOUR

4.0. RESULTS

4.1. CELL CULTURE

The plates 4.1.1 to 4.1.3 below show the non-infected vero cells (plate 4.1.1) which was used as the negative control and the vero cells (plates 4.1.2 and 4.1.3) that were infected with the virus isolates which were being investigated. The photographs for the plates were taken at a low magnification of x40.

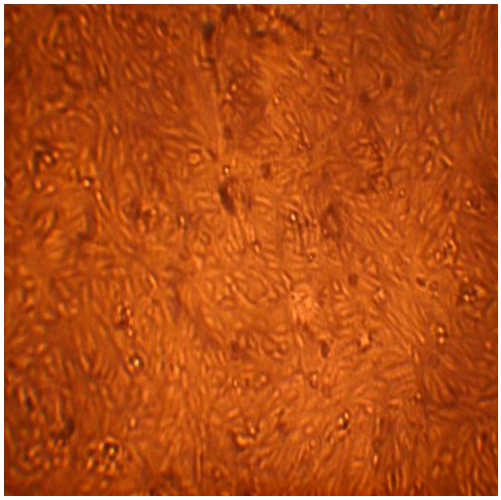


plate 4.1.1 negative control (non-infected Vero cells)

Magnification: x40

This plate shows intact monolayer that was completely compacted.

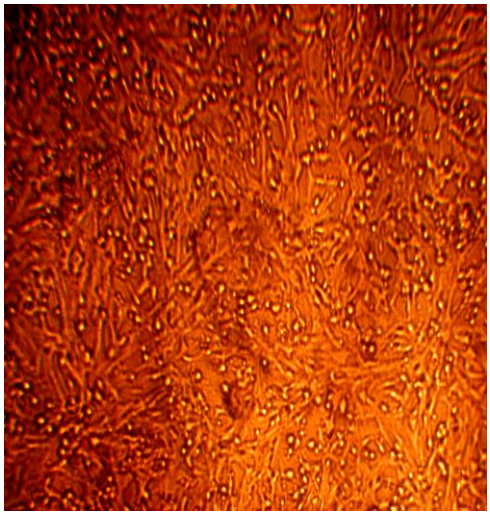


plate 4.1.2 Cells infected with Ndumu virus isolates

Magnification: x40

The plate shows CPE characterized by cells rounding up and remaining on the monolayer. The isolates: GSA/S1/936; GSA/S5/4278; BSA/S4/2265; BSA/S4/2268; BAR/S2/3526; BAR/S2/3527 and AMH001346 showed a similar type of CPE that is represented by this plate.

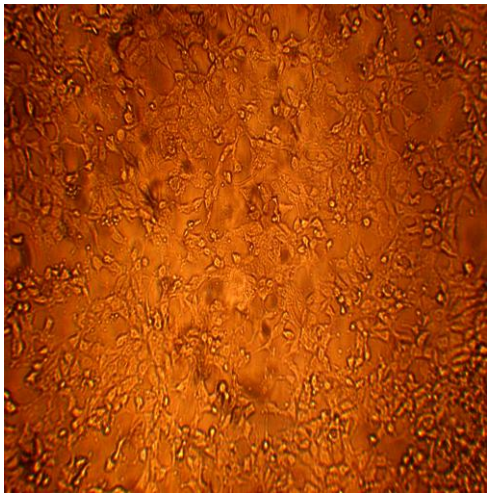


plate 4.1.3 Cells infected with virus AMH001056

Magnification: x40

The plate showing CPE characterized by cells elongating and falling off from the flask surface thus leaving a lot of spaces in between the cell monolayer. This isolate caused a different type of CPE from the other isolates.

4.2. RT-PCR RESULTS

Plate 4.2.1 shows the photograph of PCR products of the 8 isolates when they were resolved on 2% agarose gel after RT-PCR. The E1 protein of each isolate was amplified by RT-PCR as described in the methods above and the amplicons were resolved on agarose gel to separate the DNA fragments based on size and charge. This was done to confirm the identity of the virus isolates under investigation by comparing their PCR products with that of the positive control. Seven isolates produced bands of the expected size (508bp) which were similar to the positive control while one isolate (AMH001056) yielded a non-specific band. Initially, this sample was detected and identified as Ndumu virus using alphavirus genus primers and Ndumu virus primers. This sample was included in the study because it was considered a good candidate for the study of genetic diversity based on the initial RT-PCR results that were obtained. From the gel photograph (Plate 4.2.1), the negative control was negative and the positive control was positive indicating that the PCR reaction was successful.

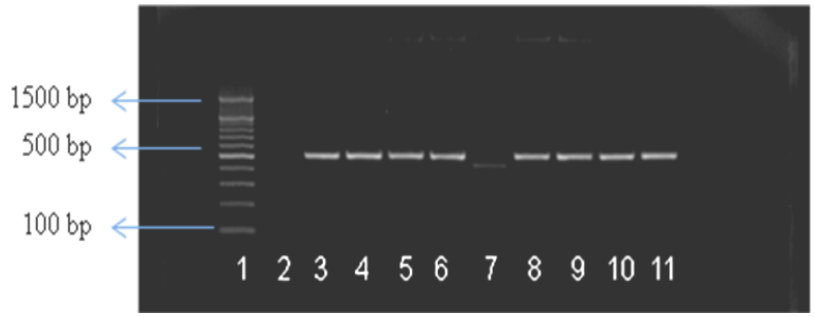


plate 4.2.1 Gel photograph of 8 isolates of Ndumu virus

Lane	**isolate**
1	100bp DNA ladder
2	Negative control
3	positive control- one of the isolates from virus stocks that were being used as a control
4	AMH001346 (AMH- Avid mosquito homogenate is an isolate from Ijara)
5	GSA/S1/936 (GSA-Garissa- isolate from Garissa)
6	BAR/S2/3526 (BAR- Baringo- isolate from Baringo)
7	AMH001056 (AMH- Avid mosquito homogenate is an isolate from Ijara)
8	BAR/S2/3527 (BAR- Baringo- isolate from Baringo)
9	BSA/S4/2265 (BSA- Busia- isolate from Busia)
10	BSA/S4/2268 (BSA- Busia- isolate from Busia)
11	GSA/S5/4278 (GSA-Garissa- isolate from Garissa)

4.3. SEQUENCING RESULTS

The SISPA products were sequenced using 454 pyrosequencer (Roche). A partial genome sequence of Ndumu virus (genbank Accession No. AF339487.1) was used as a reference. This was the longest nucleotide sequence for Ndumu virus that was available in the database and it was the one against which each of the contigs was aligned in order to come up with continuous sequences. The isolate AMH001056 was found to be Semliki forest virus after a similarity search using (blastn) (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome) against the genbank database and was included in the phylogenetic analysis.

The nucleotide sequence comparison between the seven isolates under study show that six were similar. One isolate from Baringo (BAR/S2/3526) had SNPs at two positions and the two isolates from Busia had a SNP at one position each (Appendix 1). Isolate AMH001346 had the highest number of SNPs at 49 positions of the genome (Appendix 1) and therefore the most genetically different from the other six virus isolates under study. The seven isolates under study showed 97-99% maximum identity to the reference Ndumu virus isolate when similarity search was done using Blastn. At the amino acid level, the Ijara isolate (AMH001346) was different from the other six isolates at six positions. On the other hand, it was different at 11 positions with reference to the virus from the genbank. The other six isolates were different from the reference virus at 7 positions (Appendix 2). The Ijara isolate (AMH001346) had the least similarity with the Ndumu virus reference genome and the other six isolates under study.

4.4. PHYLOGENETIC TREE ANALYSIS

This analysis was done using neighbor joining method of phylogenetic tree generation in order to study how the isolates under study and the reference virus cluster in the tree. This provides better understanding of the virus origin or ancestry in relation to other strains of the same or similar viruses.

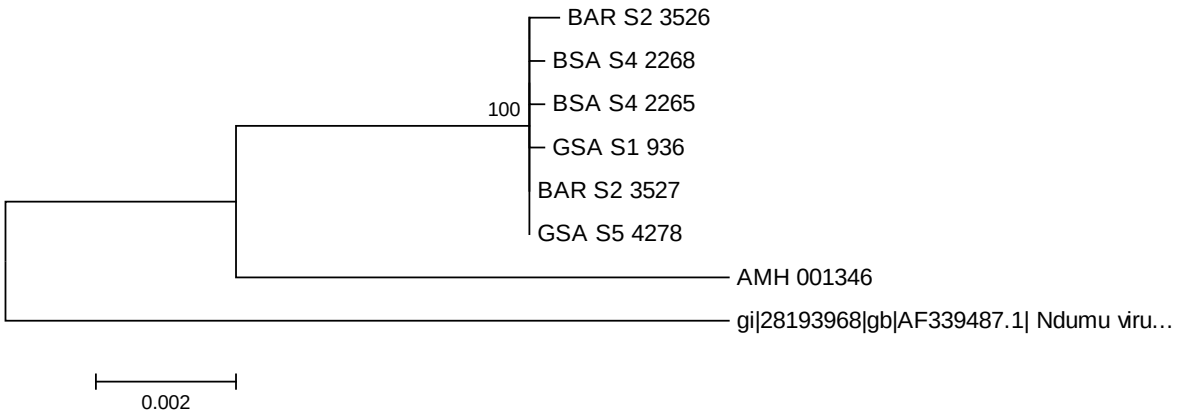


Figure 4.4.1 Neighbour Joining tree using 1000 replicates (bootstraps) on number of nucleotide sequence differences of Ndumu virus isolates and the reference strain from genbank

Legend:

BAR- Baringo

BSA- Busia

GSA- Garissa

AMH- Avid mosquito homogenate- Ijara isolate

Figure 4.4.1 shows that six isolates are nearly identical and cluster together, whereas one isolate from Ijara is distinct. The Ndumu virus reference sequence is used to root the tree.

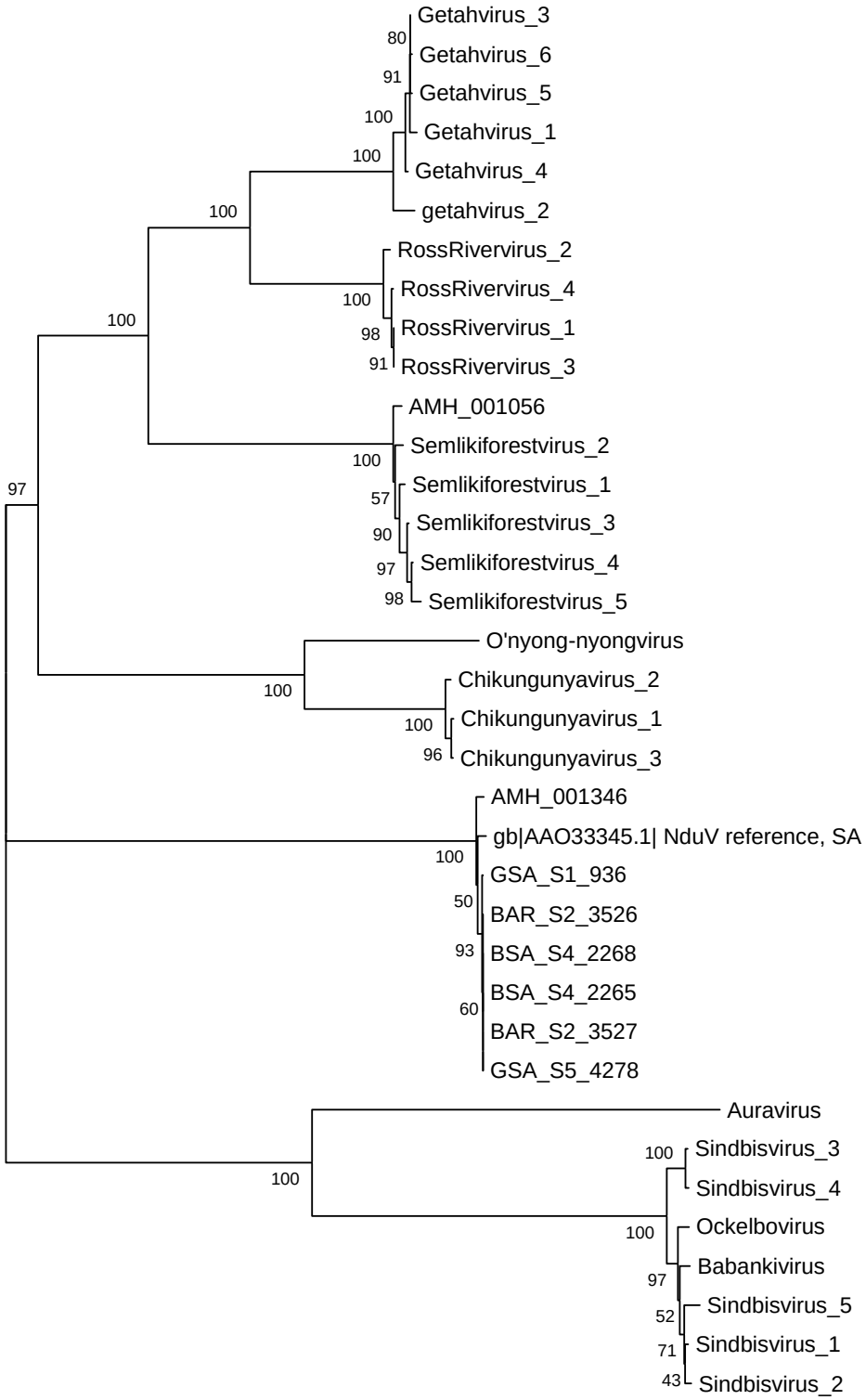


Figure 4.4.2 Neighbour Joining tree with 1000 bootstrap based on the amino acid sequences for alphaviruses

It shows the relationship of Ndumu virus with other alphaviruses based on amino-acid sequences. All the Ndumu viruses clustered together forming a clade, a pattern which is common to most alphaviruses. The isolate AMH001346 branched off first, followed by the reference virus and finally the 6 genetically closely related isolates from Garissa, Baringo and Busia.

CHAPTER FIVE

5.0. DISCUSSION, CONCLUSION, LIMITATIONS AND RECOMMENDATIONS

5.1. DISCUSSION

Alphaviruses are genetically highly diverse (Weston *et al*., 2005) creating challenges for their detection and monitoring in environmental samples such as RNA extracts and amplification from mosquitoes. Due to this diversity, separate assays and primers targeting different serogroups must be performed in order to detect them. The possibility of more than one alphavirus type/species being present in a pool of mosquitoes presents additional challenge. Alphaviruses grow fast in cell culture and are known to cause cytopathic effects after a period of time ranging from one to three days post inoculation. The Ndumu virus isolates under study caused CPE after 24 hrs post inoculation. The CPE was characterized by cells rounding up while still attached to cell culture flask surface and by the second day the whole cell sheet had been cleared from the culture surface. The isolate AMH001056, which was one of the two isolates from Ijara county, showed a different type of CPE which was characterized by cells elongating and falling off from the flask surface. This is an indicator that such difference in cytopathogenic changes could be used to differentiate alphavirus isolates at cell culture level.

Molecular analysis of viral isolates

Sequence-Independent Single Primer Amplification (SISPA) method was used to generate cDNA libraries using random primers to amplify any RNA in the sample.

This method was particularly useful for obtaining genome sequences from RNA viruses (Djikeng *et al.,* 2008). Because most sequencing methods for RNA viruses depend on RT-PCR with primers designed from pre-existing sequence data, the utility of SISPA method was found to be particularly useful for highly variable or degenerate viral families or for viruses with little available sequence information. In addition, the SISPA method was useful for uncharacterized viruses as no prior sequence information was required. This method was also found to be useful for this study because Ndumu virus had not been sequenced fully at the time this work was being done. Therefore, it was possible to sequence the entire genome even without prior sequence information. The challenge was to assemble all the reads to get a complete genome of the Ndumu virus due to unavailability of the complete genome sequence from the genbank.

As a result, this study focused on the available 4900bp section (Accession No. AF339487.1) of the Ndumu virus genome that was available in the genbank database at the time. If the complete sequence of a closely related alphavirus would be available, the contigs from the de novo assembly could have been aligned against it in order to come up with complete sequences of the isolates. Currently, work on getting full sequences of Ndumu virus is in progress and once done, the sequences will be made available in Genbank.

The results have shown that the isolates from Garissa, Busia and Baringo counties were different at very few positions at both nucleotide and protein levels despite the

geographical distance between sampling locations. One of the isolates from Ijara county (AMH001346) was genetically different (had 49 SNPs) from the isolates from the other three districts. Although Ijara is closer to Garissa county than to the other two counties, the Ijara isolate was very different from the Garissa isolates. Ijara is ecologically different considering the proximity to Tana river delta, closeness to the Indian ocean, presence of the Boni forest near the ocean and more alphaviruses have been isolated in Ijara than in Garissa (unpublished data). However, it is not very clear if the genetic diversity of the Ijara isolate is due to its geographic origin. Co-circulation of many alphaviruses provides the opportunity for viruses to evolve. The genetic divergence of this Ndumu virus isolate cannot be underscored because with time, this change in the genetic make-up of the virus may occur at a position that affects the protein which influences the virulence of the virus thus making it more virulent as was in the case of Chikungunya virus.

Chikungunya virus which was generally not fatal turned into a fatal strain due to just a single mutation that caused a single amino acid change at position 226 of E1. This led to many deaths in 2006 on Reunion Island and a widespread outbreak in India. An analysis of the virus's genetic code (Prescott, 1993) suggested that the increased severity of the 2006 outbreak might have been due to a change in the amino acid sequence, altering the virus's coat protein, which potentially allows it to multiply more easily in mosquito cells. In July 2006, a team analyzed the virus's RNA and determined the genetic changes that had occurred in various strains of the virus and

identified those genetic sequences which led to the increased virulence of recent strains (Mudur, 2006).

Phylogenetic tree analysis

The two isolates from Busia and Baringo came from different species of mosquitoes, *Coquelitidia fraseri* and *Culex rubinotus* respectively. This could be revealing that viruses isolated from the same species of mosquitoes collected from the same region might be having very little or no diversity in their genetic makeup. On the other hand, the two isolates from Garissa were isolated from two different mosquito species: *Ae. Ochraceus* and *Ae. Mcintoshi*. Despite the fact that they were isolated from the same region, there were genetic differences in some positions, though very few. The Ijara isolate AMH001346 branched off separately from the other six isolates. It was isolated from the same species of mosquito pool as the isolate GSA/S1/936 from Garissa but they separated in the phylogenetic tree. This suggests that both vector species and the region of isolation have to be the same for the genetic makeup of the virus to be identical. The Ndumu virus reference isolate was also different from the rest of the isolate. This was a South African strain while the rest were Kenyan and therefore the diversity was expected. Previous studies of Alphavirus diversification have emphasized host switching events and geographic introductions in the evolution of the alphavirus genus (Weaver, 1995; Powers *et al.*, 1997; Weaver *et al*., 1997; Brault *et al*., 1999; Powers *et a.l*, 2000).

The analysis of translated sequences of Ndumu virus isolates in relation to a number of Alphaviruses showed that alphavirus species in general make monophyletic clusters in the phylogenetic tree. This confirms the previous report by Powers and the team that viruses within a given antigenic serocomplex were usually genetically more closely related than viruses in different complexes (Powers *et al*., 2001). From figure 4.4.2, all viruses of the same species cluster together in the tree. For instance, Chikungunya viruses are grouped together and are closely related to O'nyong-nyong virus. None of the alphaviruses cluster with a different species even though they belong to the same genus. Ndumu virus formed an out group to O'nyong-nyong, Chikungunya, Ross river and Getah viruses.

5.2. CONCLUSION

This study has shown that the Ndumu virus isolates from Busia, Baringo and Garissa exhibited minimal genetic diversity despite the geographic distance separating them. The Ndumu virus isolate from Ijara county was the most distinct from all the rest. The phylogenetic analysis of the Ndumu virus isolates revealed that although Ijara and Garissa counties are closer geographically, the Ndumu virus isolates from these two counties were the most genetically distant. Nevertheless only one isolate from Ijara was sequenced, and more data will be necessary to confirm this finding.

In the study of Ndumu virus relatedness to other known alphaviruses, it was observed that Ndumu virus was distantly related to other alphaviruses and this confirms what has been reported before that alphavirus species are quite divergent (Weston *et al*., 2005).

5.3. LIMITATIONS

This study had some limitations which are worthwhile to mention. For instance, there was very little information about Ndumu virus since it has not been studied as much as other alphaviruses. This does not mean that the virus is of less importance to public health but it might be due to misdiagnosis and lack of diagnostic tools to detect the virus. Being an RNA virus, Ndumu virus has the potential to undergo genetic change with time and become dangerous to human health. From this study, the presence of SNPs in the Ijara isolate suggests that genetic changes are occurring in Ndumu virus and like in the case of Chikungunya virus; it might cause illnesses in humans after a period of time. In addition, there were no recent publications on the subject and this might have been attributed by the reasons mentioned above. The lack of complete Ndumu virus sequence in the Genbank database at the time of analysis of the data generated by this project made it difficult to assemble all the contigs generated by the sequencing method that was used in this study. The mapping process of the sequences required the availability of a reference sequence for this to be achieved. This led to the use of half of the genome for the study of genetic diversity.

5.4. RECOMMENDATIONS

More work beyond this dissertation will be done to get the complete genome sequence of Ndumu virus deposited in the Genbank database which can be used for further analysis of this virus' genome. This will be able to give the complete picture of the genetic nature of this virus which will in turn help to study the genetic

diversity of this virus based on the entire genome. Molecular genetic assessment of the genomes may lead to identification of viral elements involved in or responsible for the outbreaks in an area. This information would be useful for application in preventive and control measures, in antiviral drug development or vaccine development against the virus or alphaviruses in general.

I also recommend that more samples from the four counties be analyzed in order to make a concrete conclusion that Ndumu virus from Ijara county is actually different from those obtained from the other areas. Finally, I recommend that more specific primers be designed to be used for detection of Ndumu virus because the primers that were used in this study were not very specific because they picked Semliki forest virus and yielded a slightly smaller fragment than expected.

REFERENCES

Afjal H. K., Kouichi M., Maria del C. P., Futoshi H., Edward G. M. M. and Akira I. (**2002).** Complete nucleotide sequence of Chikungunya virus and evidence for an internal polyadenylation site. *Journal of General Virology*, **83**: 3075-3084.

Allander T, Andreasson K, Gupta S, Bjerkner A, Bogdanovic G, Persson MA, Dalianis T, Ramqvist T, Andersson B. (**2007).** Identification of a third human polyomavirus. *Journal of Virology*, **81**(8):4130-4136.

Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. **(2001).** A virus discovery method
incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(20):11609-11614.

Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B. (**2005).** Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(36):12891-12896.

Ambrose HE, Clewley JP. (**2006).** Virus discovery by sequence-independent genome
amplification. *Reviews in medical virology*, **16**(6):365-383.

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes
M, Kelley S, Liu H., Mahaffy J. M., Mueller J. E., Nulton J., Olson R., Parsons R., Rayhawk S. , Suttle C. A., Rohwer F. (**2006).** The marine viromes of four oceanic regions. *PLoS biology*, **4**(11):e368.

Banks A.L. **(2007)** Chikungunya; Indian Ocean update. ProMED. 2007 Jan 21 [cited 2007 Apr 13]. Available from *http://www.promedmail.org*, archive no 20070121.20070285.

Barret B. **(2007a).** Chikungunya;Indian Ocean update (06): Sri Lanka, India, Islands. ProMED. 2007 Mar 18 [cited 2007 Apr 13]. Available from *http://www.promedmail.org*, archive no. 20070320.20070974.

Barret B. **(2007b).** Chikungunya; Indian Ocean update (10): Indonesia (Bandarlampung). ProMED. 2007 May 30 [cited 2007 May 31]. Available from *http://www.promedmail.org*, archive no. 20070530.20071750.

Bianchi T. I., Aviles G., Monath T. P. and Sabattini M. S. (1993). Western equine encephalomyelitis: virulence markers and their epidemiological significance. *American Journal of Tropical Medicine and Hygiene,* **49:**322-328.

Brault A. C., Powers A. M., Chavez C. L., Lopez R. N., Cachon M. F., Gutierrez L. F., Kang W., Tesh R. B., Shope R. E., and Weaver S. C. (1999). Genetic and antigenic diversity among eastern equine encephalitis viruses from North, Central, and South America. *American Journal of Tropical Medicine and Hygiene,* **61**:579-586.

Breitbart M, Rohwer F. (**2005).** Method for discovering novel DNA viruses in blood using
viral particle selection and shotgun sequencing. *BioTechniques*, **39**(5)**:**729-736.

Bryant J. E., Crabtree M. B., Nam V. S., Yen N. T., Duc H. M., and Miller B. R.(2005). Isolation of arboviruses from mosquitoes collected in northern Vietnam. *American Journal of Tropical Medicine and Hygiene,* **73**: 470–473.

Büchen-Osmond C. (Ed). ICTVdB Management. **(2006)**. 00.073.0.01.017.
Ndumu virus. In: *ICTVdB - The Universal Virus Database*, version 4, Columbia University, New York, USA.

Calisher C. H. and Karabatsos N. (**1988)**. Arbovirus serogroups: definition and geographic distribution. p. In: T. P. Monath (ed.), The arboviruses: epidemiology and ecology, CRC Press, Inc., Boca Raton, Fla. **I:** 19-57.

Calisher C. H., Karabatsos N., Foster J. P., Pallansch M. and Roehrig J. T. (1990). Identification of an antigenic subtype of eastern equine encephalitis virus isolated from a human. *Journal of Clinical Microbiology,* **28**:373-374.

Calisher C. H., Shope R. E., Brandt W., Casals J., Karabatsos N., Murphy F. A., Tesh R. B. and Wiebe M. E. (**1980)**. Proposed antigenic classification of registered arboviruses. *Intervirology,* **14:**229-232.

Centre for Disease control and prevention, U.S. Government Printing Office, Washington, D.C. (1999). Biosafety in microbiological and biomedical laboratories, 4th edition, HHS Publication No. (CDC) 93-8395. **U.S. Department of Health and Human Services,** Centers for Disease control and prevention, Washington DC**.**

Charrel R.N., de Lamballerie X. and Raoult D. (**2007).** Chikungunya outbreaks—the globalization of vectorborne diseases. Northern England. *Journal of Medicine***,** **356***:*769–71.

Chomczynski P. and Sacchi N. (1987). Single Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. *Analytical Biochemistry,* **162**: 156-159.

Chretien J., Bedno S., Anyamba A., Breiman R., Njuguna C., Sang R., Sergon K., Powers A., Onyango C., Small J., Linthicum K. and Tucker C. (2006). Eco-climatic precursors to large chikungunya outbreaks in Kenya and Comoros, 2004-2005 [abstract]. International Conference on *Emerging Infectious Diseases*, March 19-22, 2006, Atlanta, Georgia.

Colpitts T. M. (2007). Characterization of the entry mechanisms utilized by the alphavirus Venezuelan Equine encephalitis virus to infect mosquito cells. pHD thesis,The University of Texas Medical Branch.

Culley AI, Lang AS, Suttle CA. (**2006).** Metagenomic analysis of coastal RNA virus communities. *Science*, **312**(5781):1795-1798.

Dalgarno L., Short N. J., Hardy C. M., Bell J. R., Strauss J. H. and Marshall I. D. (1984). Characterization of Barmah forest virus: an alphavirus with some unusual properties. *Virology,* **133**:416-426.

Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.F., Guindon S., Lefort V., Lescot M., Claverie J.M., and Gascuel O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research,* **36**(Web Server issue):W465-9.

Djikeng A., Halpin R., Kuzmickas R., DePasse J., Feldblyum J., Sengamalay N., Afonso C., Zhang X., Anderson N. G., Ghedin E. and Spiro D. J. (2008). Viral genome sequencing by random priming **.** *BioMed Central Genomics;*, **9**:5 doi:10.1186/1471-2164-9-5.

Edgar RC. **(2004).** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* **32**(5):1792-7.

Enserink M. (**2006).** Infectious diseases. Massive outbreak draws fresh attention to little-known virus. *Science,* **311**:1085.

Flauhaut A. **(2007)**. Chikungunya—Indian Ocean update (32). ProMED. 2006 Oct 14 [cited 2007 Apr 13]. Available from http://www.promedmail.org, archive no. 20061014.20062953.

Felsenstein J. **(1985).** Confidence limits on phylogenies: an approach using the bootstrap. *Evolution,* **39**:783–791.

Fiers W., Contreras R., Duerinck F., Haegeman G., Iserentant D., Merregaert J., Min Jou W., Molemans F., Raeymaekers A., Van den Berghe A., Volckaert G., Ysebaert M. (1976). Complete nucleotide-sequence of bacteriophage MS2-RNA primary and secondary structure of replicase gene, *Nature*, **260**: 500-507.

Froussard P. (**1992).** A random-PCR method (rPCR) to construct whole cDNA library from
low amounts of RNA. *Nucleic acids research,* **20**(11):2900.

Ghedin E., Sengamalay N.A., Shumway M., Zaborsky J., Feldblyum T., Subbu V., Spiro D.J., Sitz J., Koo H., Bolotov P., Dmitry D., Tatiana T., Bao Y., Kirsten G., Taylor J., Lipman D. J., Claire M. F., Taubenberger J. K. and Salzberg S. L. (**2005).** Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature*, **437**(7062):1162-1166.

Griffin D.E. (2001). General introduction—alphaviruses—chapter 1. In: Knipe, P.M.H.D.M. (Ed.), Fields Virology. Philadelphia, pp. 917–962.

Hahn C. S., Lustig S., Strauss E. G. and Strauss J. H. (**1988)**. Western equine encephalitis virus is a recombinant virus. USA *Proceedings of National academy science*, **85:**5997-6001.

Holmes E.C., Ghedin E., Miller N., Taylor J., Bao Y., St George K., Grenfell B.T., Salzberg S.L., Fraser C.M., Lipman D.J., Taubenberger, J.K. (**2005).** Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS biology*, **3**(9):e300.

Jewhurst V.A., Todd D., Rowley H.M., Walker I.W., Weston J.H., McLoughlin M.F., Graham D.A. (2004). Detection and antigenic characterization of salmonid alphavirus isolates from sera obtained from farmed Atlantic salmon, Salmo salar L., and farmed rainbow trout, Oncorhynchus mykiss (Walbaum). *Journal of Fish Diseases*, **27** (3), 143–149.

Johnson, B. K. (1988). O'nyong-nyong virus disease. *In* T. P. Monath (ed.). The arboviruses: epidemiology and ecology, CRC Press, Inc., Boca Raton, Fla. **vol. III** p. 217-223.

Johnston R. E. and Peters C. J. (1996). Alphaviruses. In: B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Virology, 3rd edition Lippincott-Raven, New York, N.Y. p. 843-898.

Jones MS, Kapoor A, Lukashov VV, Simmonds P, Hecht F, Delwart E. (2005). New DNA viruses identified in patients with acute viral infection syndrome. *Journal of virology*, **79**(13):8230-8236.

Jupp, P. G. (**1986)**. Mosquitoes of Southern Africa. Ecogilde, South Africa pp 156

Karabatsos, N. (**1975)**. Antigenic relationships of group A arboviruses by plaque reduction neutralization testing. *American Journal of Tropical Medicine and Hygiene*, **24:**527-532.

Karabatsos, N. (ed.) (**1985**). International catalogue of arboviruses. *American Society of Tropical Medicine and Hygiene,* San Antonio, TX.

Kinney R.M. and Pfeffer M. (2001). Nucleotide sequence analyses of the 26S mRNAs of viruses of the genus Alphavirus. Arbovirus Diseases Branch, Centers for Disease Control and Prevention, P.O. Box 2087, Fort Collins, CO 80522, USA.

Kinney R. M., Pfeffer M., Tsuchiya K. R., Chang G. J. and Roehrig J. T. (1998). Nucleotide sequences of the 26S mRNAs of the viruses defining the Venezuelan equine encephalitis antigenic complex. *American Journal of Tropical Medicine and Hygiene,* **59**:952-964.

Kiyomasu T., Miyazawa T., Furuya T., Shibata R., Sakai H., Sakuragi J., Fukasawa M., Maki N., Hasegawa A., Mikami T. (**1991)**. Identification of feline immunodeficiency virus rev gene activity. *Journal of Virology,* **65** (8): 4539–4542.

Kokernot R. H., McIntosh B. M. and Worth C. B. (1961). Ndumu virus, a hitherto unknown agent, isolated from culicine mosquitoes collected in northern Natal. Union of South Africa. *American Journal of Tropical Medicine and Hygiene,* **10**: 383–386.

La Linn M., Gardner J.,Warrilow D., Darnell G.A., McMahon C.R., Field I., Hyatt A.D., Slade R.W. and Suhrbier A. (2001). Arbovirus of marine mammals: a new alphavirus isolated from the elephant seal louse, Lepidophthirus macrorhini. *Journal of Virology,* **75** (9), 4103–4109.

Levinson R. S., Strauss J. H. and Strauss E. G. (1990). Complete sequence of the genomic RNA of o'nyong-nyong virus and its use in the construction of alphavirus phylogenetic trees. *Virology,* **175:**110-123.

Marshall I. D., Woodroofe G. M. and Hirsch S. (1982). Viruses recovered from mosquitoes and wildlife serum collected in the Murray Valley of South-eastern Australia, February 1974, during an epidemic of encephalitis. *Australian Journal of Experimental Biology and Medical Science,* **60**:457-470.

Meissner J. D., Huang C. Y., Pfeffer M. and Kinney R. M. (1999). Sequencing of prototype viruses in the Venezuelan equine encephalitis antigenic complex. *Virus Research,* **64**:43-59

Mudur G. **(2006).** Failure to control mosquitoes has led to two fever epidemics in India. *British medical journal*, **333**:773.

Pfeffer M., Kinney R. M. and Kaaden O. R. (1998). The alphavirus 3'-nontranslated region: size heterogeneity and arrangement of repeated sequence elements. *Virology*, **240**: 100-108.

Powers A.M., Brault A.C., ShirakoY., Strauss E.G., Kang W., Strauss J.H. and Weaver S.C. (2001). Evolutionary relationships and systematics of the alphaviruses. The following generic references are cited in the most recent *ICTV* Report. *Journal of Virology*, **75**: 10118-10131.

Powers A. M., Brault A. C., Tesh R. B., and Weaver S. C.(2000). Re-emergence of chikungunya and o'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. *Journal of General Virology*, **81**:471-479.

Powers A. M and Logue C. H. (2007). Changing patterns of Chikungunya virus: re-emergence of a zoonotic arbovirus. *Journal of General Virology*, **88**: 2363-2377.

Powers A. M., Oberste M. S., Brault A. C., Rico-Hesse R., Schmura S. M., Smith J. F., Kang W., Sweeney W. P. and Weaver S. C. (**1997)**. Repeated emergence of epidemic/epizootic Venezuelan equine encephalitis from a single genotype of enzootic subtype ID virus. *Journal of Virology*, **71:**6697-6705.
Prescott L. (1993). Microbiology. Wm. C. Brown Publishers. 0-697-01372-3.

Reinert,J. F., Harbach R. E. and Kitching I. J. (2004). Phylogeny and classification of Aedini (Diptera: *Culicidae)*, based on morphological characters of all life stages. *Zoological Journal of the Linnean Society*, **142**: 289–368

Reyes GR, Kim JP. (**1991).** Sequence-independent, single-primer amplification (SISPA) of
complex DNA populations. *Molecular and cellular probes*, **5**(6):473-481.

Salas R. A., Garcia C. Z., Liria J., Barrera R., Navarro J. C., Medina G., Vasquez C., Fernandez Z., and Weaver S. C. (2001). Ecological studies of enzootic Venezuelan equine encephalitis in north-central Venezuela, 1997-1998. *American Journal of Tropical Medicine and Hygiene*, **64**:84-92[Abstract].

Sang R. C., **Ahmed O., Faye O., Kelly C.L.H., Yahaya A.A., Mmadi I., Toilibou A., Sergon K., Brown J., Agata N., Yakouide A., Ball M.D., Breiman R.F., Miller B.R. and Powers A.M.(2008).** Entomologic Investigations of a Chikungunya Virus Epidemic in the Union of the Comoros, 2005. *American Journal of Tropical Medicine and Hygiene;***78**:77-82.

Schlesinger S. and Schlesinger M. J. (1996). *Togaviridae*: the viruses and their replication. In: B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Fields' virology, 3rd Edition Raven Press, New York, N.Y. p. 825-842.

Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney MC, Lavenir R., Pardigon N., Reynes J., Pettinelli F., Biscornet L., Diancourt L., Michel S., Duquerroy S., Guigon G., Frenkiel M., Bréhin A., Cubito N., Desprès P., Kunst F., Rey F.A., Zeller H., Brisse S.(2006). Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *Public Library of Science Medicine*; **3** (7):e263

Shin S.Y., Kim K.S., Lee Y.S., Chung Y.S., Park K.S., Cheon D.S., Na B.K., Kang Y., Cheong H.M., Moon Y., Jee-Hye C., Hang-Eui C., Na-Young M., Jin-Sook S., Young-Hoon P., Youngmee J., Jae-Deuk Y., Chul-Yong S. and Kwang-Ho L.. (**2003).** Identification of enteroviruses by using monoclonal antibodies against a putative common epitope. *Journal of clinical microbiology*, **41**(7):3028-3034.

Sidwell R.W. and Smee D.F. (2003). Viruses of the Bunya- and Togaviridae families: potential as bioterrorism agents and means of control. *Antiviral Research*, **57** (1–2): 101–111.

Strauss J. H. and Strauss E. G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiological Reviews*, 58:491-562.

Tamura K, Dudley J, Nei M & Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24:**1596-1599**.**

Tesh R. B. (**1982)**. Arthritides caused by mosquito-borne viruses. *Annual Review of Medicine,* **33:**31-40.

Travassos da Rosa A.P., Turell M.J., Watts D.M., Powers A.M., Vasconcelos P.F., Jones J.W., Klein T.A., Dohm D.J., Shope R.E., Degallier N., Popov V.L., Russell K.L., Weaver S.C., Guzman H., Calampa C., Brault A.C., Lemon A.P. and Tesh R.B. (2001). Trocara virus: a newly recognized alphavirus (Togaviridae) isolated from mosquitoes in the Amazon Basin. *American Journal of Tropical Medicine and Hygiene,* **64** (1–2): 93–97.

Tsai T.E., Weaver S.C.and Monath T.P. (2002). Alphaviruses, In: Richman, D.D. (Ed.), Clinical Virology, 2nd ed. ASM Press, pp. 1177–1210.

van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC,
Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B. (**2004).** Identification of a new human coronavirus. *Nature medicine*, **10**(4):368-373.

Van Regenmortel, M. H. (1990). Virus species, a much overlooked but essential concept in virus classification. *Intervirology,* **31**:241-254.

Van Regenmortel M. H. V. (2000). Introduction to the species concept in virus taxonomy, p. 3-16. In M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeogh, C. R. Pringle, and R. B. Wickner (ed.), Virus taxonomy. Classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, Inc., San Diego, Calif.

Venter J.C., Remington K., Heidelberg J.F., Halpern A.L., Rusch D., Eisen J.A., Wu D., Paulsen I., Nelson K.E., Nelson W., Fouts D. E., Levy S., Knap A. H., Lomas M. W., Nealson K., Owen W., Peterson J., Hoffman J., Parsons R., Baden-Tillson H., Pfannkoch C., Yu-Hui R. and Smith H. O. (**2004).** Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**(5667):66-74.

Villoing S., Bearzotti M., Chilmonczyk S., Castric J. and Bremont M. (2000). Rainbow trout sleeping disease virus is an atypical alphavirus. *Journal of Virology,* **74** (1): 173–183.

Vogel P., Fritz D.L., Kuehl K., Davis K.J.and Geisbert T. (1997). The agents of biological warfare. *Journal of the American Medical Association,* **278** (5): 438–439.

Wages D.P., Ficken M.D., Guy J.S., Cummings T.S. and Jennings S.R. (1993). Egg-production drop in turkeys associated with alphaviruses: eastern equine encephalitis virus and Highlands J virus. *Avian diseases,* **37** (4): 1163–1166.

Wang E., Barrera R., Boshell J., Ferro C., Freier J. E., Navarro J. C., Salas R., Vasquez C., and Weaver S. C. (1999). Genetic and phenotypic changes accompanying the emergence of epizootic subtype IC Venezuelan equine encephalitis viruses from an enzootic subtype ID progenitor. *Journal of Virology,* **73**:4266-4271

Weaver S. C. (**1995)**. Evolution of alphaviruses. In: A. J. Gibbs, C. H. Calisher, and F. Garcia-Arenal (ed.), Molecular basis of virus evolution. Cambridge University Press, Cambridge, England. p. 501-530

Weaver S. C. (**1998)**. Recurrent emergence of Venezuelan equine encephalomyelitis, In: W. M. Scheld, and J. Hughes (ed.), Emerging infections I. ASM Press, Washington, D.C. p. 27-42.

Weaver S. C., Hagenbaugh A., Bellew L. A., Netesov S. V., Volchkov V. E., Chang G.-J. J., Clarke D. K., Gousset L., Scott T. W., Trent D. W. and Holland J. J. (**1993)**. A comparison of the nucleotide sequences of eastern and western equine encephalomyelitis viruses with those of other alphaviruses and related RNA viruses. *Virology,* **197:**375-390.

Weaver S. C., Bellew L. A., Hagenbaugh A., Mallampalli V., . Holland J. J and Scott T. W. (1994). Evolution of alphaviruses in the eastern equine encephalomyelitis complex. *Journal of Virology,* **68:**158-169.

Weaver S. C., Bellew L. A. and Rico-Hesse R. (**1992)**. Phylogenetic analysis of alphaviruses in the Venezuelan equine encephalitis complex and identification of the source of epizootic viruses. *Virology,* **191:**282-290.

Weaver S. C., Dalgarno L., Frey T. K., Huang H. V., Kinney R. M., Rice C. M., Roehrig J. T., Shope R. E. and Strauss E. G. (**2000)**. Family Togaviridae. In: M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeogh, C. R. Pringle, and R. B. Wickner (ed.), Virus taxonomy. Classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, Inc., San Diego, Calif. p. 879-889

Weaver S. C., Kang W., Shirako Y., Rumenapf T., Strauss E. G., and Strauss J. H. (**1997)**. Recombinational history and molecular evolution of western equine encephalomyelitis complex alphaviruses. *Journal of Virology,* **71:**613-623.

Weaver S. C., Hagenbaugh A., Bellew L. A. and Calisher C. H. (1992). Genetic characterization of an antigenic subtype of eastern equine encephalomyelitis virus. *Archives of Virology*, **127**:305-314

Weston J.H., Graham D.A., Branson E., Rowley H.M., Walker I.W., Jewhurst V.A., Jewhurst H.L., Todd D. (2005). Nucleotide sequence variation in salmonid alphaviruses from outbreaks of salmon pancreas disease and sleeping disease. *Diseases of Aquatic Organisms*, **66** (2): 105–111.

Weston J.H., Welsh M.D., McLoughlin M.F. and Todd D. (1999). Salmon pancreas disease virus, an alphavirus infecting farmed Atlantic salmon, Salmo salar L. *Virology,* **256**:188-195.

Womack, M. (1993). "The yellow fever mosquito, *Aedes aegypti*." Wing Beats **5** (4): 4.

World Health Organization. (**2006).** Chikungunya and dengue, south-west Indian Ocean. Weekly Epidemiology Records, **81:**105–16, *Virology*, **256** (2): 188.

APPENDICES

APPENDIX 1: Multiple sequence alignment for Ndumu virus SaAR strain and isolates under study

```
GSA_S1_936/1-4689
TTTAAGTTTGGGGCCATGATGAAGTCCGGAATGTTTCTGACGTTATTTGT 50
AMH_001346/1-4687
TTCAAGTTTGGGGCCATGATGAAGTCCGGAATGTTTCTGACGTTATTTGT 50
BAR_S2_3526/1-4689
TTTAAGTTTGGGGCCATGATGAAGTCCGGAATGTTTCTGACGTTATTTGT 50
BSA_S4_2268/1-4689
TTTAAGTTTGGGGCCATGATGAAGTCCGGAATGTTTCTGACGTTATTTGT 50
GSA_S5_4278/1-4689
TTTAAGTTTGGGGCCATGATGAAGTCCGGAATGTTTCTGACGTTATTTGT 50
gi|28193968|gb|AF339487.1|/1-4689
TTTAAGTTTGGGGCCATGATGAAGTCCGGAATGTTTCTGACGTTATTTGT 50
BAR_S2_3527/1-4689
TTTAAGTTTGGGGCCATGATGAAGTCCGGAATGTTTCTGACGTTATTTGT 50
BSA_S4_2265/1-4689
TTTAAGTTTGGGGCCATGATGAAGTCCGGAATGTTTCTGACGTTATTTGT 50
                                                  **
**********************************************

GSA_S1_936/1-4689
TAACACGCTGCTTAACGTCGTTATAGCTAGTCGCGTCCTGGAATCTAAAT 100
AMH_001346/1-4687
TAACACGCTGCTTAACGTCGTTATAGCTAGTCGCGTCCTGGAATCTAAAT 100
BAR_S2_3526/1-4689
TAACACGCTGCTTAACGTCGTTATAGCTAGTCGCGTCCTGGAATCTAAAT 100
BSA_S4_2268/1-4689
TAACACGCTGCTTAACGTCGTTATAGCTAGTCGCGTCCTGGAATCTAAAT 100
GSA_S5_4278/1-4689
TAACACGCTGCTTAACGTCGTTATAGCTAGTCGCGTCCTGGAATCTAAAT 100
gi|28193968|gb|AF339487.1|/1-4689
TAACACGCTGCTTAACGTCGTTATAGCTAGTCGCGTCCTAGAATCCAAAT 100
BAR_S2_3527/1-4689
TAACACGCTGCTTAACGTCGTTATAGCTAGTCGCGTCCTGGAATCTAAAT 100
BSA_S4_2265/1-4689
TAACACGCTGCTTAACGTCGTTATAGCTAGTCGCGTCCTGGAATCTAAAT 100

************************************ ***** ****

GSA_S1_936/1-4689
TGACGGGGTCGCGATGTGCCGCCTTCATTGGGGACGATAACATCGTGCAT 150
AMH_001346/1-4687
TGACGGGGTCGCGATGTGCCGCCTTCATTGGGGACGATAACATCGTGCAT 150
BAR_S2_3526/1-4689
TGACGGGGTCGCGATGTGCCGCCTTCATTGGGGACGATAACATCGTGCAT 150
BSA_S4_2268/1-4689
TGACGGGGTCGCGATGTGCCGCCTTCATTGGGGACGATAACATCGTGCAT 150
```

```
GSA_S5_4278/1-4689
TGACGGGGTCGCGATGTGCCGCCTTCATTGGGGACGATAACATCGTGCAT 150
gi|28193968|gb|AF339487.1|/1-4689
TGACGGGGTCGCGATGTGCCGCCTTCATTGGGGACGATAACATCGTGCAT 150
BAR_S2_3527/1-4689
TGACGGGGTCGCGATGTGCCGCCTTCATTGGGGACGATAACATCGTGCAT 150
BSA_S4_2265/1-4689
TGACGGGGTCGCGATGTGCCGCCTTCATTGGGGACGATAACATCGTGCAT 150

**************************************************

GSA_S1_936/1-4689
GGCGTGGTCTCAGATAAGTTGATGGCAGAAAGGTGTGCCACCTGGATGAA 200
AMH_001346/1-4687
GGCGTGGTCTCAGATAAGTTGATGGCAGAAAGGTGTGCCACCTGGATGAA 200
BAR_S2_3526/1-4689
GGCGTGGTCTCAGATAAGTTGATGGCAGAAAGGTGTGCCACCTGGATGAA 200
BSA_S4_2268/1-4689
GGCGTGGTCTCAGATAAGTTGATGGCAGAAAGGTGTGCCACCTGGATGAA 200
GSA_S5_4278/1-4689
GGCGTGGTCTCAGATAAGTTGATGGCAGAAAGGTGTGCCACCTGGATGAA 200
gi|28193968|gb|AF339487.1|/1-4689
GGCGTGGTCTCAGATAAGTTGATGGCAGAAAGGTGTGCCACCTGGATGAA 200
BAR_S2_3527/1-4689
GGCGTGGTCTCAGATAAGTTGATGGCAGAAAGGTGTGCCACCTGGATGAA 200
BSA_S4_2265/1-4689
GGCGTGGTCTCAGATAAGTTGATGGCAGAAAGGTGTGCCACCTGGATGAA 200

**************************************************

GSA_S1_936/1-4689
CATGGAGGTCAAAATTATCGACGCAGTCATCGGAGAAAAACACCCGTATT 250
AMH_001346/1-4687
CATGGAGGTCAAAATTATCGACGCAGTCATCGGAGAAAAACACCCGTATT 250
BAR_S2_3526/1-4689
CATGGAGGTCAAAATTATCGACGCAGTCATCGGAGAAAAACACCCGTATT 250
BSA_S4_2268/1-4689
CATGGAGGTCAAAATTATCGACGCAGTCATCGGAGAAAAACACCCGTATT 250
GSA_S5_4278/1-4689
CATGGAGGTCAAAATTATCGACGCAGTCATCGGAGAAAAACACCCGTATT 250
gi|28193968|gb|AF339487.1|/1-4689
CATGGAGGTGAAAATTATCGACGCAGTCATCGGAGAAAAACACCCGTATT 250
BAR_S2_3527/1-4689
CATGGAGGTCAAAATTATCGACGCAGTCATCGGAGAAAAACACCCGTATT 250
BSA_S4_2265/1-4689
CATGGAGGTCAAAATTATCGACGCAGTCATCGGAGAAAAACACCCGTATT 250
                                                *********
**************************************

GSA_S1_936/1-4689
TCTGCGGCGGGTTCATCCTACAGGATGCTGTGACCGGCACGGCGTGCCGA 300
AMH_001346/1-4687
TCTGCGGCGGGTTCATCCTACAGGATGCTGTGACCGGCACGGCGTGCCGA 300
```

```
BAR_S2_3526/1-4689
TCTGCGGCGGGTTCATCCTACAGGATGCTGTGACCGGCACGGCGTGCCGA 300
BSA_S4_2268/1-4689
TCTGCGGCGGGTTCATCCTACAGGATGCTGTGACCGGCACGGCGTGCCGA 300
GSA_S5_4278/1-4689
TCTGCGGCGGGTTCATCCTACAGGATGCTGTGACCGGCACGGCGTGCCGA 300
gi|28193968|gb|AF339487.1|/1-4689
TCTGCGGCGGGTTCATCCTACAGGATGCTGTGACCGGCACGGCGTGCCGA 300
BAR_S2_3527/1-4689
TCTGCGGCGGGTTCATCCTACAGGATGCTGTGACCGGCACGGCGTGCCGA 300
BSA_S4_2265/1-4689
TCTGCGGCGGGTTCATCCTACAGGATGCTGTGACCGGCACGGCGTGCCGA 300

************************************************

GSA_S1_936/1-4689
GTATCTGACCCACTGAAGAGACTGTTTAAGTTGGGTAAACCACTGCCTGC 350
AMH_001346/1-4687
GTATCCGACCCACTGAAGAGACTGTTTAAGTTGGGTAAACCACTGCCTGC 350
BAR_S2_3526/1-4689
GTATCTGACCCACTGAAGAGACTGTTTAAGTTGGGTAAACCACTGCCTGC 350
BSA_S4_2268/1-4689
GTATCTGACCCACTGAAGAGACTGTTTAAGTTGGGTAAACCACTGCCTGC 350
GSA_S5_4278/1-4689
GTATCTGACCCACTGAAGAGACTGTTTAAGTTGGGTAAACCACTGCCTGC 350
gi|28193968|gb|AF339487.1|/1-4689
GTATCCGACCCATTGAAGAGACTGTTTAAGTTGGGTAAACCACTGCCTGC 350
BAR_S2_3527/1-4689
GTATCTGACCCACTGAAGAGACTGTTTAAGTTGGGTAAACCACTGCCTGC 350
BSA_S4_2265/1-4689
GTATCTGACCCACTGAAGAGACTGTTTAAGTTGGGTAAACCACTGCCTGC 350
                                                *****  ******
************************************

GSA_S1_936/1-4689
GGACGATGAGCAGGATGAGGACCGCAGACGAGCACTCCGTGACGAGGTGA 400
AMH_001346/1-4687
GGATGATGAGCAGGATGAGGACCGCAGACGAGCACTCCGTGATGAGGTGA 400
BAR_S2_3526/1-4689
GGACGATGAGCAGGATGAGGACCGCAGACGAGCACTCCGTGACGAGGTGA 400
BSA_S4_2268/1-4689
GGACGATGAGCAGGATGAGGACCGCAGACGAGCACTCCGTGACGAGGTGA 400
GSA_S5_4278/1-4689
GGACGATGAGCAGGATGAGGACCGCAGACGAGCACTCCGTGACGAGGTGA 400
gi|28193968|gb|AF339487.1|/1-4689
GGACGATGAGCAGGATGAGGACCGCAGACGAGCACTCCGTGACGAGGTGA 400
BAR_S2_3527/1-4689
GGACGATGAGCAGGATGAGGACCGCAGACGAGCACTCCGTGACGAGGTGA 400
BSA_S4_2265/1-4689
GGACGATGAGCAGGATGAGGACCGCAGACGAGCACTCCGTGACGAGGTGA 400
                                                ***
************************************ *******
```

```
GSA_S1_936/1-4689
TGAAATGGTTTAGGGTAGGTCTGCGGTCTGAGGTGTGTGCCGCCGTTTAT 450
AMH_001346/1-4687
TGAAATGGTTTAGGGTAGGTCTGCGGTCTGAGGTGTGTGCCGCCGTTTAT 450
BAR_S2_3526/1-4689
TGAAATGGTTTAGGGTAGGTCTGCGGTCTGAGGTGTGTGCCGCCGTTTAT 450
BSA_S4_2268/1-4689
TGAAATGGTTTAGGGTAGGTCTGCGGTCTGAGGTGTGTGCCGCCGTTTAT 450
GSA_S5_4278/1-4689
TGAAATGGTTTAGGGTAGGTCTGCGGTCTGAGGTGTGTGCCGCCGTTTAT 450
gi|28193968|gb|AF339487.1|/1-4689
TGAGATGGTTTAGGGTAGGTCTGCGGTCTGAGGTGTGTGCTGCAGTTTAT 450
BAR_S2_3527/1-4689
TGAAATGGTTTAGGGTAGGTCTGCGGTCTGAGGTGTGTGCCGCCGTTTAT 450
BSA_S4_2265/1-4689
TGAAATGGTTTAGGGTAGGTCTGCGGTCTGAGGTGTGTGCCGCCGTTTAT 450
                                                  ***
********************************** ** ******

GSA_S1_936/1-4689
TCCAGGTACGGCGTGCAGGGGCTGGATGTTGCTTTGATGGCTATGGCAAC 500
AMH_001346/1-4687
TCCAGGTATGGCGTGCAGGGGCTGGATGTTGCTTTGATGGCTATGGCAAC 500
BAR_S2_3526/1-4689
TCCAGGTACGGCGTGCAGGGGCTGGATGTTGCTTTGATGGCTATGGCAAC 500
BSA_S4_2268/1-4689
TCCAGGTACGGCGTGCAGGGGCTGGATGTTGCTTTGATGGCTATGGCAAC 500
GSA_S5_4278/1-4689
TCCAGGTACGGCGTGCAGGGGCTGGATGTTGCTTTGATGGCTATGGCAAC 500
gi|28193968|gb|AF339487.1|/1-4689
TCCAGGTACGGCGTGCAGGGGCTGGATGTTGCTTTGATGGCTATGGCAAC 500
BAR_S2_3527/1-4689
TCCAGGTACGGCGTGCAGGGGCTGGATGTTGCTTTGATGGCTATGGCAAC 500
BSA_S4_2265/1-4689
TCCAGGTACGGCGTGCAGGGGCTGGATGTTGCTTTGATGGCTATGGCAAC 500
                                                  ********
****************************************

GSA_S1_936/1-4689
CTTGTCGAAGACTAGGAAGCACTTCGACATGATTAGGGGACCCGTAAGGG 550
AMH_001346/1-4687
CTTGTCGAAGACTAGGAAGCACTTCGACATGATTAGGGGACCCGTAAGGG 550
BAR_S2_3526/1-4689
CTTGTCGAAGACTAGGAAGCACTTCGACATGATTAGGGGACCCGTAAGGG 550
BSA_S4_2268/1-4689
CTTGTCGAAGACTAGGAAGCACTTCGACATGATTAGGGGACCCGTAAGGG 550
GSA_S5_4278/1-4689
CTTGTCGAAGACTAGGAAGCACTTCGACATGATTAGGGGACCCGTAAGGG 550
gi|28193968|gb|AF339487.1|/1-4689
CCTGTCGAAGACTAGGAAGCACTTCGACATGATTAGGGGACCCGTAAGGG 550
BAR_S2_3527/1-4689
CTTGTCGAAGACTAGGAAGCACTTCGACATGATTAGGGGACCCGTAAGGG 550
```

```
BSA_S4_2265/1-4689
CTTGTCGAAGACTAGGAAGCACTTCGACATGATTAGGGGACCCGTAAGGG 550
                                                  *
**********************************************

GSA_S1_936/1-4689
TTCTCTACGGTGGTCCTAAATTGTAGCAGCCACAGAGTGACGCTACTACA 600
AMH_001346/1-4687
TTCTCTACGGTGGTCCTAAATTGTAGCAGCCACAGAGTGACGCTACTACA 600
BAR_S2_3526/1-4689
TTCTCTACGGTGGTCCTAAATTGTAGCAGCCACAGAGTGACGCTACTACA 600
BSA_S4_2268/1-4689
TTCTCTACGGTGGTCCTAAATTGTGGCAGCCACAGAGTGACGCTACTACA 600
GSA_S5_4278/1-4689
TTCTCTACGGTGGTCCTAAATTGTAGCAGCCACAGAGTGACGCTACTACA 600
gi|28193968|gb|AF339487.1|/1-4689
TTCTCTACGGTGGTCCTAAATTGTAGCAGCCACAGAGTGACGCTACTACA 600
BAR_S2_3527/1-4689
TTCTCTACGGTGGTCCTAAATTGTAGCAGCCACAGAGTGACGCTACTACA 600
BSA_S4_2265/1-4689
TTCTCTACGGTGGTCCTAAATTGTAGCAGCCACAGAGTGACGCTACTACA 600

********************** ************************

GSA_S1_936/1-4689
CTGTGCCTGCTACGCCATGGATTTCTTCAGAGGACCTTTCGCTATGCCAT 650
AMH_001346/1-4687
CTGTGCCTGCTACGCCATGGATTTCTTCAGAGGACCTTTTGCTATGCCAT 650
BAR_S2_3526/1-4689
CTGTGCCTGCTACGCCATGGATTTCTTCAGAGGACCTTTCGCTATGCCAT 650
BSA_S4_2268/1-4689
CTGTGCCTGCTACGCCATGGATTTCTTCAGAGGACCTTTCGCTATGCCAT 650
GSA_S5_4278/1-4689
CTGTGCCTGCTACGCCATGGATTTCTTCAGAGGACCTTTCGCTATGCCAT 650
gi|28193968|gb|AF339487.1|/1-4689
CTGTGCCTGCTACGCCATGGATTTCTTCAGAGGACCTTTCGCTATGCCGT 650
BAR_S2_3527/1-4689
CTGTGCCTGCTACGCCATGGATTTCTTCAGAGGACCTTTCGCTATGCCAT 650
BSA_S4_2265/1-4689
CTGTGCCTGCTACGCCATGGATTTCTTCAGAGGACCTTTCGCTATGCCAT 650

************************************ ******** *

GSA_S1_936/1-4689
TCCGCCAGCCTTACCGCCGCCCGTTTGCAATGCCGCCACGCCCGCGACCT 700
AMH_001346/1-4687
TCCGCCAGCCTTACCGCCGCCCGTTTGCAATGCCGCCACGCCCACGACCT 700
BAR_S2_3526/1-4689
TCCGCCAGCCTTACCGCCGCCCGTTTGCAATGCCGCCACGCCCGCGACCT 700
BSA_S4_2268/1-4689
TCCGCCAGCCTTACCGCCGCCCGTTTGCAATGCCGCCACGCCCGCGACCT 700
GSA_S5_4278/1-4689
TCCGCCAGCCTTACCGCCGCCCGTTTGCAATGCCGCCACGCCCGCGACCT 700
```

```
gi|28193968|gb|AF339487.1|/1-4689
TCCGCCAGCCTTACCGCCGCCCGTTTGCAATGCCGCCACGCCCACGACCT 700
BAR_S2_3527/1-4689
TCCGCCAGCCTTACCGCCGCCCGTTTGCAATGCCGCCACGCCCGCGACCT 700
BSA_S4_2265/1-4689
TCCGCCAGCCTTACCGCCGCCCGTTTGCAATGCCGCCACGCCCGCGACCT 700

****************************************** ******

GSA_S1_936/1-4689
AGGTTTAACCCGGTTGTGGAACAAGGTAGGCAAATTCAACAACTTATACA 750
AMH_001346/1-4687
AGGTTTAACCCGGTTGTGGAACAAGGTAGGCAGATTCAACAACTTATACA 750
BAR_S2_3526/1-4689
AGGTTTAACCCGGTTGTGGAACAAGGTAGGCAAATTCAACAACTTATACA 750
BSA_S4_2268/1-4689
AGGTTTAACCCGGTTGTGGAACAAGGTAGGCAAATTCAACAACTTATACA 750
GSA_S5_4278/1-4689
AGGTTTAACCCGGTTGTGGAACAAGGTAGGCAAATTCAACAACTTATACA 750
gi|28193968|gb|AF339487.1|/1-4689
AGGTTTAACCCGGTTGTGGAACAAGGTAGGCAAATTCAACAACTAATACA 750
BAR_S2_3527/1-4689
AGGTTTAACCCGGTTGTGGAACAAGGTAGGCAAATTCAACAACTTATACA 750
BSA_S4_2265/1-4689
AGGTTTAACCCGGTTGTGGAACAAGGTAGGCAAATTCAACAACTTATACA 750

******************************** *********** *****

GSA_S1_936/1-4689
GGCGGTAGGTAGTCTGGCGCTTGCGCAAAGACCGCGAGCACAGCCGAGCA 800
AMH_001346/1-4687
GGCGGTAGGTAGTCTGGCGCTTGCGCAAAGACCGCGCGCACAGCCGAGCA 800
BAR_S2_3526/1-4689
GGCGGTAGGTAGTCTGGCGCTTGCGCAAAGACCGCGAGCACAGCCGAGCA 800
BSA_S4_2268/1-4689
GGCGGTAGGTAGTCTGGCGCTTGCGCAAAGACCGCGAGCACAGCCGAGCA 800
GSA_S5_4278/1-4689
GGCGGTAGGTAGTCTGGCGCTTGCGCAAAGACCGCGAGCACAGCCGAGCA 800
gi|28193968|gb|AF339487.1|/1-4689
GGCGGTAGGTAGTCTGGCGCTTGCGCAAAGACCGCGCGCACAGCCGAGCA 800
BAR_S2_3527/1-4689
GGCGGTAGGTAGTCTGGCGCTTGCGCAAAGACCGCGAGCACAGCCGAGCA 800
BSA_S4_2265/1-4689
GGCGGTAGGTAGTCTGGCGCTTGCGCAAAGACCGCGAGCACAGCCGAGCA 800

********************************** *************

GSA_S1_936/1-4689
GACCACCGCGTCAGAACCGAAGAAGAAAGCCCCAACAACCAACGCAGTCT 850
AMH_001346/1-4687
GACCACCGCGTCAGAACCGAAGAAGAAAGCCCCAACAACCAACGCAGTCT 850
BAR_S2_3526/1-4689
GACCACCGCGTCAGAACCGAAGAAGAAAGCCCCAACAACCAACGCAGTCT 850
```

```
BSA_S4_2268/1-4689
GACCACCGCGTCAGAACCGAAGAAGAAAGCCCCAACAACCAACGCAGTCT 850
GSA_S5_4278/1-4689
GACCACCGCGTCAGAACCGAAGAAGAAAGCCCCAACAACCAACGCAGTCT 850
gi|28193968|gb|AF339487.1|/1-4689
GACCGCCGCGTCAGAACCGAAGAAGAAAGCCACAACAACCAACGCAGTCT 850
BAR_S2_3527/1-4689
GACCACCGCGTCAGAACCGAAGAAGAAAGCCCCAACAACCAACGCAGTCT 850
BSA_S4_2265/1-4689
GACCACCGCGTCAGAACCGAAGAAGAAAGCCCCAACAACCAACGCAGTCT 850
                                                ****
************************** ******************

GSA_S1_936/1-4689
CAGCAAAAGAGCAAGAAGCAGAAACCAAAGAGCAACCAGAAGAACGCTCA 900
AMH_001346/1-4687
CAGCAAAAGAGCAAGAAGCAGAAACCAAAGAGTAACCAGAAGAACGCTCA 900
BAR_S2_3526/1-4689
CAGCAAAAGAGCAAGAAGCAGAAACCAAAGAGCAACCAGAAGAACGCTCA 900
BSA_S4_2268/1-4689
CAGCAAAAGAGCAAGAAGCAGAAACCAAAGAGCAACCAGAAGAACGCTCA 900
GSA_S5_4278/1-4689
CAGCAAAAGAGCAAGAAGCAGAAACCAAAGAGCAACCAGAAGAACGCTCA 900
gi|28193968|gb|AF339487.1|/1-4689
CAGCAAAAGAGCAAGAAGCAGAAACCAAAGAGCAATCAGAAGAATGCTCA 900
BAR_S2_3527/1-4689
CAGCAAAAGAGCAAGAAGCAGAAACCAAAGAGCAACCAGAAGAACGCTCA 900
BSA_S4_2265/1-4689
CAGCAAAAGAGCAAGAAGCAGAAACCAAAGAGCAACCAGAAGAACGCTCA 900

******************************** ** ******** *****

GSA_S1_936/1-4689
AGTGCAAACAAAGAAACAGGATAAGACAAAGAAGAAGCCGGGTAAAAGGG 950
AMH_001346/1-4687
AGTGCAAACAAAGAAACAGGATAAGACAAAGAAGAAGCCGGGTAAAAGGG 950
BAR_S2_3526/1-4689
AGTGCAAACAAAGAAACAGGATAAGACAAAGAAGAAGCCGGGTAAAAGGG 950
BSA_S4_2268/1-4689
AGTGCAAACAAAGAAACAGGATAAGACAAAGAAGAAGCCGGGTAAAAGGG 950
GSA_S5_4278/1-4689
AGTGCAAACAAAGAAACAGGATAAGACAAAGAAGAAGCCGGGTAAAAGGG 950
gi|28193968|gb|AF339487.1|/1-4689
AGTGCAAACAAAGAAACAGGATAAGACAAAGAAGAAGCCGGGTAAAAGGG 950
BAR_S2_3527/1-4689
AGTGCAAACAAAGAAACAGGATAAGACAAAGAAGAAGCCGGGTAAAAGGG 950
BSA_S4_2265/1-4689
AGTGCAAACAAAGAAACAGGATAAGACAAAGAAGAAGCCGGGTAAAAGGG 950

************************************************

GSA_S1_936/1-4689
AAAGGAAGTGCATGAAAATAGAGAGTGATTGTATCTTTGAAGTGAAGCTA 1000
```

```
AMH_001346/1-4687
AAAGGAAGTGCATGAAAATAGAGAGTGATTGTATCTTTGAAGTGAAGCTA 1000
BAR_S2_3526/1-4689
AAAGGAAGTGCATGAAAATAGAGAGTGATTGTATCTTTGAAGTGAAGCTA 1000
BSA_S4_2268/1-4689
AAAGGAAGTGCATGAAAATAGAGAGTGATTGTATCTTTGAAGTGAAGCTA 1000
GSA_S5_4278/1-4689
AAAGGAAGTGCATGAAAATAGAGAGTGATTGTATCTTTGAAGTGAAGCTA 1000
gi|28193968|gb|AF339487.1|/1-4689
AAAGGAAGTGTATGAAAATAGAGAGTGATTGTATCTTTGAAGTGAAGTTA 1000
BAR_S2_3527/1-4689
AAAGGAAGTGCATGAAAATAGAGAGTGATTGTATCTTTGAAGTGAAGCTA 1000
BSA_S4_2265/1-4689
AAAGGAAGTGCATGAAAATAGAGAGTGATTGTATCTTTGAAGTGAAGCTA 1000
                                                 **********
********************************** **

GSA_S1_936/1-4689
GATGGAGAGGTGACAGGATATGCATGTTTGGTTGGGGATAAAGTGATGAA 1050
AMH_001346/1-4687
GATGGAGTGGTGACAGGATATGCATGTTTGGTTGGGGATAAAGTGATGAA 1050
BAR_S2_3526/1-4689
GATGGAGAGGTGACAGGATATGCATGTTTGGTTGGGGATAAAGTGATGAA 1050
BSA_S4_2268/1-4689
GATGGAGAGGTGACAGGATATGCATGTTTGGTTGGGGATAAAGTGATGAA 1050
GSA_S5_4278/1-4689
GATGGAGAGGTGACAGGATATGCATGTTTGGTTGGGGATAAAGTGATGAA 1050
gi|28193968|gb|AF339487.1|/1-4689
GATGGAGAGGTGACGGGATATGCATGTTTGGTTGGGGATAAAGTGATGAA 1050
BAR_S2_3527/1-4689
GATGGAGAGGTGACAGGATATGCATGTTTGGTTGGGGATAAAGTGATGAA 1050
BSA_S4_2265/1-4689
GATGGAGAGGTGACAGGATATGCATGTTTGGTTGGGGATAAAGTGATGAA 1050
                                                 ******* ******
**********************************

GSA_S1_936/1-4689
ACCAGCCCACGTGAAAGGGGTTATCCAAAATGATGACCTTGCTAAACTCA 1100
AMH_001346/1-4687
ACCAGCCCACGTGAAAGGGGTTATCCAAAATGATGACCTTGCTAAACTCA 1100
BAR_S2_3526/1-4689
ACCAGCCCACGTGAAAGGGGTTATCCAAAATGATGACCTTGCTAAACTCA 1100
BSA_S4_2268/1-4689
ACCAGCCCACGTGAAAGGGGTTATCCAAAATGATGACCTTGCTAAACTCA 1100
GSA_S5_4278/1-4689
ACCAGCCCACGTGAAAGGGGTTATCCAAAATGATGACCTTGCTAAACTCA 1100
gi|28193968|gb|AF339487.1|/1-4689
ACCAGCCCACGTAAAAGGGGTTATCCAAAATGATGAACTTGCTAAACTCA 1100
BAR_S2_3527/1-4689
ACCAGCCCACGTGAAAGGGGTTATCCAAAATGATGACCTTGCTAAACTCA 1100
BSA_S4_2265/1-4689
ACCAGCCCACGTGAAAGGGGTTATCCAAAATGATGACCTTGCTAAACTCA 1100
```

```
                                                          ************
********************** *************

GSA_S1_936/1-4689
GCTTCAAGAAATCTAGCAAGTATGACCTAGAGTGTGCACAGATACCAGTG 1150
AMH_001346/1-4687
GCCTCAAGAAATCTAGCAAGTATGACCTAGAGTGTGCACAGATACCGGTG 1150
BAR_S2_3526/1-4689
GCTTCAAGAAATCTAGCAAGTATGACCTAGAGTGTGCACAGATACCAGTG 1150
BSA_S4_2268/1-4689
GCTTCAAGAAATCTAGCAAGTATGACCTAGAGTGTGCACAGATACCAGTG 1150
GSA_S5_4278/1-4689
GCTTCAAGAAATCTAGCAAGTATGACCTAGAGTGTGCACAGATACCAGTG 1150
gi|28193968|gb|AF339487.1|/1-4689
GCTTCAAGAAATCTAGCAAGTATGACCTAGAGTGTGCACAGATACCGGTG 1150
BAR_S2_3527/1-4689
GCTTCAAGAAATCTAGCAAGTATGACCTAGAGTGTGCACAGATACCAGTG 1150
BSA_S4_2265/1-4689
GCTTCAAGAAATCTAGCAAGTATGACCTAGAGTGTGCACAGATACCAGTG 1150
                                                        **
****************************************** ***

GSA_S1_936/1-4689
AAGATGCGTTCTGATGCATCCAAGTACACTCACGAGAAGCCAGAGGGTCA 1200
AMH_001346/1-4687
AAGATGCGTTCTGATGCATCCAAGTACACTCACGAGAAGCCTGAGGGTCA 1200
BAR_S2_3526/1-4689
AAGATGCGTTCTGATGCATCCAAGTACACTCACGAGAAGCCAGAGGGTCA 1200
BSA_S4_2268/1-4689
AAGATGCGTTCTGATGCATCCAAGTACACTCACGAGAAGCCAGAGGGTCA 1200
GSA_S5_4278/1-4689
AAGATGCGTTCTGATGCATCCAAGTACACTCACGAGAAGCCAGAGGGTCA 1200
gi|28193968|gb|AF339487.1|/1-4689
AAGATGCGTTCCGATGCATCCAAGTACACTCACGAGAAGCCTGAGGGTCA 1200
BAR_S2_3527/1-4689
AAGATGCGTTCTGATGCATCCAAGTACACTCACGAGAAGCCAGAGGGTCA 1200
BSA_S4_2265/1-4689
AAGATGCGTTCTGATGCATCCAAGTACACTCACGAGAAGCCAGAGGGTCA 1200
                                                    ***********
**************************** ********

GSA_S1_936/1-4689
TTACAATTGGCACCACGGAGCAGTCCAGTATACTAACGGAAGGTTCACCA 1250
AMH_001346/1-4687
TTACAATTGGCACCACGGAGCAGTCCAGTATACTAACGGAAGGTTCACCA 1250
BAR_S2_3526/1-4689
TTACAATTGGCACCACGGAGCAGTCCAGTATACTAACGGAAGGTTCACCA 1250
BSA_S4_2268/1-4689
TTACAATTGGCACCACGGAGCAGTCCAGTATACTAACGGAAGGTTCACCA 1250
GSA_S5_4278/1-4689
TTACAATTGGCACCACGGAGCAGTCCAGTATACTAACGGAAGGTTCACCA 1250
gi|28193968|gb|AF339487.1|/1-4689
TTACAATTGGCACCACGGAGCAGTCCAGTATACTAACGGAAGGTTCACCA 1250
```

```
BAR_S2_3527/1-4689
TTACAATTGGCACCACGGAGCAGTCCAGTATACTAACGGAAGGTTCACCA 1250
BSA_S4_2265/1-4689
TTACAATTGGCACCACGGAGCAGTCCAGTATACTAACGGAAGGTTCACCA 1250

**************************************************

GSA_S1_936/1-4689
TTCCGACTGGAGCAGGCAAGCCAGGAGATAGCGGTAGACCGATCTTCGAT 1300
AMH_001346/1-4687
TTCCGACTGGAGCAGGCAAGCCAGGAGATAGCGGTAGACCGATCTTCGAT 1300
BAR_S2_3526/1-4689
TTCCGACTGGAGCAGGCAAGCCAGGAGATAGCGGTAGACCGATCTTCGAT 1300
BSA_S4_2268/1-4689
TTCCGACTGGAGCAGGCAAGCCAGGAGATAGCGGTAGACCGATCTTCGAT 1300
GSA_S5_4278/1-4689
TTCCGACTGGAGCAGGCAAGCCAGGAGATAGCGGTAGACCGATCTTCGAT 1300
gi|28193968|gb|AF339487.1|/1-4689
TTCCGACTGGAGCAGGCAAGCCAGGAGATAGCGGTAGACCGATCTTCGAT 1300
BAR_S2_3527/1-4689
TTCCGACTGGAGCAGGCAAGCCAGGAGATAGCGGTAGACCGATCTTCGAT 1300
BSA_S4_2265/1-4689
TTCCGACTGGAGCAGGCAAGCCAGGAGATAGCGGTAGACCGATCTTCGAT 1300

**************************************************

GSA_S1_936/1-4689
AACAAGGGAAGAGTCGTCGCAATAGTGCTGGGAGGGGCCAACGAAGGAGC 1350
AMH_001346/1-4687
AACAAGGGAAGAGTCGTCGCAATAGTGCTGGGAGGGGCCAACGAAGGAGC 1350
BAR_S2_3526/1-4689
AACAAGGGAAGAGTCGTCGCAATAGTGCTGGGAGGGGCCAACGAAGGAGC 1350
BSA_S4_2268/1-4689
AACAAGGGAAGAGTCGTCGCAATAGTGCTGGGAGGGGCCAACGAAGGAGC 1350
GSA_S5_4278/1-4689
AACAAGGGAAGAGTCGTCGCAATAGTGCTGGGAGGGGCCAACGAAGGAGC 1350
gi|28193968|gb|AF339487.1|/1-4689
AACAAGGGAAGAGTCGTCGCAATAGTGCTGGGAGGGGCCAACGAAGGAGC 1350
BAR_S2_3527/1-4689
AACAAGGGAAGAGTCGTCGCAATAGTGCTGGGAGGGGCCAACGAAGGAGC 1350
BSA_S4_2265/1-4689
AACAAGGGAAGAGTCGTCGCAATAGTGCTGGGAGGGGCCAACGAAGGAGC 1350

**************************************************

GSA_S1_936/1-4689
CCGAACAGCTCTATCAGTGGTCACGTGGAACAAAGACATGGTGACCAGGG 1400
AMH_001346/1-4687
CCGAACAGCTCTATCAGTGGTCACGTGGAACAAAGACATGGTGACCAGGG 1400
BAR_S2_3526/1-4689
CCGAACAGCTCTATCAGTGGTCACGTGGAACAAAGACATGGTGACCAGGG 1400
BSA_S4_2268/1-4689
CCGAACAGCTCTATCAGTGGTCACGTGGAACAAAGACATGGTGACCAGGG 1400
```

```
GSA_S5_4278/1-4689
CCGAACAGCTCTATCAGTGGTCACGTGGAACAAAGACATGGTGACCAGGG 1400
gi|28193968|gb|AF339487.1|/1-4689
CCGAACAGCTCTATCAGTGGTCACGTGGAACAAAGACATGGTGACCAGGG 1400
BAR_S2_3527/1-4689
CCGAACAGCTCTATCAGTGGTCACGTGGAACAAAGACATGGTGACCAGGG 1400
BSA_S4_2265/1-4689
CCGAACAGCTCTATCAGTGGTCACGTGGAACAAAGACATGGTGACCAGGG 1400

**************************************************

GSA_S1_936/1-4689
TTACACCTGAGGGATCAGTCGAATGGTCCGCGGCGGCCACCTGTGTCCTT 1450
AMH_001346/1-4687
TTACACCTGAGGGATCCGTCGAATGGTCCGCGGCGGCCACCTGCGTCCTT 1450
BAR_S2_3526/1-4689
TTACACCTGAGGGATCAGTCGAATGGTCCGCGGCGGCCACCTGTGTCCTT 1450
BSA_S4_2268/1-4689
TTACACCTGAGGGATCAGTCGAATGGTCCGCGGCGGCCACCTGTGTCCTT 1450
GSA_S5_4278/1-4689
TTACACCTGAGGGATCAGTCGAATGGTCCGCGGCGGCCACCTGTGTCCTT 1450
gi|28193968|gb|AF339487.1|/1-4689
TTACACCAGAGGGATCCGTCGAATGGTCCGCGGCGGCCACCTGCGTTCTT 1450
BAR_S2_3527/1-4689
TTACACCTGAGGGATCAGTCGAATGGTCCGCGGCGGCCACCTGTGTCCTT 1450
BSA_S4_2265/1-4689
TTACACCTGAGGGATCAGTCGAATGGTCCGCGGCGGCCACCTGTGTCCTT 1450
                                                 ******* ********
************************* ** ***

GSA_S1_936/1-4689
GGTGCTGCCATCTTTTCCTGCCTGTCTCCGCCTGTGTGCTATGATGAAGA 1500
AMH_001346/1-4687
GGTGCTGCCATCTTTTCCTGCCTGTCTCCGCCTGTGTGCTATGATGAAGA 1500
BAR_S2_3526/1-4689
GGTGCTGCCATCTTTTCCTGCCTGTCTCCGCCTGTGTGCTATGATGAAGA 1500
BSA_S4_2268/1-4689
GGTGCTGCCATCTTTTCCTGCCTGTCTCCGCCTGTGTGCTATGATGAAGA 1500
GSA_S5_4278/1-4689
GGTGCTGCCATCTTTTCCTGCCTGTCTCCGCCTGTGTGCTATGATGAAGA 1500
gi|28193968|gb|AF339487.1|/1-4689
GGTGCTGCCATCTTTTCCTGCCTGTCTCCGCCTGTGTGCTATGATGAAGA 1500
BAR_S2_3527/1-4689
GGTGCTGCCATCTTTTCCTGCCTGTCTCCGCCTGTGTGCTATGATGAAGA 1500
BSA_S4_2265/1-4689
GGTGCTGCCATCTTTTCCTGCCTGTCTCCGCCTGTGTGCTATGATGAAGA 1500

**************************************************

GSA_S1_936/1-4689
ACCGGAGAAAACTCTAACGATGCTGAGTATGAACGTGGATAACCCGGACT 1550
AMH_001346/1-4687
ACCGGAGAAAACTCTAACGATGCTGAGTATGAACGTGGATAACCCGGACT 1550
```

```
BAR_S2_3526/1-4689
ACCGGAGAAAACTCTAACGATGCTGAGTATGAACGTGGATAACCCGGACT 1550
BSA_S4_2268/1-4689
ACCGGAGAAAACTCTAACGATGCTGAGTATGAACGTGGATAACCCGGACT 1550
GSA_S5_4278/1-4689
ACCGGAGAAAACTCTAACGATGCTGAGTATGAACGTGGATAACCCGGACT 1550
gi|28193968|gb|AF339487.1|/1-4689
ACCGGAAAAAACTCTAACGATGCTGAGTATGAACGTGGATAACCCAGACT 1550
BAR_S2_3527/1-4689
ACCGGAGAAAACTCTAACGATGCTGAGTATGAACGTGGATAACCCGGACT 1550
BSA_S4_2265/1-4689
ACCGGAGAAAACTCTAACGATGCTGAGTATGAACGTGGATAACCCGGACT 1550
                                                ******
************************************ ****

GSA_S1_936/1-4689
ATTACTCGTTGTTTGAAGCCGCGCTTGACTGCCGCTCTCGTAGGCACCGC 1600
AMH_001346/1-4687
ATTACTCGTTGTTTGAGGCCGCGCTTGATTGCCGCTCTCGTAGGCACCGC 1600
BAR_S2_3526/1-4689
ATTACTCGTTGTTTGAAGCCGCGCTTGACTGCCGCTCTCGTAGGCACCGC 1600
BSA_S4_2268/1-4689
ATTACTCGTTGTTTGAAGCCGCGCTTGACTGCCGCTCTCGTAGGCACCGC 1600
GSA_S5_4278/1-4689
ATTACTCGTTGTTTGAAGCCGCGCTTGACTGCCGCTCTCGTAGGCACCGC 1600
gi|28193968|gb|AF339487.1|/1-4689
ATTACTCGTTGTTTGAAGCCGCGCTTGACTGCCGCTCTCGCAGGCACCGC 1600
BAR_S2_3527/1-4689
ATTACTCGTTGTTTGAAGCCGCGCTTGACTGCCGCTCTCGTAGGCACCGC 1600
BSA_S4_2265/1-4689
ATTACTCGTTGTTTGAAGCCGCGCTTGACTGCCGCTCTCGTAGGCACCGC 1600
                                                ****************
*********** *********** *********

GSA_S1_936/1-4689
AGAGCTGCTCAGCACTTCAACGAGTATAAAGTCACCCAGCCTTACATCGC 1650
AMH_001346/1-4687
AGAGCTGCTCAGCACTTCAACGAGTATAAGGTCACCCAGCCTTACATCGC 1650
BAR_S2_3526/1-4689
AGAGCTGCTCAGCACTTCAACGAGTATAAAGTCACCCAGCCTTACATCGC 1650
BSA_S4_2268/1-4689
AGAGCTGCTCAGCACTTCAACGAGTATAAAGTCACCCAGCCTTACATCGC 1650
GSA_S5_4278/1-4689
AGAGCTGCTCAGCACTTCAACGAGTATAAAGTCACCCAGCCTTACATCGC 1650
gi|28193968|gb|AF339487.1|/1-4689
AGAGCTGCTCAGCACTTCAACGAGTATAAGGTCACCCAGCCTTACATTGC 1650
BAR_S2_3527/1-4689
AGAGCTGCTCAGCACTTCAACGAGTATAAAGTCACCCAGCCTTACATCGC 1650
BSA_S4_2265/1-4689
AGAGCTGCTCAGCACTTCAACGAGTATAAAGTCACCCAGCCTTACATCGC 1650

**************************** ***************** **
```

```
GSA_S1_936/1-4689
TTACTGTCCGGACTGTGGTAATGGAGAGTCATGCCAGTCTCCGGTGGCTA 1700
AMH_001346/1-4687
TTACTGTCCGGACTGTGGTAATGGAGAGTCATGCCAGTCTCCGGTGGCTA 1700
BAR_S2_3526/1-4689
TTACTGTCCGGACTGTGGTAATGGAGAGTCATGCCAGTCTCCGGTGGCTA 1700
BSA_S4_2268/1-4689
TTACTGTCCGGACTGTGGTAATGGAGAGTCATGCCAGTCTCCGGTGGCTA 1700
GSA_S5_4278/1-4689
TTACTGTCCGGACTGTGGTAATGGAGAGTCATGCCAGTCTCCGGTGGCTA 1700
gi|28193968|gb|AF339487.1|/1-4689
TTACTGTCCGGACTGTGGTAATGGAGAGTCATGCCAGTCTCCGGTGGCTA 1700
BAR_S2_3527/1-4689
TTACTGTCCGGACTGTGGTAATGGAGAGTCATGCCAGTCTCCGGTGGCTA 1700
BSA_S4_2265/1-4689
TTACTGTCCGGACTGTGGTAATGGAGAGTCATGCCAGTCTCCGGTGGCTA 1700

**************************************************

GSA_S1_936/1-4689
TTGAAAGGATTCGTAATGAAGCTACAGATGGGACCCTGAAGGTGCAATTT 1750
AMH_001346/1-4687
TTGAAAGGATTCGTAATGAGGCTACAGATGGGACCCTGAAGGTGCAATTT 1750
BAR_S2_3526/1-4689
TTGAAAGGATTCGTAATGAAGCTACAGATGGGACCCTGAAGGTGCAATTT 1750
BSA_S4_2268/1-4689
TTGAAAGGATTCGTAATGAAGCTACAGATGGGACCCTGAAGGTGCAATTT 1750
GSA_S5_4278/1-4689
TTGAAAGGATTCGTAATGAAGCTACAGATGGGACCCTGAAGGTGCAATTT 1750
gi|28193968|gb|AF339487.1|/1-4689
TTGAAAGGATTCGCAATGAAGCTACAGATGGGACCCTGAAGGTGCAATTT 1750
BAR_S2_3527/1-4689
TTGAAAGGATTCGTAATGAAGCTACAGATGGGACCCTGAAGGTGCAATTT 1750
BSA_S4_2265/1-4689
TTGAAAGGATTCGTAATGAAGCTACAGATGGGACCCTGAAGGTGCAATTT 1750
                                          ************ *****
*****************************

GSA_S1_936/1-4689
GCCTTGCAGGTGGGGATTAAGAAAGATCACAGCACAGCACCTACTTACCT 1800
AMH_001346/1-4687
GCCTTGCAGGTGGGGATTAAGAAAGATCACAGCACAGCACCTACTTACCT 1800
BAR_S2_3526/1-4689
GCCTTGCAGGTGGGGATTAAGAAAGATCACAGCACAGCACCTACTTACCT 1800
BSA_S4_2268/1-4689
GCCTTGCAGGTGGGGATTAAGAAAGATCACAGCACAGCACCTACTTACCT 1800
GSA_S5_4278/1-4689
GCCTTGCAGGTGGGGATTAAGAAAGATCACAGCACAGCACCTACTTACCT 1800
gi|28193968|gb|AF339487.1|/1-4689
GCCTTGCAGGTGGGGATTAAGAAGGATCACAGCACTGCACCTACTTACCT 1800
BAR_S2_3527/1-4689
GCCTTGCAGGTGGGGATTAAGAAAGATCACAGCACAGCACCTACTTACCT 1800
```

```
BSA_S4_2265/1-4689
GCCTTGCAGGTGGGGATTAAGAAAGATCACAGCACAGCACCTACTTACCT 1800

********************** ********** ************

GSA_S1_936/1-4689
AAGGTACATGGAGGGACACGAGGTGAAGGATGCACTGAACAACCTCAGAG 1850
AMH_001346/1-4687
AAGGTACATGGAGGGACACGAGGTGAAGGATGCACTGAACAACCTTAGAG 1850
BAR_S2_3526/1-4689
AAGGTACATGGAGGGACACGAGGTGAAGGATGCACTGAACAACCTCAGAG 1850
BSA_S4_2268/1-4689
AAGGTACATGGAGGGACACGAGGTGAAGGATGCACTGAACAACCTCAGAG 1850
GSA_S5_4278/1-4689
AAGGTACATGGAGGGACACGAGGTGAAGGATGCACTGAACAACCTCAGAG 1850
gi|28193968|gb|AF339487.1|/1-4689
AAGGTACATGGAGGGACACGAGGTGAAGGATGCACTGAACAACCTCAGAG 1850
BAR_S2_3527/1-4689
AAGGTACATGGAGGGACACGAGGTGAAGGATGCACTGAACAACCTCAGAG 1850
BSA_S4_2265/1-4689
AAGGTACATGGAGGGACACGAGGTGAAGGATGCACTGAACAACCTCAGAG 1850

****************************************** ****

GSA_S1_936/1-4689
TGTATACTTCGGCCCAATGCGAGGTGACTGCGACTATGGGTCATTTTGCT 1900
AMH_001346/1-4687
TGTATACTTCGGCCCAATGCGAGGTGACTGCGACTATGGGTCATTTTGCT 1900
BAR_S2_3526/1-4689
TGTATACTTCGGCCCAATGCGAGGTGACTGCGACTATGGGTCATTTTGCT 1900
BSA_S4_2268/1-4689
TGTATACTTCGGCCCAATGCGAGGTGACTGCGACTATGGGTCATTTTGCT 1900
GSA_S5_4278/1-4689
TGTATACTTCGGCCCAATGCGAGGTGACTGCGACTATGGGTCATTTTGCT 1900
gi|28193968|gb|AF339487.1|/1-4689
TGTATACTTCGGCCCAATGTGAGGTGACTGCGACTATGGGTCATTTCGCT 1900
BAR_S2_3527/1-4689
TGTATACTTCGGCCCAATGCGAGGTGACTGCGACTATGGGTCATTTTGCT 1900
BSA_S4_2265/1-4689
TGTATACTTCGGCCCAATGCGAGGTGACTGCGACTATGGGTCATTTTGCT 1900
                                                  ******************
************************** ***

GSA_S1_936/1-4689
CTGGCTCGGTGTCCTCCTGGGGACAGCCTGAGTGTTACGTTCATGGACGC 1950
AMH_001346/1-4687
CTGGCTCGGTGTCCTCCTGGGGACAGCCTGAGTGTTACGTTCATGGACGC 1950
BAR_S2_3526/1-4689
CTGGCTCGGTGTCCTCCTGGGGACAGCCTGAGTGTTACGTTCATGGACGC 1950
BSA_S4_2268/1-4689
CTGGCTCGGTGTCCTCCTGGGGACAGCCTGAGTGTTACGTTCATGGACGC 1950
GSA_S5_4278/1-4689
CTGGCTCGGTGTCCTCCTGGGGACAGCCTGAGTGTTACGTTCATGGACGC 1950
```

```
gi|28193968|gb|AF339487.1|/1-4689
CTGGCTCGGTGTCCTCCTGGGGACAGCCTGAGTGTTACGTTCATGGACGC 1950
BAR_S2_3527/1-4689
CTGGCTCGGTGTCCTCCTGGGGACAGCCTGAGTGTTACGTTCATGGACGC 1950
BSA_S4_2265/1-4689
CTGGCTCGGTGTCCTCCTGGGGACAGCCTGAGTGTTACGTTCATGGACGC 1950

**************************************************

GSA_S1_936/1-4689
CTCCAAGAAAAACAGGATGTGTAGGGTCCCATTTGTACATAAGCTTCCCT 2000
AMH_001346/1-4687
CTCCAAGAAAAACAGGATGTGTAGGGTCCCATTTGTACATAAGCTTCCCT 2000
BAR_S2_3526/1-4689
CTCCAAGAAAAACAGGATGTGTAGGGTCCCATTTGTACATAAGCTTCCCT 2000
BSA_S4_2268/1-4689
CTCCAAGAAAAACAGGATGTGTAGGGTCCCATTTGTACATAAGCTTCCCT 2000
GSA_S5_4278/1-4689
CTCCAAGAAAAACAGGATGTGTAGGGTCCCATTTGTACATAAGCTTCCCT 2000
gi|28193968|gb|AF339487.1|/1-4689
CTCCAAGAAAAACAGGATGTGTAGGGTCCCATTTGTACATAAACTTCCCT 2000
BAR_S2_3527/1-4689
CTCCAAGAAAAACAGGATGTGTAGGGTCCCATTTGTACATAAGCTTCCCT 2000
BSA_S4_2265/1-4689
CTCCAAGAAAAACAGGATGTGTAGGGTCCCATTTGTACATAAGCTTCCCT 2000

****************************************** *******

GSA_S1_936/1-4689
TCTTAGGTAGAGAGAAAAACAGTGCCCGCCACTACCACGGCAAAGATGTT 2050
AMH_001346/1-4687
TCTTAGGCAGAGAGAAAAACAGTGCCCGCCACTACCACGGCAAAGATGTT 2050
BAR_S2_3526/1-4689
TCTTAGGTAGAGAGAAAAACAGTGCCCGCCACTACCACGGCAAAGATGTT 2050
BSA_S4_2268/1-4689
TCTTAGGTAGAGAGAAAAACAGTGCCCGCCACTACCACGGCAAAGATGTT 2050
GSA_S5_4278/1-4689
TCTTAGGTAGAGAGAAAAACAGTGCCCGCCACTACCACGGCAAAGATGTT 2050
gi|28193968|gb|AF339487.1|/1-4689
TCTTAGGCAGAGAGAAAAACAGTGCCCGCCGCTACCACGGCAAAGATGTT 2050
BAR_S2_3527/1-4689
TCTTAGGTAGAGAGAAAAACAGTGCCCGCCACTACCACGGCAAAGATGTT 2050
BSA_S4_2265/1-4689
TCTTAGGTAGAGAGAAAAATAGTGCCCGCCACTACCACGGCAAAGATGTT 2050
                                                 ******* ***********
********** ******************

GSA_S1_936/1-4689
ACGTGTACCACCTATGCCAGACGGACAGATGTCACAGATGAGAATCTGGA 2100
AMH_001346/1-4687
ACGTGTACCACCTATGCCAGACGGACAGATGTCACAGATGAGAATCTGGA 2100
BAR_S2_3526/1-4689
ACGTGTACCACCTATGCCAGACGGACAGATGTCACAGATGAGAATCTGGA 2100
```

```
BSA_S4_2268/1-4689
ACGTGTACCACCTATGCCAGACGGACAGATGTCACAGATGAGAATCTGGA 2100
GSA_S5_4278/1-4689
ACGTGTACCACCTATGCCAGACGGACAGATGTCACAGATGAGAATCTGGA 2100
gi|28193968|gb|AF339487.1|/1-4689
ACGTGTACCACCTATGCCAGACGGACAGATGTCACAGATGAGAACCTGGA 2100
BAR_S2_3527/1-4689
ACGTGTACCACCTATGCCAGACGGACAGATGTCACAGATGAGAATCTGGA 2100
BSA_S4_2265/1-4689
ACGTGTACCACCTATGCCAGACGGACAGATGTCACAGATGAGAATCTGGA 2100

******************************************** *****

GSA_S1_936/1-4689
GATGCATGTGCCACCGCACATACCGGATGCGACGCTTGTTAAAATACAAT 2150
AMH_001346/1-4687
GATGCATGTGCCACCGCACATACCGGACGCGACGCTTGTTAAAATACAAT 2150
BAR_S2_3526/1-4689
GATGCATGTGCCACCGCACATACCGGATGCGACGCTTGTTAAAATACAAT 2150
BSA_S4_2268/1-4689
GATGCATGTGCCACCGCACATACCGGATGCGACGCTTGTTAAAATACAAT 2150
GSA_S5_4278/1-4689
GATGCATGTGCCACCGCACATACCGGATGCGACGCTTGTTAAAATACAAT 2150
gi|28193968|gb|AF339487.1|/1-4689
GATGCATGTGCCACCGCACATACCGGACGCGACGCTTGTTAAAGTACAAT 2150
BAR_S2_3527/1-4689
GATGCATGTGCCACCGCACATACCGGATGCGACGCTTGTTAAAATACAAT 2150
BSA_S4_2265/1-4689
GATGCATGTGCCACCGCACATACCGGATGCGACGCTTGTTAAAATACAAT 2150

************************** ************** ******

GSA_S1_936/1-4689
CAAATAATGTTACAATCACACCGCCGTCGGGAGCGACGGTAAGGTATAAT 2200
AMH_001346/1-4687
CACATAATGTTACAATCACACCGCCGTCGGGAGCGACGGTAAGGTATAAT 2200
BAR_S2_3526/1-4689
CAAATAATGTTACAATCACACCGCCGTCGGGAGCGACGGTAAGGTATAAT 2200
BSA_S4_2268/1-4689
CAAATAATGTTACAATCACACCGCCGTCGGGAGCGACGGTAAGGTATAAT 2200
GSA_S5_4278/1-4689
CAAATAATGTTACAATCACACCGCCGTCGGGAGCGACGGTAAGGTATAAT 2200
gi|28193968|gb|AF339487.1|/1-4689
TAAATAATGTTACAATCACACCGCCGTCGGGAGCGACGGTGAGGTATAAT 2200
BAR_S2_3527/1-4689
CAAATAATGTTACAATCACACCGCCGTCGGGAGCGACGGTAAGGTATAAT 2200
BSA_S4_2265/1-4689
CAAATAATGTTACAATCACACCGCCGTCGGGAGCGACGGTAAGGTATAAT 2200
                                                  *
********************************** *********

GSA_S1_936/1-4689
TGTTCATGCAAAGACGGTAACAGGCAATCAGAGACCAGCCGTGAAGCCGT 2250
```

```
AMH_001346/1-4687
TGTTCATGCAAAGACGGTAACAGGCAATCAGAGACCAGCCGTGAAGCCTT 2250
BAR_S2_3526/1-4689
TGTTCATGCAAAGACGGTAACAGGCAATCAGAGACCAGCCGTGAAGCCGT 2250
BSA_S4_2268/1-4689
TGTTCATGCAAAGACGGTAACAGGCAATCAGAGACCAGCCGTGAAGCCGT 2250
GSA_S5_4278/1-4689
TGTTCATGCAAAGACGGTAACAGGCAATCAGAGACCAGCCGTGAAGCCGT 2250
gi|28193968|gb|AF339487.1|/1-4689
TGTTCATGCAAAGACGGTAACAGGCAATCAGAGACCAGCCGTGAAGTCGT 2250
BAR_S2_3527/1-4689
TGTTCATGCAAAGACGGTAACAGGCAATCAGAGACCAGCCGTGAAGCCGT 2250
BSA_S4_2265/1-4689
TGTTCATGCAAAGACGGTAACAGGCAATCAGAGACCAGCCGTGAAGCCGT 2250

********************************************** * *

GSA_S1_936/1-4689
GTTGTCCGGGTGCGCTGAAGCAAAGTGCCACGCCGCGGTGGTGGATGGGA 2300
AMH_001346/1-4687
GTTGTCCGGGTGCGCTGAAGCAAAGTGCCACGCCGCGGTGGTGGATGGGA 2300
BAR_S2_3526/1-4689
GTTGTCCGGGTGCGCTGAAGCAAAGTGCCACGCCGCGGTGGTGGATGGGA 2300
BSA_S4_2268/1-4689
GTTGTCCGGGTGCGCTGAAGCAAAGTGCCACGCCGCGGTGGTGGATGGGA 2300
GSA_S5_4278/1-4689
GTTGTCCGGGTGCGCTGAAGCAAAGTGCCACGCCGCGGTGGTGGATGGGA 2300
gi|28193968|gb|AF339487.1|/1-4689
GTTGTCCGGGTGCGCTGAAGCAAAGTGCCACGCCGCGGTGGTGGATGGGA 2300
BAR_S2_3527/1-4689
GTTGTCCGGGTGCGCTGAAGCAAAGTGCCACGCCGCGGTGGTGGATGGGA 2300
BSA_S4_2265/1-4689
GTTGTCCGGGTGCGCTGAAGCAAAGTGCCACGCCGCGGTGGTGGATGGGA 2300

**************************************************

GSA_S1_936/1-4689
AGGTGTGGCAATACAGCTCTAGATTCGTGCCTAGGCACGAAAAGACAGAG 2350
AMH_001346/1-4687
AGGTGTGGCAATATAGCTCTAGATTCGTGCCTAGGCATGAAAAGACAGAG 2350
BAR_S2_3526/1-4689
AGGTGTGGCAATACAGCTCTAGATTCGTGCCTAGGCACGAAAAGACAGAG 2350
BSA_S4_2268/1-4689
AGGTGTGGCAATACAGCTCTAGATTCGTGCCTAGGCACGAAAAGACAGAG 2350
GSA_S5_4278/1-4689
AGGTGTGGCAATACAGCTCTAGATTCGTGCCTAGGCACGAAAAGACAGAG 2350
gi|28193968|gb|AF339487.1|/1-4689
AGGTGTGGCAGTATAGCTCTAGATTCGTGCCTAGGCACGAAAAGACAGAG 2350
BAR_S2_3527/1-4689
AGGTGTGGCAATACAGCTCTAGATTCGTGCCTAGGCACGAAAAGACAGAG 2350
BSA_S4_2265/1-4689
AGGTGTGGCAATACAGCTCTAGATTCGTGCCTAGGCACGAAAAGACAGAG 2350
```

```
                                                          ********** **
********************** ************

GSA_S1_936/1-4689
CTAAAGGGTAAGGTACACGTTCCTTTCCCTTTGACTAACACCACATGTAG 2400
AMH_001346/1-4687
CTAAAGGGTAAGGTACACGTTCCTTTCCCTCTGACTAACACCACATGTAG 2400
BAR_S2_3526/1-4689
CTAAAGGGTAAGGTACACGTTCCTTTCCCTTTGACTAACACCACATGTAG 2400
BSA_S4_2268/1-4689
CTAAAGGGTAAGGTACACGTTCCTTTCCCTTTGACTAACACCACATGTAG 2400
GSA_S5_4278/1-4689
CTAAAGGGTAAGGTACACGTTCCTTTCCCTTTGACTAACACCACATGTAG 2400
gi|28193968|gb|AF339487.1|/1-4689
CTAAAGGGTAAGGTACACGTTCCTTTCCCTTTGACTAACACCACATGTAG 2400
BAR_S2_3527/1-4689
CTAAAGGGTAAGGTACACGTTCCTTTCCCTTTGACTAACACCACATGTAG 2400
BSA_S4_2265/1-4689
CTAAAGGGTAAGGTACACGTTCCTTTCCCTTTGACTAACACCACATGTAG 2400

***************************** ******************

GSA_S1_936/1-4689
AGTAAGTGTTGCAACTAAGCCAATAGAAAAACATGCTAGAGGATTGCTTA 2450
AMH_001346/1-4687
AGTAAGTGTTGCAGCTAAGCCAATAGAAAAACATGCTAGAGGATTGCTTA 2450
BAR_S2_3526/1-4689
AGTAAGTGTTGCAACTAAGCCAATAGAAAAACATGCTAGAGGATTGCTTA 2450
BSA_S4_2268/1-4689
AGTAAGTGTTGCAACTAAGCCAATAGAAAAACATGCTAGAGGATTGCTTA 2450
GSA_S5_4278/1-4689
AGTAAGTGTTGCAACTAAGCCAATAGAAAAACATGCTAGAGGATTGCTTA 2450
gi|28193968|gb|AF339487.1|/1-4689
AGTAAGTGTAGCAACTAAGCCAATAGAAAAACATGCTAGAGGATTGCTTA 2450
BAR_S2_3527/1-4689
AGTAAGTGTTGCAACTAAGCCAATAGAAAAACATGCTAGAGGATTGCTTA 2450
BSA_S4_2265/1-4689
AGTAAGTGTTGCAACTAAGCCAATAGAAAAACATGCTAGAGGATTGCTTA 2450
                                                  ********* ***
**********************************

GSA_S1_936/1-4689
CGCTAGTGTTGCATCCCGAATACCCGACCCTGCTTAGCTATAGATACTTA 2500
AMH_001346/1-4687
CGCTAGTGTTGCATCCCGAATACCCGACCCTACTTAGCTATAGATACTTA 2500
BAR_S2_3526/1-4689
CGCTAGTGTTGCATCCCGAATACCCGACCCTGCTTAGCTATAGATACTTA 2500
BSA_S4_2268/1-4689
CGCTAGTGTTGCATCCCGAATACCCGACCCTGCTTAGCTATAGATACTTA 2500
GSA_S5_4278/1-4689
CGCTAGTGTTGCATCCCGAATACCCGACCCTGCTTAGCTATAGATACTTA 2500
gi|28193968|gb|AF339487.1|/1-4689
CGCTAGTGTTGCATCCCGAATACCCGACCCTGCTTAGCTATAGATATTTA 2500
```

```
BAR_S2_3527/1-4689
CGCTAGTGTTGCATCCCGAATACCCGACCCTGCTTAGCTATAGATACTTA 2500
BSA_S4_2265/1-4689
CGCTAGTGTTGCATCCCGAATACCCGACCCTGCTTAGCTATAGATACTTA 2500

****************************** ************** ***

GSA_S1_936/1-4689
GGAGCAGAACCACGCCCGGTTAACAGATGGATATCAGAGCCAACGGAGGT 2550
AMH_001346/1-4687
GGAGCAGAACCACGCCCGGTTAACAGATGGATATCAGAGCCAACGGAGGT 2550
BAR_S2_3526/1-4689
GGAGCAGAACCACGCCCGGTTAACAGATGGATATCAGAGCCAACGGAGGT 2550
BSA_S4_2268/1-4689
GGAGCAGAACCACGCCCGGTTAACAGATGGATATCAGAGCCAACGGAGGT 2550
GSA_S5_4278/1-4689
GGAGCAGAACCACGCCCGGTTAACAGATGGATATCAGAGCCAACGGAGGT 2550
gi|28193968|gb|AF339487.1|/1-4689
GGAGCAGAACCACGCCCGGTTAACAGATGGATATCAGAGCCAACGGAGGT 2550
BAR_S2_3527/1-4689
GGAGCAGAACCACGCCCGGTTAACAGATGGATATCAGAGCCAACGGAGGT 2550
BSA_S4_2265/1-4689
GGAGCAGAACCACGCCCGGTTAACAGATGGATATCAGAGCCAACGGAGGT 2550

**************************************************

GSA_S1_936/1-4689
GCAGATTCCCGTGAGCAAAGAGGGGATAGAATATAGATGGGGCAACAATG 2600
AMH_001346/1-4687
GCAGATTCCCGTGAGCAAAGAGGGGATAGAATATAGATGGGGCAACAATG 2600
BAR_S2_3526/1-4689
GCAGATTCCCGTGAGCAAAGAGGGGATAGAATATAGATGGGGCAACAATG 2600
BSA_S4_2268/1-4689
GCAGATTCCCGTGAGCAAAGAGGGGATAGAATATAGATGGGGCAACAATG 2600
GSA_S5_4278/1-4689
GCAGATTCCCGTGAGCAAAGAGGGGATAGAATATAGATGGGGCAACAATG 2600
gi|28193968|gb|AF339487.1|/1-4689
GCAGATTCCCGTGAGCAAAGAGGGGATAGAATATAGATGGGGCAACAATG 2600
BAR_S2_3527/1-4689
GCAGATTCCCGTGAGCAAAGAGGGGATAGAATATAGATGGGGCAACAATG 2600
BSA_S4_2265/1-4689
GCAGATTCCCGTGAGCAAAGAGGGGATAGAATATAGATGGGGCAACAATG 2600

**************************************************

GSA_S1_936/1-4689
AACCTGAAAGAAGGTGGGCGCAGCACACGACCACCGGATCCGCGCATGGA 2650
AMH_001346/1-4687
AACCTGAAAGAAGGTGGGCGCAGCACACGACCACCGGATCCGCGCATGGA 2650
BAR_S2_3526/1-4689
AACCTGAAAGAAGGTGGGCGCAGCACACGACCACCGGATCCGCGCATGGA 2650
BSA_S4_2268/1-4689
AACCTGAAAGAAGGTGGGCGCAGCACACGACCACCGGATCCGCGCATGGA 2650
```

```
GSA_S5_4278/1-4689
AACCTGAAAGAAGGTGGGCGCAGCACACGACCACCGGATCCGCGCATGGA 2650
gi|28193968|gb|AF339487.1|/1-4689
AACCTGAAAGAAGGTGGGCGCAGCACACGACCACCGGATCCGCGCATGGA 2650
BAR_S2_3527/1-4689
AACCTGAAAGAAGGTGGGCGCAGCACACGACCACCGGATCCGCGCATGGA 2650
BSA_S4_2265/1-4689
AACCTGAAAGAAGGTGGGCGCAGCACACGACCACCGGATCCGCGCATGGA 2650

**************************************************

GSA_S1_936/1-4689
CACCCACACGAGATAATTACGTACTACTACCACTCACACCCGACAACCAC 2700
AMH_001346/1-4687
CACCCACACGAGATAATTACGTACTACTACCACTCACACCCGACAACCAC 2700
BAR_S2_3526/1-4689
CACCCACACGAGATAATTACGTACTACTACCACTCACACCCGACAACCAC 2700
BSA_S4_2268/1-4689
CACCCACACGAGATAATTACGTACTACTACCACTCACACCCGACAACCAC 2700
GSA_S5_4278/1-4689
CACCCACACGAGATAATTACGTACTACTACCACTCACACCCGACAACCAC 2700
gi|28193968|gb|AF339487.1|/1-4689
CACCCACACGAGATAATTACGTACTACTACCACTCACACCCGACAACCAC 2700
BAR_S2_3527/1-4689
CACCCACACGAGATAATTACGTACTACTACCACTCACACCCGACAACCAC 2700
BSA_S4_2265/1-4689
CACCCACACGAGATAATTACGTACTACTACCACTCACACCCGACAACCAC 2700

**************************************************

GSA_S1_936/1-4689
AGTGGTGGCGTGCGTGACGGCGGCTGTCGTGACTCTTGCAATGACGTGCG 2750
AMH_001346/1-4687
AGTGGTGGCGTGCGTGACGGCAGCTGTCGTGACTCTTGTAATGATGTGCG 2750
BAR_S2_3526/1-4689
AGTGGTGGCGTGCGTGACGGCGGCTGTCGTGACTCTTGCAATGATGTGCG 2750
BSA_S4_2268/1-4689
AGTGGTGGCGTGCGTGACGGCGGCTGTCGTGACTCTTGCAATGATGTGCG 2750
GSA_S5_4278/1-4689
AGTGGTGGCGTGCGTGACGGCGGCTGTCGTGACTCTTGCAATGATGTGCG 2750
gi|28193968|gb|AF339487.1|/1-4689
AGTGGTGGCGTGCGTGACGGCGGCTGCCGTGACTCTTGTAATGATGTGCG 2750
BAR_S2_3527/1-4689
AGTGGTGGCGTGCGTGACGGCGGCTGTCGTGACTCTTGCAATGATGTGCG 2750
BSA_S4_2265/1-4689
AGTGGTGGCGTGCGTGACGGCGGCTGTCGTGACTCTTGCAATGATGTGCG 2750
                                                    **********************
**** *********** ***** *****

GSA_S1_936/1-4689
TGGGGTGCAGTGCTTGTAGAGTCGCTAGAACACGATGCTTGACTCCATAT 2800
AMH_001346/1-4687
TGGGGTGCAGTGCTTGTAGAGTCGCTAGAACACGATGCTTGACTCCATAT 2800
```

```
BAR_S2_3526/1-4689
TGGGGTGCAGTGCTTGTAGAGTCGCTAGAACACGATGCTTGACTCCATAT 2800
BSA_S4_2268/1-4689
TGGGGTGCAGTGCTTGTAGAGTCGCTAGAACACGATGCTTGACTCCATAT 2800
GSA_S5_4278/1-4689
TGGGGTGCAGTGCTTGTAGAGTCGCTAGAACACGATGCTTGACTCCATAT 2800
gi|28193968|gb|AF339487.1|/1-4689
TGGGGTGCAGTGCTTGTAGAGTCGCTAGAACACGATGCTTGACTCCATAT 2800
BAR_S2_3527/1-4689
TGGGGTGCAGTGCTTGTAGAGTCGCTAGAACACGATGCTTGACTCCATAT 2800
BSA_S4_2265/1-4689
TGGGGTGCAGTGCTTGTAGAGTCGCTAGAACACGATGCTTGACTCCATAT 2800

**************************************************

GSA_S1_936/1-4689
GTACTGGCGCCGGGATCTCGCGTCCCTTTGATTCTGGGGCTGTTGTGCTG 2850
AMH_001346/1-4687
GTACTGGCGCCGGGATCTCGCGTCCCTTTGATTCTGGGGCTGTTGTGCTG 2850
BAR_S2_3526/1-4689
GTACTGGCGCCGGGATCTCGCGTCCCTTTGATTCTGGGGCTGTTGTGCTG 2850
BSA_S4_2268/1-4689
GTACTGGCGCCGGGATCTCGCGTCCCTTTGATTCTGGGGCTGTTGTGCTG 2850
GSA_S5_4278/1-4689
GTACTGGCGCCGGGATCTCGCGTCCCTTTGATTCTGGGGCTGTTGTGCTG 2850
gi|28193968|gb|AF339487.1|/1-4689
GTACTGGCGCCGGGATCTCGCGTCCCTCTGATTCTGGGGCTGTTGTGCTG 2850
BAR_S2_3527/1-4689
GTACTGGCGCCGGGATCTCGCGTCCCTTTGATTCTGGGGCTGTTGTGCTG 2850
BSA_S4_2265/1-4689
GTACTGGCGCCGGGATCTCGCGTCCCTTTGATTCTGGGGCTGTTGTGCTG 2850

************************** **********************

GSA_S1_936/1-4689
CGCCAAGGGTGCAAGGGCCGAACAATACGAGCCGACACCGTGGTGGGTTT 2900
AMH_001346/1-4687
CGCCAAGGGTGCAAGGGCCGAACAATACGAGCCGACACCGTGGTGGGTTT 2900
BAR_S2_3526/1-4689
CGCCAAGGGTGCAAGGGCCGAACAATACGAGCCGACACCGTGGTGGGTTT 2900
BSA_S4_2268/1-4689
CGCCAAGGGTGCAAGGGCCGAACAATACGAGCCGACACCGTGGTGGGTTT 2900
GSA_S5_4278/1-4689
CGCCAAGGGTGCAAGGGCCGAACAATACGAGCCGACACCGTGGTGGGTTT 2900
gi|28193968|gb|AF339487.1|/1-4689
CGCCAAGGGTGCAAGGGCCGAACAATACGAGCCGACACCGTGGTGGGTTT 2900
BAR_S2_3527/1-4689
CGCCAAGGGTGCAAGGGCCGAACAATACGAGCCGACACCGTGGTGGGTTT 2900
BSA_S4_2265/1-4689
CGCCAAGGGTGCAAGGGCCGAACAATACGAGCCGACACCGTGGTGGGTTT 2900

**************************************************
```

```
GSA_S1_936/1-4689
GGGCACCCGCTGCGTTGGTACTCCTATGTGGATTGAGGAAGTGTCTTTGC 2950
AMH_001346/1-4687
GGGCACCCGCTGCGTTGGTACTCCTATGTGGATTGAGGAAGTGTCTTTGC 2950
BAR_S2_3526/1-4689
GGGCACCCGCTGCGTTGGTACTCCTATGTGGATTGAGGAAGTGTCTTTGC 2950
BSA_S4_2268/1-4689
GGGCACCCGCTGCGTTGGTACTCCTATGTGGATTGAGGAAGTGTCTTTGC 2950
GSA_S5_4278/1-4689
GGGCACCCGCTGCGTTGGTACTCCTATGTGGATTGAGGAAGTGTCTTTGC 2950
gi|28193968|gb|AF339487.1|/1-4689
GGGCACCCGCTGCGTTGGTACTCCTATGTGGATTGAGGAAGTGTCTATGC 2950
BAR_S2_3527/1-4689
GGGCACCCGCTGCGTTGGTACTCCTATGTGGATTGAGGAAGTGTCTTTGC 2950
BSA_S4_2265/1-4689
GGGCACCCGCTGCGTTGGTACTCCTATGTGGATTGAGGAAGTGTCTTTGC 2950

******************************************** ***

GSA_S1_936/1-4689
TTGACTTTTTTAGTGATACTAGGCCTGGCCAGCCCACCGACCCAGGCCTA 3000
AMH_001346/1-4687
TTGACTTTTTTAGTGATACTGGGCCTGGCCAGCCCACCGACCCAGGCCTA 3000
BAR_S2_3526/1-4689
TTGACTTTTTTAGTGATACTAGGCCTGGCCAGCCCACCGACCCAGGCCTA 3000
BSA_S4_2268/1-4689
TTGACTTTTTTAGTGATACTAGGCCTGGCCAGCCCACCGACCCAGGCCTA 3000
GSA_S5_4278/1-4689
TTGACTTTTTTAGTGATACTAGGCCTGGCCAGCCCACCGACCCAGGCCTA 3000
gi|28193968|gb|AF339487.1|/1-4689
TTGACTTTTTTAGTGATACTAGGCCTGGCCAGCCCACCGACCCAGGCCTA 3000
BAR_S2_3527/1-4689
TTGACTTTTTTAGTGATACTAGGCCTGGCCAGCCCACCGACCCAGGCCTA 3000
BSA_S4_2265/1-4689
TTGACTTTTTTAGTGATACTAGGCCTGGCCAGCCCACCGACCCAGGCCTA 3000
                                                    ********************
****************************

GSA_S1_936/1-4689
CGAGCACACGGCTGTGATGTCGAATCAGGTGGGAGTACCTTACAAGGCCC 3050
AMH_001346/1-4687
CGAGCACACGGCTGTGATGTCGAATCAGGTGGGAGTACCTTACAAGGCCC 3050
BAR_S2_3526/1-4689
CGAGCACACGGCTGTGATGTCGAATCAGGTGGGAGTACCTTACAAGGCCC 3050
BSA_S4_2268/1-4689
CGAGCACACGGCTGTGATGTCGAATCAGGTGGGAGTACCTTACAAGGCCC 3050
GSA_S5_4278/1-4689
CGAGCACACGGCTGTGATGTCGAATCAGGTGGGAGTACCTTACAAGGCCC 3050
gi|28193968|gb|AF339487.1|/1-4689
CGAGCACACGGCTGTGATGTCGAATCAGGTGGGAGTACCCTACAAGGCCC 3050
BAR_S2_3527/1-4689
CGAGCACACGGCTGTGATGTCGAATCAGGTGGGAGTACCTTACAAGGCCC 3050
```

```
BSA_S4_2265/1-4689
CGAGCACACGGCTGTGATGTCGAATCAGGTGGGAGTACCTTACAAGGCCC 3050

************************************* **********

GSA_S1_936/1-4689
TGATCAACAAGCCCGGATTTGCGCCCATGATTCTGCAAATAGAAGTGCTC 3100
AMH_001346/1-4687
TGATCAACAAGCCCGGATTTGCGCCCATGATCCTGCAAATAGAAGTGCTC 3100
BAR_S2_3526/1-4689
TGATCAACAAGCCCGGATTTGCGCCCATGATTCTGCAAATAGAAGTGCTC 3100
BSA_S4_2268/1-4689
TGATCAACAAGCCCGGATTTGCGCCCATGATTCTGCAAATAGAAGTGCTC 3100
GSA_S5_4278/1-4689
TGATCAACAAGCCCGGATTTGCGCCCATGATTCTGCAAATAGAAGTGCTC 3100
gi|28193968|gb|AF339487.1|/1-4689
TGATCAACAAGCCCGGATTTGCGCCCATGATTCTGCAAATAGAAGTGCTC 3100
BAR_S2_3527/1-4689
TGATCAACAAGCCCGGATTTGCGCCCATGATTCTGCAAATAGAAGTGCTC 3100
BSA_S4_2265/1-4689
TGATCAACAAGCCCGGATTTGCGCCCATGATTCTGCAAATAGAAGTGCTC 3100

****************************** ******************

GSA_S1_936/1-4689
CAGTCTAGTCTGATACCATCTTTGGAACTAGATTATATCACATGTGAGTA 3150
AMH_001346/1-4687
CAGTCTAGTCTGATACCGTCTTTGGAACTAGATTATATCACATGTGAGTA 3150
BAR_S2_3526/1-4689
CAGTCTAGTCTGATACCATCTTTGGAACTAGATTATATCACATGTGAGTA 3150
BSA_S4_2268/1-4689
CAGTCTAGTCTGATACCATCTTTGGAACTAGATTATATCACATGTGAGTA 3150
GSA_S5_4278/1-4689
CAGTCTAGTCTGATACCATCTTTGGAACTAGATTATATCACATGTGAGTA 3150
gi|28193968|gb|AF339487.1|/1-4689
CAGTCTAGTCTGATACCATCTTTGGAACTAGATTATATCACATGTGAGTA 3150
BAR_S2_3527/1-4689
CAGTCTAGTCTGATACCATCTTTGGAACTAGATTATATCACATGTGAGTA 3150
BSA_S4_2265/1-4689
CAGTCTAGTCTGATACCATCTTTGGAACTAGATTATATCACATGTGAGTA 3150
                                                  **********
********************************

GSA_S1_936/1-4689
CAAGACCGTAGTGCCTTCCCCTTTTGTCAAGTGCTGCGGATCTGTAGAGT 3200
AMH_001346/1-4687
CAAGACTGTAGTGCCTTCCCCTTTTGTCAAGTGCTGCGGATCCGTAGAAT 3200
BAR_S2_3526/1-4689
CAAGACCGTAGTGCCTTCCCCTTTTGTCAAGTGCTGCGGATCTGTAGAGT 3200
BSA_S4_2268/1-4689
CAAGACCGTAGTGCCTTCCCCTTTTGTCAAGTGCTGCGGATCTGTAGAGT 3200
GSA_S5_4278/1-4689
CAAGACCGTAGTGCCTTCCCCTTTTGTCAAGTGCTGCGGATCTGTAGAGT 3200
```

```
gi|28193968|gb|AF339487.1|/1-4689
CAAGACTGTAGTGCCTTCCCCTTTCGTTAAGTGCTGCGGATCTGTAGAAT 3200
BAR_S2_3527/1-4689
CAAGACCGTAGTGCCTTCCCCTTTTGTCAAGTGCTGCGGATCTGTAGAGT 3200
BSA_S4_2265/1-4689
CAAGACCGTAGTGCCTTCCCCTTTTGTCAAGTGCTGCGGATCTGTAGAGT 3200
                                                  ******
**************** ** ************** ***** *

GSA_S1_936/1-4689
GCACGGGACGGAGTATGCCCGACTACCAGTGTAGGGTTGATTCTGGAGTG 3250
AMH_001346/1-4687
GCACGGGACGGAGTATGCCCGACTACCAGTGTAGGGTTGATTCTGGAGTG 3250
BAR_S2_3526/1-4689
GCACGGGACGGAGTATGCCCGACTACCAGTGTAGGGTTGATTCTGGAGTG 3250
BSA_S4_2268/1-4689
GCACGGGACGGAGTATGCCCGACTACCAGTGTAGGGTTGATTCTGGAGTG 3250
GSA_S5_4278/1-4689
GCACGGGACGGAGTATGCCCGACTACCAGTGTAGGGTTGATTCTGGAGTG 3250
gi|28193968|gb|AF339487.1|/1-4689
GCACGGGACGGAGTATGCCCGATTACCAGTGTAGGGTTGATTCTGGAGTG 3250
BAR_S2_3527/1-4689
GCACGGGACGGAGTATGCCCGACTACCAGTGTAGGGTTGATTCTGGAGTG 3250
BSA_S4_2265/1-4689
GCACGGGACGGAGTATGCCCGACTACCAGTGTAGGGTTGATTCTGGAGTG 3250
                                                  **********************
**************************

GSA_S1_936/1-4689
TACCCGTACATGTGGGGAGGGGCATACTGCTTTTGCTCTAGTGAGAACAC 3300
AMH_001346/1-4687
TACCCGTACATGTGGGGAGGGGCATACTGCTTTTGCTCTAGTGAGAACAC 3300
BAR_S2_3526/1-4689
TACCCGTACATGTGGGGAGGGGCATACTGCTTTTGCTCTAGTGAGAACAC 3300
BSA_S4_2268/1-4689
TACCCGTACATGTGGGGAGGGGCATACTGCTTTTGCTCTAGTGAGAACAC 3300
GSA_S5_4278/1-4689
TACCCGTACATGTGGGGAGGGGCATACTGCTTTTGCTCTAGTGAGAACAC 3300
gi|28193968|gb|AF339487.1|/1-4689
TACCCGTACATGTGGGGAGGGGCATACTGCTTTTGCTCTAGTGAGAACAC 3300
BAR_S2_3527/1-4689
TACCCGTACATGTGGGGAGGGGCATACTGCTTTTGCTCTAGTGAGAACAC 3300
BSA_S4_2265/1-4689
TACCCGTACATGTGGGGAGGGGCATACTGCTTTTGCTCTAGTGAGAACAC 3300

**********************************************

GSA_S1_936/1-4689
TCAGATGAGTGAAGCCTACGTAGAACGTGCAGAAGTGTGTAAGCACGAAC 3350
AMH_001346/1-4687
TCAGATGAGTGAAGCCTACGTAGAACGTGCAGAAGTGTGTAAGCACGAAC 3350
BAR_S2_3526/1-4689
TCAGATGAGTGAAGCCTACGTAGAACGTGCAGAAGTGTGTAAGCACGAAC 3350
```

```
BSA_S4_2268/1-4689
TCAGATGAGTGAAGCCTACGTAGAACGTGCAGAAGTGTGTAAGCACGAAC 3350
GSA_S5_4278/1-4689
TCAGATGAGTGAAGCCTACGTAGAACGTGCAGAAGTGTGTAAGCACGAAC 3350
gi|28193968|gb|AF339487.1|/1-4689
TCAGATGAGTGAAGCCTACGTAGAACGTGCAGAAGTGTGTAAGCACGAAC 3350
BAR_S2_3527/1-4689
TCAGATGAGTGAAGCCTACGTAGAACGTGCAGAAGTGTGTAAGCACGAAC 3350
BSA_S4_2265/1-4689
TCAGATGAGTGAAGCCTACGTAGAACGTGCAGAAGTGTGTAAGCACGAAC 3350

**************************************************

GSA_S1_936/1-4689
ACGCACTGGCTTACAAAACACAGACAGCGTCACTTACAGCCACCCTAAAA 3400
AMH_001346/1-4687
ACGCACTGGCTTACAAAACACAGACAGCGTCACTTACAGCCACCCTAAAA 3400
BAR_S2_3526/1-4689
ACGCACTGGCTTACAAAACACAGACAGCGTCACTTACAGCCACCCTAAAA 3400
BSA_S4_2268/1-4689
ACGCACTGGCTTACAAAACACAGACAGCGTCACTTACAGCCACCCTAAAA 3400
GSA_S5_4278/1-4689
ACGCACTGGCTTACAAAACACAGACAGCGTCACTTACAGCCACCCTAAAA 3400
gi|28193968|gb|AF339487.1|/1-4689
ACGCATTGGCTTACAAAACACAGACAGCGTCACTTACAGCCACCCTAAAA 3400
BAR_S2_3527/1-4689
ACGCACTGGCTTACAAAACACAGACAGCGTCACTTACAGCCACCCTAAAA 3400
BSA_S4_2265/1-4689
ACGCACTGGCTTACAAAACACAGACAGCGTCACTTACAGCCACCCTAAAA 3400
                                                *****
*****************************************

GSA_S1_936/1-4689
GTGACGTTGGGAAACATCACTCAAGGAGCTAAGGCTTACGTAAATGGGGA 3450
AMH_001346/1-4687
GTGACGTTGGGAAACATCACTCAAGGAGCTAAGGCTTATGTAAATGGGGA 3450
BAR_S2_3526/1-4689
GTGACGTTGGGAAACATCACTCAAGGAGCTAAGGCTTACGTAAATGGGGA 3450
BSA_S4_2268/1-4689
GTGACGTTGGGAAACATCACTCAAGGAGCTAAGGCTTACGTAAATGGGGA 3450
GSA_S5_4278/1-4689
GTGACGTTGGGAAACATCACTCAAGGAGCTAAGGCTTACGTAAATGGGGA 3450
gi|28193968|gb|AF339487.1|/1-4689
GTGACGTTGGGAAACATCACTCAAGGAGCTAAGGCTTACGTAAATGGGGA 3450
BAR_S2_3527/1-4689
GTGACGTTGGGAAACATCACTCAAGGAGCTAAGGCTTACGTAAATGGGGA 3450
BSA_S4_2265/1-4689
GTGACGTTGGGAAACATCACTCAAGGAGCTAAGGCTTACGTAAATGGGGA 3450

************************************ ***********

GSA_S1_936/1-4689
GACACCGTTCCAAATTGGCCAGGCTAAATTTGTTTTAGGACCTATATCCG 3500
```

```
AMH_001346/1-4687
GACACCGTTCCAAATCGGCCAGGCTAAATTTGTTTTAGGACCTATATCCG 3500
BAR_S2_3526/1-4689
GACACCGTTCCAAATTGGCCAGGCTAAATTTGTTTTAGGACCTATATCCG 3500
BSA_S4_2268/1-4689
GACACCGTTCCAAATTGGCCAGGCTAAATTTGTTTTAGGACCTATATCCG 3500
GSA_S5_4278/1-4689
GACACCGTTCCAAATTGGCCAGGCTAAATTTGTTTTAGGACCTATATCCG 3500
gi|28193968|gb|AF339487.1|/1-4689
GACACCGTTCCAAATTGGCCAGGCTAAATTTGTTTTAGGACCGATATCCG 3500
BAR_S2_3527/1-4689
GACACCGTTCCAAATTGGCCAGGCTAAATTTGTTTTAGGACCTATATCCG 3500
BSA_S4_2265/1-4689
GACACCGTTCCAAATTGGCCAGGCTAAATTTGTTTTAGGACCTATATCCG 3500
                                                    ***************
************************** *******

GSA_S1_936/1-4689
CGGCGTGGTCGCCATTTAACCCTAAGATAGTAGTATATAAGGATGATGTA 3550
AMH_001346/1-4687
CGGCGTGGTCGCCATTTAACCCTAAGATAGTAGTATATAAGGATGATGTA 3550
BAR_S2_3526/1-4689
CGGCGTGGTCGCCATTTAACCCTAAGATAGTAGTATATAAGGATGATGTA 3550
BSA_S4_2268/1-4689
CGGCGTGGTCGCCATTTAACCCTAAGATAGTAGTATATAAGGATGATGTA 3550
GSA_S5_4278/1-4689
CGGCGTGGTCGCCATTTAACCCTAAGATAGTAGTATATAAGGATGATGTA 3550
gi|28193968|gb|AF339487.1|/1-4689
CGGCGTGGTCGCCATTTAACCCTAAGATAGTAGTATATAAGGATGATGTA 3550
BAR_S2_3527/1-4689
CGGCGTGGTCGCCATTTAACCCTAAGATAGTAGTATATAAGGATGATGTA 3550
BSA_S4_2265/1-4689
CGGCGTGGTCGCCATTTAACCCTAAGATAGTAGTATATAAGGATGATGTA 3550

**************************************************

GSA_S1_936/1-4689
TATAATTATGACTTTCCGGCCTACGGTGCCGGGCAGCCAGGCCGTTTTGG 3600
AMH_001346/1-4687
TATAATTATGACTTTCCGGCCTACGGTGCCGGGCAACCAGGCCGTTTTGG 3600
BAR_S2_3526/1-4689
TATAATTATGACTTTCCGGCCTACGGTGCCGGGCAGCCAGGCCGTTTTGG 3600
BSA_S4_2268/1-4689
TATAATTATGACTTTCCGGCCTACGGTGCCGGGCAGCCAGGCCGTTTTGG 3600
GSA_S5_4278/1-4689
TATAATTATGACTTTCCGGCCTACGGTGCCGGGCAGCCAGGCCGTTTTGG 3600
gi|28193968|gb|AF339487.1|/1-4689
TATAATTATGATTTTCCGGCCTACGGTGCCGGGCAGCCAGGCCGTTTTGG 3600
BAR_S2_3527/1-4689
TATAATTATGACTTTCCGGCCTACGGTGCCGGGCAGCCAGGCCGTTTTGG 3600
BSA_S4_2265/1-4689
TATAATTATGACTTTCCGGCCTACGGTGCCGGGCAGCCAGGCCGTTTTGG 3600
```

```
                                                          ***********
********************** **************

GSA_S1_936/1-4689
GGATATCCAGAGTAGGACAGTAGATAGTAAGGACCTGTATGCAAGGACAC 3650
AMH_001346/1-4687
GGATATCCAGAGTAGGACAGTAGATAGTAAGGACCTGTATGCAAGGACAC 3650
BAR_S2_3526/1-4689
GGATATCCAGAGTAGGACAGTAGATAGTAAGGACCTGTATGCAAGGACAC 3650
BSA_S4_2268/1-4689
GGATATCCAGAGTAGGACAGTAGATAGTAAGGACCTGTATGCAAGGACAC 3650
GSA_S5_4278/1-4689
GGATATCCAGAGTAGGACAGTAGATAGTAAGGACCTGTATGCAAGGACAC 3650
gi|28193968|gb|AF339487.1|/1-4689
GGATATCCAGAGTAGGACAGTAGATAGTAAGGACCTGTATGCAAGGACAC 3650
BAR_S2_3527/1-4689
GGATATCCAGAGTAGGACAGTAGATAGTAAGGACCTGTATGCAAGGACAC 3650
BSA_S4_2265/1-4689
GGATATCCAGAGTAGGACAGTAGATAGTAAGGACCTGTATGCAAGGACAC 3650

**************************************************

GSA_S1_936/1-4689
ACCTAAGGCTAGACAGACCGGCATCTGGTAACATACATGTACCATATACC 3700
AMH_001346/1-4687
ACCTAAGGCTAGACAGACCGGCATCTGGTAATATACATGTACCATATACC 3700
BAR_S2_3526/1-4689
ACCTAAGGCTAGACAGACCGGCATCTGGTAACATACATGTACCATATACC 3700
BSA_S4_2268/1-4689
ACCTAAGGCTAGACAGACCGGCATCTGGTAACATACATGTACCATATACC 3700
GSA_S5_4278/1-4689
ACCTAAGGCTAGACAGACCGGCATCTGGTAACATACATGTACCATATACC 3700
gi|28193968|gb|AF339487.1|/1-4689
ACCTAAGGCTAGACAGACCGGCATCTGGTAACATACATGTACCATATACC 3700
BAR_S2_3527/1-4689
ACCTAAGGCTAGACAGACCGGCATCTGGTAACATACATGTACCATATACC 3700
BSA_S4_2265/1-4689
ACCTAAGGCTAGACAGACCGGCATCTGGTAACATACATGTACCATATACC 3700

****************************** ******************

GSA_S1_936/1-4689
CAGATACCATCGGGGTTTAAATATTGGATGCAAGAAAAAGGGGAGCCTCT 3750
AMH_001346/1-4687
CAGATACCATCGGGGTTTAAATATTGGATGCAAGAAAAAGGGGAGCCTCT 3750
BAR_S2_3526/1-4689
CAGATACCATCGGGGTTTAAATATTGGATGCAAGAAAAAGGGGAGCCTCT 3750
BSA_S4_2268/1-4689
CAGATACCATCGGGGTTTAAATATTGGATGCAAGAAAAAGGGGAGCCTCT 3750
GSA_S5_4278/1-4689
CAGATACCATCGGGGTTTAAATATTGGATGCAAGAAAAAGGGGAGCCTCT 3750
gi|28193968|gb|AF339487.1|/1-4689
CAGATACCATCGGGGTTTAAATATTGGATGCAAGAAAAAGGGGAGCCTCT 3750
```

```
BAR_S2_3527/1-4689
CAGATACCATCGGGGTTTAAATATTGGATGCAAGAAAAAGGGGAGCCTCT 3750
BSA_S4_2265/1-4689
CAGATACCATCGGGGTTTAAATATTGGATGCAAGAAAAAGGGGAGCCTCT 3750

**************************************************

GSA_S1_936/1-4689
GAACAGGCAAGCAGCATTTGGCTGTGTGGTAAAGACGAACCCAGTGAGAG 3800
AMH_001346/1-4687
GAACAGGCAAGCAGCATTCGGCTGTGTGGTAAAGACGAACCCAGTGAGAG 3800
BAR_S2_3526/1-4689
GAACAGGCAAGCAGCATTTGGCTGTGTGGTAAAGACGAACCCAGTGAGAG 3800
BSA_S4_2268/1-4689
GAACAGGCAAGCAGCATTTGGCTGTGTGGTAAAGACGAACCCAGTGAGAG 3800
GSA_S5_4278/1-4689
GAACAGGCAAGCAGCATTTGGCTGTGTGGTAAAGACGAACCCAGTGAGAG 3800
gi|28193968|gb|AF339487.1|/1-4689
GAACAGGCAAGCAGCATTTGGCTGTGTGGTAAAGACGAACCCAGTGAGAG 3800
BAR_S2_3527/1-4689
GAACAGGCAAGCAGCATTTGGCTGTGTGGTAAAGACGAACCCAGTGAGAG 3800
BSA_S4_2265/1-4689
GAACAGGCAAGCAGCATTTGGCTGTGTGGTAAAGACGAACCCAGTGAGAG 3800
                                                  ******************
******************************

GSA_S1_936/1-4689
CAGATAACTGCGCTTATGGCAACATTCCCATTAGCGTGGACATCCCGGAT 3850
AMH_001346/1-4687
CAGATAACTGCGCTTATGGCAACATTCCCATTAGCGTGGACATCCCGGAT 3850
BAR_S2_3526/1-4689
CAGATAACTGCGCTTATGGCAACATTCCCATTAGCGTGGACATCCCGGAT 3850
BSA_S4_2268/1-4689
CAGATAACTGCGCTTATGGCAACATTCCCATTAGCGTGGACATCCCGGAT 3850
GSA_S5_4278/1-4689
CAGATAACTGCGCTTATGGCAACATTCCCATTAGCGTGGACATCCCGGAT 3850
gi|28193968|gb|AF339487.1|/1-4689
CAGATAACTGCGCTTATGGCAACATTCCCATTAGCGTGGACATCCCGGAT 3850
BAR_S2_3527/1-4689
CAGATAACTGCGCTTATGGCAACATTCCCATTAGCGTGGACATCCCGGAT 3850
BSA_S4_2265/1-4689
CAGATAACTGCGCTTATGGCAACATTCCCATTAGCGTGGACATCCCGGAT 3850

**************************************************

GSA_S1_936/1-4689
GCGATGTTCACCAGAGTAGTTGATATACCAGCGGTATCTCATCTGCAATG 3900
AMH_001346/1-4687
GCGATGTTCACCAGAGTAGTTGATATACCAGCGGTATCTCATCTGCAATG 3900
BAR_S2_3526/1-4689
GCGATGTTCACCAGAGTAGTTGATATACCAGCGGTATCTCATCTGCAATG 3900
BSA_S4_2268/1-4689
GCGATGTTCACCAGAGTAGTTGATATACCAGCGGTATCTCATCTGCAATG 3900
```

```
GSA_S5_4278/1-4689
GCGATGTTCACCAGAGTAGTTGATATACCAGCGGTATCTCATCTGCAATG 3900
gi|28193968|gb|AF339487.1|/1-4689
GCGATGTTCACCAGAGTAGTTGATATACCAGCGGTATCCCATCTGCAATG 3900
BAR_S2_3527/1-4689
GCGATGTTCACCAGAGTAGTTGATATACCAGCGGTATCTCATCTGCAATG 3900
BSA_S4_2265/1-4689
GCGATGTTCACCAGAGTAGTTGATATACCAGCGGTATCTCATCTGCAATG 3900

************************************ ***********

GSA_S1_936/1-4689
CATAGATGTCAGCTGCGCCCACTCGGCAGGATTCGGCGGAGTGGCGAAGC 3950
AMH_001346/1-4687
TATAGATGTCAGCTGCGCTCACTCGGCAGGATTCGGCGGAGTGGCGAAGC 3950
BAR_S2_3526/1-4689
CATAGATGTCAGCTGCGCCCACTCGGCAGGATTCGGCGGAGTGGCGAAGC 3950
BSA_S4_2268/1-4689
CATAGATGTCAGCTGCGCCCACTCGGCAGGATTCGGCGGAGTGGCGAAGC 3950
GSA_S5_4278/1-4689
CATAGATGTCAGCTGCGCCCACTCGGCAGGATTCGGCGGAGTGGCGAAGC 3950
gi|28193968|gb|AF339487.1|/1-4689
CATAGATGTCAGCTGCGCCCACTCGGCAGGATTTGGCGGAGTGGCGAAGC 3950
BAR_S2_3527/1-4689
CATAGATGTCAGCTGCGCCCACTCGGCAGGATTCGGCGGAGTGGCGAAGC 3950
BSA_S4_2265/1-4689
CATAGATGTCAGCTGCGCCCACTCGGCAGGATTCGGCGGAGTGGCGAAGC 3950
                                                 ******************
************** ****************

GSA_S1_936/1-4689
TGAGGTTCATGGCAGACCGCACAGGGAAGTGTGCCATCCACACGGCATCG 4000
AMH_001346/1-4687
TGAGGTTCATGGCAGACCGCACAGGGAAGTGTGCCATCCACACGGCATCG 4000
BAR_S2_3526/1-4689
TGAGGTTCATGGCAGACCGCACAGGGAAGTGTGCCATCCACACGGCATCG 4000
BSA_S4_2268/1-4689
TGAGGTTCATGGCAGACCGCACAGGGAAGTGTGCCATCCACACGGCATCG 4000
GSA_S5_4278/1-4689
TGAGGTTCATGGCAGACCGCACAGGGAAGTGTGCCATCCACACGGCATCG 4000
gi|28193968|gb|AF339487.1|/1-4689
TGAGGTTTATGGCAGACCGCACAGGGAAGTGTGCCATCCACACGGCATCG 4000
BAR_S2_3527/1-4689
TGAGGTTCATGGCAGACCGCACAGGGAAGTGTGCCATCCACACGGCATCG 4000
BSA_S4_2265/1-4689
TGAGGTTCATGGCAGACCGCACAGGGAAGTGTGCCATCCACACGGCATCG 4000
                                                 *******
****************************************

GSA_S1_936/1-4689
AAAATAGTTGCTATTCAGGAGGCAACGGTGGACTTACCCGAGATAGCAGT 4050
AMH_001346/1-4687
AAAATAGTTGCTATTCAGGAGGCAACGGTGGACTTACCCGAGATAGCAGT 4050
```

```
BAR_S2_3526/1-4689
AAAATAGTTGCTATTCAGGAGGCAACGGTGGACTTACCCGAGATAGCAGT 4050
BSA_S4_2268/1-4689
AAAATAGTTGCTATTCAGGAGGCAACGGTGGACTTACCCGAGATAGCAGT 4050
GSA_S5_4278/1-4689
AAAATAGTTGCTATTCAGGAGGCAACGGTGGACTTACCCGAGATAGCAGT 4050
gi|28193968|gb|AF339487.1|/1-4689
AAAATAGTTGCTATTCAGGAGGCAACGGTGGACTTACCCGAGATAGCAGT 4050
BAR_S2_3527/1-4689
AAAATAGTTGCTATTCAGGAGGCAACGGTGGACTTACCCGAGATAGCAGT 4050
BSA_S4_2265/1-4689
AAAATAGTTGCTATTCAGGAGGCAACGGTGGACTTACCCGAGATAGCAGT 4050

**************************************************

GSA_S1_936/1-4689
GGGCGCAGGAGGTGGAACAGCAGAAATAACGTTCCACTACTCAACAGCCT 4100
AMH_001346/1-4687
GGGCGCAGGAGGTGGAACAGCAGAAATAACGTTCCACTACTCAACAGCCT 4100
BAR_S2_3526/1-4689
GGGCGCAGGAGGTGGAACAGCAGAAATAACGTTCCACTACTCAACAGCCT 4100
BSA_S4_2268/1-4689
GGGCGCAGGAGGTGGAACAGCAGAAATAACGTTCCACTACTCAACAGCCT 4100
GSA_S5_4278/1-4689
GGGCGCAGGAGGTGGAACAGCAGAAATAACGTTCCACTACTCAACAGCCT 4100
gi|28193968|gb|AF339487.1|/1-4689
GGGCGCAGGAGGTGGAACAGCAGAAATAACGTTCCACTACTCAACAGCCT 4100
BAR_S2_3527/1-4689
GGGCGCAGGAGGTGGAACAGCAGAAATAACGTTCCACTACTCAACAGCCT 4100
BSA_S4_2265/1-4689
GGGCGCAGGAGGTGGAACAGCAGAAATAACGTTCCACTACTCAACAGCCT 4100

**************************************************

GSA_S1_936/1-4689
TGGCATCGCCGTCCTTTGACGTCCAACTGTGTTCTGCACACGTGACGTGC 4150
AMH_001346/1-4687
TGGCATCGCCGTCCTTTGACGTCCAACTGTGTTCTGCACACGTGACGTGC 4150
BAR_S2_3526/1-4689
TGGCATCGCCGTCCTTTGACGTCCAACTGTGTTCTGCACACGTGACGTGC 4150
BSA_S4_2268/1-4689
TGGCATCGCCGTCCTTTGACGTCCAACTGTGTTCTGCACACGTGACGTGC 4150
GSA_S5_4278/1-4689
TGGCATCGCCGTCCTTTGACGTCCAACTGTGTTCTGCACACGTGACGTGC 4150
gi|28193968|gb|AF339487.1|/1-4689
TGGCATCGCCGTCCTTTGACGTCCAACTGTGTTCTGCACACGTGACGTGC 4150
BAR_S2_3527/1-4689
TGGCATCGCCGTCCTTTGACGTCCAACTGTGTTCTGCACACGTGACGTGC 4150
BSA_S4_2265/1-4689
TGGCATCGCCGTCCTTTGACGTCCAACTGTGTTCTGCACACGTGACGTGC 4150

**************************************************
```

```
GSA_S1_936/1-4689
GATACCAAGTGCGAACCGCCGAAGGATCACATTGTCCCTTTCGCGGCTCA 4200
AMH_001346/1-4687
GATACCAAGTGCGAACCGCCGAAGGATCACATTGTCCCTTTCGCGGCTCA 4200
BAR_S2_3526/1-4689
GATACCAAGTGCGAACCGCCGAAGGATCACATTGTCCCTTTCGCGGCTCA 4200
BSA_S4_2268/1-4689
GATACCAAGTGCGAACCGCCGAAGGATCACATTGTCCCTTTCGCGGCTCA 4200
GSA_S5_4278/1-4689
GATACCAAGTGCGAACCGCCGAAGGATCACATTGTCCCTTTCGCGGCTCA 4200
gi|28193968|gb|AF339487.1|/1-4689
GATACCAAGTGCGAACCGCCGAAGGATCACATTGTCCCTTTCGCGGCTCA 4200
BAR_S2_3527/1-4689
GATACCAAGTGCGAACCGCCGAAGGATCACATTGTCCCTTTCGCGGCTCA 4200
BSA_S4_2265/1-4689
GATACCAAGTGCGAACCGCCGAAGGATCACATTGTCCCTTTCGCGGCTCA 4200

**************************************************

GSA_S1_936/1-4689
GCACATGTCGAACGACATGCCTAACTTGTCCGCCACTGCGATGAATTGGG 4250
AMH_001346/1-4687
GCACATGTCGAACGACATGCCTAACTTGTCCGCCACTGCGATGAATTGGG 4250
BAR_S2_3526/1-4689
GCACATGTCGAACGACATGCCTAACTTGTCCGCCACTGCGATGAATTGGG 4250
BSA_S4_2268/1-4689
GCACATGTCGAACGACATGCCTAACTTGTCCGCCACTGCGATGAATTGGG 4250
GSA_S5_4278/1-4689
GCACATGTCGAACGACATGCCTAACTTGTCCGCCACTGCGATGAATTGGG 4250
gi|28193968|gb|AF339487.1|/1-4689
GCACATGTCGAACGACATGCCTAACCTGTCCGCCACTGCGATGAATTGGG 4250
BAR_S2_3527/1-4689
GCACATGTCGAACGACATGCCTAACTTGTCCGCCACTGCGATGAATTGGG 4250
BSA_S4_2265/1-4689
GCACATGTCGAACGACATGCCTAACTTGTCCGCCACTGCGATGAATTGGG 4250

************************ ************************

GSA_S1_936/1-4689
TGACCGGATTGGGAACAAGTATTGGGACGTTTGTCTTCTTGTTCCTGGGG 4300
AMH_001346/1-4687
TGACCGGATTGGGAACAAGTATTGGGACATTTGTCTTCTTGTTCCTGGGG 4300
BAR_S2_3526/1-4689
TGACCGGATTGGGAACAAGTATTGGGACGTTTGTCTTCTTGTTCCTGGGG 4300
BSA_S4_2268/1-4689
TGACCGGATTGGGAACAAGTATTGGGACGTTTGTCTTCTTGTTCCTGGGG 4300
GSA_S5_4278/1-4689
TGACCGGATTGGGAACAAGTATTGGGACGTTTGTCTTCTTGTTCCTGGGG 4300
gi|28193968|gb|AF339487.1|/1-4689
TGACCGGATTGGGAACGAGTATTGGGACATTTGTCTTCTTGTTCCTGGGG 4300
BAR_S2_3527/1-4689
TGACCGGATTGGGAACAAGTATTGGGACGTTTGTCTTCTTGTTCCTGGGG 4300
```

```
BSA_S4_2265/1-4689
TGACCGGATTGGGAACAAGTATTGGGACGTTTGTCTTCTTGTTCCTGGGG 4300
                                                    ****************
********** ********************

GSA_S1_936/1-4689
ATACTACTGGTGGTCACCATAATTAGATGTTTTACCAGATAACTTTGGTG 4350
AMH_001346/1-4687
ATACTACTGGTGGTCACCATAATTAGATGTTTTACCAGATAACTTTGGTG 4350
BAR_S2_3526/1-4689
ATACTACTGGTGGTCACCATAATTAGATGTTTTACCAGATAACTTTGGTG 4350
BSA_S4_2268/1-4689
ATACTACTGGTGGTCACCATAATTAGATGTTTTACCAGATAACTTTGGTG 4350
GSA_S5_4278/1-4689
ATACTACTGGTGGTCACCATAATTAGATGTTTTACCAGATAACTTTGGTG 4350
gi|28193968|gb|AF339487.1|/1-4689
ATACTACTGGTGGTCACCATAATTAGATGTTTTACCAGATAACTTTGGTG 4350
BAR_S2_3527/1-4689
ATACTACTGGTGGTCACCATAATTAGATGTTTTACCAGATAACTTTGGTG 4350
BSA_S4_2265/1-4689
ATACTACTGGTGGTCACCATAATTAGATGTTTTACCAGATAACTTTGGTG 4350

**************************************************

GSA_S1_936/1-4689
TGGTCCCAGCATGCAGAGGTATTATAGATAACTTAGTTAGGTATTATAGG 4400
AMH_001346/1-4687
TGGTCCCAGCATGCAGAGGTATTATAGATAACTTAGTTAGGTATTATAGG 4400
BAR_S2_3526/1-4689
TGGTCCCAGCATGCAGAGGTATTATAGATAACTTAGTTAGGTATTATAGG 4400
BSA_S4_2268/1-4689
TGGTCCCAGCATGCAGAGGTATTATAGATAACTTAGTTAGGTATTATAGG 4400
GSA_S5_4278/1-4689
TGGTCCCAGCATGCAGAGGTATTATAGATAACTTAGTTAGGTATTATAGG 4400
gi|28193968|gb|AF339487.1|/1-4689
TGGTCCCAGCATGCTGAGGTATTATAGATAACTTAGTTAGGTATTATAGG 4400
BAR_S2_3527/1-4689
TGGTCCCAGCATGCAGAGGTATTATAGATAACTTAGTTAGGTATTATAGG 4400
BSA_S4_2265/1-4689
TGGTCCCAGCATGCAGAGGTATTATAGATAACTTAGTTAGGTATTATAGG 4400
                                                    **************
**********************************

GSA_S1_936/1-4689
TAACTTAGGTGTAAGCAGAAAAATGGAAAACCGAATAAAAAGTTAGAGTA 4450
AMH_001346/1-4687
TAACTTAGGTGTAAGCAGAAAAATGGAAAACCGAATAAAAAGTTAGAGTA 4450
BAR_S2_3526/1-4689
TAACTTAGGTGTAAGCAGAAAAATGGAAAACCGAATAAAAAGTTAGAGTA 4450
BSA_S4_2268/1-4689
TAACTTAGGTGTAAGCAGAAAAATGGAAAACCGAATAAAAAGTTAGAGTA 4450
GSA_S5_4278/1-4689
TAACTTAGGTGTAAGCAGAAAAATGGAAAACCGAATAAAAAGTTAGAGTA 4450
```

```
gi|28193968|gb|AF339487.1|/1-4689
TAACTTAGGTGTAAGCAGAAAAATGGAAAACCGAATAAAAAGTTAGAGTA 4450
BAR_S2_3527/1-4689
TAACTTAGGTGTAAGCAGAAAAATGGAAAACCGAATAAAAAGTTAGAGTA 4450
BSA_S4_2265/1-4689
TAACTTAGGTGTAAGCAGAAAAATGGAAAACCGAATAAAAAGTTAGAGTA 4450

**************************************************

GSA_S1_936/1-4689
AGTTATGTAAATTAGAAAATAAGTTTTGATAGTGGTAGTTAGGTGTAAGC 4500
AMH_001346/1-4687
AGTAATGTAAATTAGAAAATAAGTTTTGATAGTGGTAGTTAGGTGTAAGC 4500
BAR_S2_3526/1-4689
AGTTATGTAAATTAGAAAATAAGTTTTGATAGTGGTAGTTAGGTGTAAGC 4500
BSA_S4_2268/1-4689
AGTTATGTAAATTAGAAAATAAGTTTTGATAGTGGTAGTTAGGTGTAAGC 4500
GSA_S5_4278/1-4689
AGTTATGTAAATTAGAAAATAAGTTTTGATAGTGGTAGTTAGGTGTAAGC 4500
gi|28193968|gb|AF339487.1|/1-4689
AGTAGGTGTAAATTAGAAAATAAGTTTTGATAGTGGTAGTTAGGTGTAAGC 4500
BAR_S2_3527/1-4689
AGTTATGTAAATTAGAAAATAAGTTTTGATAGTGGTAGTTAGGTGTAAGC 4500
BSA_S4_2265/1-4689
AGTTATGTAAATTAGAAAATAAGTTTTGATAGTGGTAGTTAGGTGTAAGC 4500
                                                  ***
**********************************************

GSA_S1_936/1-4689
AGAAAACAGAAAAATGAATAAAAAGCTAGAGTAAGTAGTTAGCTGCATAT 4550
AMH_001346/1-4687
AGAAAACAGAAAAATGAATAAAAAGCTAGAGTAAGTAGTTAGCTGCATAT 4550
BAR_S2_3526/1-4689
AGAAAACAGAAAAATGAATAAAAAGCTAGAGTAAGTAGTTAGCTGCATAT 4550
BSA_S4_2268/1-4689
AGAAAACAGAAAAATGAATAAAAAGCTAGAGTAAGTAGTTAGCTGCATAT 4550
GSA_S5_4278/1-4689
AGAAAACAGAAAAATGAATAAAAAGCTAGAGTAAGTAGTTAGCTGCATAT 4550
gi|28193968|gb|AF339487.1|/1-4689
AGAAAACAGAAAAACGAATAAAAAGCTAGAGTAAGTAGTTAGCTGCATAT 4550
BAR_S2_3527/1-4689
AGAAAACAGAAAAATGAATAAAAAGCTAGAGTAAGTAGTTAGCTGCATAT 4550
BSA_S4_2265/1-4689
AGAAAACAGAAAAATGAATAAAAAGCTAGAGTAAGTAGTTAGCTGCATAT 4550
                                                  **************
**********************************

GSA_S1_936/1-4689
AGAGGTAGTATAGGTGTAAGCAGAAAATGGAAAAATAGTAAAAAGTTAGA 4600
AMH_001346/1-4687
AGAGGTAGTATAGGTGTAAGCAGAAAATGAAAAATTAGTAAAA--TTAGA 4598
BAR_S2_3526/1-4689
AGAGGTAGTATAGGTGTAAGCAGAAAATGGAAAAATAGTAAAAAGTTAGA 4600
```

```
BSA_S4_2268/1-4689
AGAGGTAGTATAGGTGTAAGCAGAAAATGGAAAAATAGTAAAAAGTTAGA 4600
GSA_S5_4278/1-4689
AGAGGTAGTATAGGTGTAAGCAGAAAATGGAAAAATAGTAAAAAGTTAGA 4600
gi|28193968|gb|AF339487.1|/1-4689
AGAGGTAGTATAGGTGTAAGCAGAAAATGGAAAACCAGTAAAAAGTTAGA 4600
BAR_S2_3527/1-4689
AGAGGTAGTATAGGTGTAAGCAGAAAATGGAAAAATAGTAAAAAGTTAGA 4600
BSA_S4_2265/1-4689
AGAGGTAGTATAGGTGTAAGCAGAAAATGGAAAAATAGTAAAAAGTTAGA 4600

**************************** ****  *******  *****

GSA_S1_936/1-4689
GTAAGTAGTAGAGGATAAGTTTTGTTATGCATTAGATAGTTTGCTTAGAT 4650
AMH_001346/1-4687
GTAAGTAGTAGAGAATAAGTTTTGCTATGCATTAGATAGTTTGCTTTGAT 4648
BAR_S2_3526/1-4689
GTAAGTAGTAGAGGATACGTTTTGTTATGCATTAGATAGTTTGCTTAGAT 4650
BSA_S4_2268/1-4689
GTAAGTAGTAGAGGATAAGTTTTGTTATGCATTAGATAGTTTGCTTAGAT 4650
GSA_S5_4278/1-4689
GTAAGTAGTAGAGGATAAGTTTTGTTATGCATTAGATAGTTTGCTTAGAT 4650
gi|28193968|gb|AF339487.1|/1-4689
GTAAGTAGTAGAGAATAAGTTTTGCTATGCATTAGATAGGTTGCTTTGAT 4650
BAR_S2_3527/1-4689
GTAAGTAGTAGAGGATAAGTTTTGTTATGCATTAGATAGTTTGCTTAGAT 4650
BSA_S4_2265/1-4689
GTAAGTAGTAGAGGATAAGTTTTGTTATGCATTAGATAGTTTGCTTAGAT 4650
                                                    ************* ***
****** ************** ****** ***

GSA_S1_936/1-4689
TCTATAGAAAATAGTAGATGCTTATAGAGGTAGTATAGG 4689
AMH_001346/1-4687
TCTATAGAAAATAGTAGATGCTTATAGAGGTAGTATAGG 4687
BAR_S2_3526/1-4689
TCTATAGAAAATAGTAGATGCTTATAGAGGTAGTATACG 4689
BSA_S4_2268/1-4689
TCTATAGAAAATAGTAGATGCTTATAGAGGTAGTATAGG 4689
GSA_S5_4278/1-4689
TCTATAGAAAATAGTAGATGCTTATAGAGGTAGTATAGG 4689
gi|28193968|gb|AF339487.1|/1-4689
TTTATAGAAAATAGTAGATGCTTATAGAGGTAGTATAGG 4689
BAR_S2_3527/1-4689
TCTATAGAAAATAGTAGATGCTTATAGAGGTAGTATAGG 4689
BSA_S4_2265/1-4689
TCTATAGAAAATAGTAGATGCTTATAGAGGTAGTATAGG 4689
                                          *
********************************** *
```

APPENDIX 2: Multiple protein sequence alignment for Ndumu viruses and other alphaviruses


-K.KKKPQKA.AKKNE.QKK-NENK..PPK..
Getahvirus_4
.NYIPTQTFYGR.WRP.PA.RPWRV.MQPAP.M.IPELQIPIV.AQ.M....S..SA.TTK.NGK.PK
-K.KKKPQKA.AKKNE.QKK-NENK..PPK..
Getahvirus_5
.NYIPTQTFYGR.WRP.PAYRPWRV.MQPAP.M.IPELQTPIV.AQ.M....S..SA.TTK.NGK.PK
-K.KKKPQKA.AKKNE.QKK-NENK..PPK..
Getahvirus_6
.NYIPTQTFYGR.WRP.PAYRPWRV.MQPAP.M.IPELQTPIV.AQ.M....S..SA.TTK.NGK.PK
-K.KKKPQKA.AKKNE.QKK-NENK..PPK..
Semlikiforestvirus_1 .NYIPTQTFYGR.WRP.PAARPW.LQATPVA...P-
-----DF.AQ.M....S..NA.TMR.NAI.PARP.KPKKKKTTKPK.KTQPK.I.GKTQQQKK.
Semlikiforestvirus_2 .NYIPTQTFYGR.WRP.PAARPW.LQATPVA...P-
-----DF.AQ.M....S..NA.TMR.NAI.PTRP.KPKKKKTTKPK.KTQPK.I.GKTQQQKK.
Semlikiforestvirus_3 .NYIPTQTFYGR.WRP.PAARPW.LQATPVA...P-
-----DF.AQ.M....S..NA.TMR.NAI.PAGP.KPKKKKTTKPK.KTQPK.I.GKTQQQKK.
Semlikiforestvirus_4 .NYIPTQTFYGR.WRP.PAARPW.LQATPVA...P-
-----DF.AQ.M....S..NA.TMR.NAI.PARP.KPKKKKTTKPK.KTQPK.INGKTQQQKK.
Semlikiforestvirus_5 .NYIPTQTFYGR.WRP.PAARPW.LQATPVA...P-
-----DF.AQ.M....S..NA.TMR.NAI.PARP.KPKKKKTTKPK.KTPPK.INGKTQQQKK.
Babankivirus .NRG-
FFNMFGR.PFPAPTAMWR..R.RQAA.MPAR-----
NGLAS.....TT..SA.VIG.AT.P.NP..RPPP.Q-.K.A.K.PPKPK.PKTQEKKK.
Sindbisvirus_1 .NRG-
FFNMLGR.PFPAPTAMWR..R.RQAA.MPAR-----
NGLAS.....TT..SA.VIG.AT.P.TP..RPPP.Q-.K.A.K.PPKPK.PKTQEKKK.
Sindbisvirus_2 .NRG-
FFNMLGR.PFPAPTAMWR..R.RQAA.MPAR-----
NGLAS.....TT..SA.VIG.AT.P.TP..RPPP.Q-.K.A.K.PPKPK.PKTQEKKK.
Sindbisvirus_3 .NRG-
FFNMLGR.PFPAPTAMWR..R.RQAA.MPAR-----
NGLAS.....TT..SA.VIG.AT.P..P..RPPP.Q-.K.A.K.PPKPK.PKTQEKKK.
Sindbisvirus_4 .NRG-
FFNMLGR.PFPAPTAMWR..R.RQAA.MPAR-----
NGLAS.....TT..SA.VIG.AT.P..P..RPPP.Q-.K.A.K.PPKPK.PKTQEKKK.
Sindbisvirus_5 .NRG-
FFNMLGR.PFPAPTAMWR..R.RQAA.MPAR-----
NGLAS.....TT..SA.VIG.AT.P.TP..RPPP.Q-.K.A.K.PPKRK.PKTQEKKK.
Auravirus .NSV-FYNPFGRGAYAQP.I.WR..R.--
AA.APRP-----
SGLTT.....TR..RA.V.DNAT.R.RPA.RTRP.KP.T.K.KPKK.NQ.PPQQQKKG.
Ockelbovirus .NRG-
FFNMLGR.PFPAPTAMWR..R.RQAA.MPAR-----
NGLAS.....TT..SA.VIG.AT.P.NP..RPPP.Q-.K.A.K.PPKPK.PKPQEKKK.
110 120 130
140 150 160 170 180 190 200
....|....|....|....|....|....|....|....|....|....|....|....|...
.|....|....|....|....|....|....
gi|28193970|gb|AAO33345.1| str NAQVQTKKQDKTKKKPGKRERKCMKIESDCIFEVK-
LDGEVTGYACLVGDKVMKPAHVKGVIQNDELAKLSFKKSSKYDLECAQIPVKMRSDASKYTHEK
AMH_001346 .....-
...V.....D.....L.....
BAR_S2_3526 .....-
.....D.....


BSA_S4_2268
BSA_S4_2265
GSA_S1_936
BAR_S2_3527
GSA_S5_4278
AMH_001056
O'nyong-nyongvirus
Chikungunyavirus_1
Chikungunyavirus_2
Chikungunyavirus_3
RossRivervirus_1
RossRivervirus_2
RossRivervirus_3
RossRivervirus_4
Getahvirus_1
getahvirus_2
Getahvirus_3
Getahvirus_4
Getahvirus_5
Getahvirus_6
Semlikiforestvirus_1
Semlikiforestvirus_2
Semlikiforestvirus_3
Semlikiforestvirus_4
Semlikiforestvirus_5
Babankivirus
Sindbisvirus_1


Sindbisvirus_2
Sindbisvirus_3
Sindbisvirus_4
Sindbisvirus_5
Auravirus
Ockelbovirus
210 220 230
240 250 260 270 280 290 300
gi|28193970|gb|AAO33345.1|
PEGHYNWHHGAVQYTNGRFTIPTGAGKPGDSGRPIFDNKGRVVAIVLGGANEGARTALSVVTWN-
KDMVTRVTPEGSVEWSAAAT----CVLGAAIFSCL
AMH_001346
BAR_S2_3526
BSA_S4_2268
BSA_S4_2265
GSA_S1_936
BAR_S2_3527
GSA_S5_4278
AMH_001056
O'nyong-nyongvirus
Chikungunyavirus_1

..I..KI....AE...L.IP--VM.L.ANTT.P.S
Chikungunyavirus_2
...Y..........SG...-
..I..KI....AE...L.IP--VM.L.ANTT.P.S
Chikungunyavirus_3
...Y..........SG...-
..I..KI....AE...L.IP--VM.L.ANTT.P.S
RossRivervirus_1
..............SG...T-
...........TE.....LM---M.I.ANTS.P.S
RossRivervirus_2
..............SG...T-
...........TE.....LM---M.I.ANTS.P.S
RossRivervirus_3
..............SG...T-
...........TE.....LM---M.I.ANTS.P.S
RossRivervirus_4
..............SG...T-
...........TE.....LM---M.I.ANTS.P.S
Getahvirus_1
..............SG...T-
......Y....TE.....LM---M...ANVT.P.S
getahvirus_2
..............SG...T-
......Y....TE.....LM---M...ANVT.P.S
Getahvirus_3
..............SG...T-
......Y....TE.....LM---M...ANVT.P.S
Getahvirus_4
..............SG...T-
......Y....TE.....LM---M...ANVT.P.S
Getahvirus_5
..............SG...T-
......Y....TE.....LM---M...ANVT.P.S
Getahvirus_6
..............SG...T-
......Y....TE.....LM---M...ANVT.P.S
Semlikiforestvirus_1
..............SG.....................................S..........-
...........E....PLIT-AM...AN.T.P.F
Semlikiforestvirus_2
..............SG.....................................S..........-
...........E....PLIT-AM...TN.T.P.F
Semlikiforestvirus_3
..............SG.....................................S..........-
...........E....PLIT-AM...AN.T.P.F
Semlikiforestvirus_4
..............SG.....................................S..........-
...........E....PLIT-AM...AN.T.P.F
Semlikiforestvirus_5
..............SG.....................................S..........-
...........E....PLIT-AM...AN.T.P.F
Babankivirus
...F..........SG......R.V.GR.......M..S...........D..T..........S.GK
TIKT....TE.....PLVTAM.L..NVS.P.N
Sindbisvirus_1

```
...F.........SG......R.V.GR.......M..S...........D..T.........S.GK
TIKT....TE.....PLVTAM.L..NVS.P.N
Sindbisvirus_2
...F.........SG......R.V.GR.......M..S...........D..T.........S.GK
TIKT....TE.....PLVTAM.L..NVS.P.N
Sindbisvirus_3
...F.........SG......R.V.GR.......M..S...........D..T.........S.GK
TIKT....TE.....PLVTAM.L..NVS.P.D
Sindbisvirus_4
...F.........SG......R.V.GR.......M..S...........D..T.........S.GK
TIKT....TE.....PLVTAM.L..NVS.P.D
Sindbisvirus_5
...F.........SG......R.V.GR.......M..S...........D..T.........S.GK
TIKT....TE.....PLVTAM.LP.NVS.P.N
Auravirus
..VF.........FSG........V.G........L..S.K...........VPG.........K.GA
AIKT.H.DT....R.--ITAM.I.QNVT.P.D
Ockelbovirus
...F.........SG......R.V.GR.......M..S...........D..T.........S.GK
TIKT....TE.....PLVTAM.L..NVS.P.N

                                          310          320          330
340        350        360        370        380        390        400

....|....|....|....|....|....|....|....|....|....|....|....|....|...
.|....|....|....|....|....|....|
gi|28193970|gb|AAO33345.1|           str                          SP---
PVCYDEEPEKTLTMLSMNVDNPDYYSLFEAALDCRSR-RHRR-
AAQHFNEYKVTQPYIAYCPDCGNGESCQSPVAIERIRNEATDGTLKVQFALQ
AMH_001346                                                        ..---
.......................................-....-
.....................................................
BAR_S2_3526                                                       ..---
.......................................-....-
.....................................................
BSA_S4_2268                                                       ..---
.......................................-....-
.....................................................
BSA_S4_2265                                                       ..---
.......................................-....-
.....................................................
GSA_S1_936                                                        ..---
.......................................-....-
.....................................................
BAR_S2_3527                                                       ..---
.......................................-....-
.....................................................
GSA_S5_4278                                                       ..---
.......................................-....-
.....................................................
AMH_001056
Q.PCA.C..ENNA.A..R..ED...R.G..D.LQ...T..NGT....SVS....V..A.R......A.
..A.H..H......A..S.....M..I..SA.
O'nyong-nyongvirus
Q.PCA.C..EKK..E..R..ED..MQ.G..Q.LDS..A.SQ.-
.QK.N.REN..V....R..L.H.....E.H..H..I.L....S.........I.VS..
```

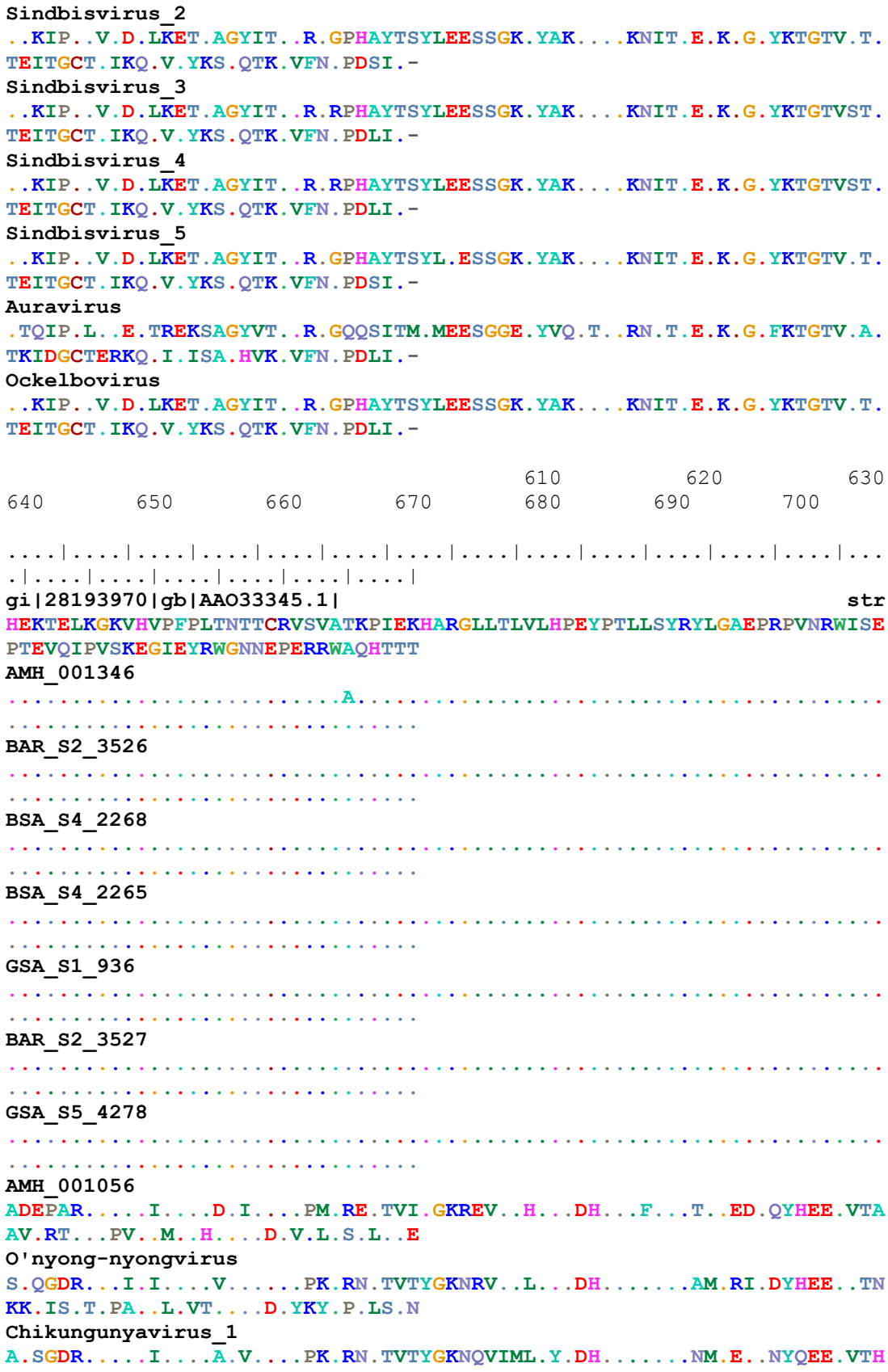
```
Chikungunyavirus_1
Q.PCT.C..EK.....R..ED..MS.G..Q.LQ.S.T.SP.-
.R..SIKD...V..A.R..L.H.....E.H..H....L.............I.VS..
Chikungunyavirus_2
Q.PCT.C..EK.....R..ED..MS.G..Q.LQ.S.T.SP.-
.Q..SIKD...V..A.R..L.H.....E.H..H....L.............I.VS..
Chikungunyavirus_3
Q.PCT.C..EK.....R..ED..MS.G..Q.LQ.S.T.SP.-
.R..SIKD...V..A.R..L.H.....E.H..H....L.............I.VS..
RossRivervirus_1
..PCY.C..EKQ..Q..R..ED..NR.G..E.L..SMT..N.S....SVTE...V..A.R..L...A.
..D.YF.Y......K..D..S..M..I.VSA.
RossRivervirus_2
..PCY.C..EKQ..Q..R..ED..NR.G..E.L..SMT..NGS....SVTE...V..A.R..L.H.A.
..D.YF.Y......K..D..S..M..I.VSA.
RossRivervirus_3
..PCY.C..EKQ..Q..R..ED..NR.G..E.L..SMT..N.S....SVTE...V..A.R..L.X.A.
..D.YF.Y......K..D..S..M..I.VSA.
RossRivervirus_4
..PCY.C..EKQ..Q..R..ED..NR.G..E.L..SMT..N.S....SVTE...V..A.R..L.H.A.
..D.YF.Y......K..D..S..M..I.VSA.
Getahvirus_1
E.ACA.C..EKQ..Q..R..ED...R.G..D.L..TMT.NNSA....SVTE...V..A.K..L...A.
..D.QF.Y......K..D..S..MI.I.V.A.
getahvirus_2
E.ACA.C..EKQ..Q..R..ED...R.G..D.L..TMT.NNSA....SVTE...V..A.K..L...A.
..D.Q..Y......K..D..S..MI.I.V.A.
Getahvirus_3
E.ACA.C..EKQ..Q..R..ED...R.G..D.L..TMT.NNSA....SVTE...V..A.K..L...A.
..D.QF.Y......K..D..S..MI.I.V.A.
Getahvirus_4
E.ACA.C..EKQ..Q..R..ED...R.G..D.L..TMT.NNSA....SVTE...V..A.K..L...A.
..D.QF.Y......K..D..S..MI.I.V.A.
Getahvirus_5
E.ACA.C..EKQ..Q..R..ED...R.G..D.L..TMT.NNSA....SVTK...V..A.K..L...A.
..D.QF.Y......K..D..S..MI.I.V.A.
Getahvirus_6
E.ACA.C..EKQ..Q..R..ED...R.G..D.L..TMT.NNSA....SVTE...V..A.K..L...A.
..D.QF.Y......K..D..S..MI.I.V.A.
Semlikiforestvirus_1
Q.PCV.C..ENNA.A..R..ED...R.G..D.LQ...T..NGT....SVS....V..A.R......A.
..A.H..H......A..S.....M..I..SA.
Semlikiforestvirus_2
Q.PCA.C..ENNA.A..R..ED...R.G..D.LQ...T..NGT....SVS....V..A.R......A.
..A.H..H......A..S.....M..I..SA.
Semlikiforestvirus_3
Q.PCV.C..ENNA.A..R..ED...R.G..D.LQ...T..NGT....SVS....V..A.R......A.
..A.H..H......A..S.....M..I..SA.
Semlikiforestvirus_4
Q.PCV.C..ENNA.A..R..ED...R.G..D.LQ...T..NGT....SVS....V..A.R......A.
..A.H..H......AV.S.....M..I..SA.
Semlikiforestvirus_5
Q.PCV.C..ENNA.A..R..ED...R.G..D.LQ...T..NGT....SVS....V..A.R......A.
..A.H..H......AV.S.....M..I..SA.
Babankivirus                                                  R.---
.T..TR..SRA.DI.EE..NHEA.DT.LN.I.R.G.SG.SK.-SVTDD--
```

```
FTL.S..LGT.SY.HHT.P.F..IK..QVWD..D.N.IRI.TSA.
Sindbisvirus_1                                                  R.---
.T..TR..SRA.DI.EE..NHEA.DT.LN.I.R.G.SG.SK.-SVTDD--
FTL.S..LGT.SY.HHT.P.F..IK..QVWD..D.N.IRI.TSA.
Sindbisvirus_2                                                  R.---
.T..TR..SRA.DI.EE..NHEA.DT.LN.I.R.G.SG.SK.-SVTDD--
FTL.S..LGT.SY.HHT.P.F..IK..QVWD..D.N.IRI.TSA.
Sindbisvirus_3                                                  R.---
.T..TR..SRA.DI.EE..NHEA.DT.LN.I.R.G.SG.SK.-SVIDG--
FTL.S..LGT.SY.HHT.P.F...K..QVWD..D.N.IRI.TSA.
Sindbisvirus_4                                                  R.---
.T..TR..SRA.DI.EE..NHEA.DT.LN.I.R.G.SG.SK.-SVIDD--
FTL.S..LGT.SY.HHTVP.F...K..QVWD..D.N.IRI.TSA.
Sindbisvirus_5                                                  R.---
.T..TR..SRA.DI.EE..NHEA.DT.LN.I.R.G.SG.SKKKSVTDD--
FTL.S..LGT.SY.HHT.P.F.LIK..QVWD..D.N.IRI.TSA.
Auravirus                                                       R.---
.T..NRN.DL.....ET..NH.S.DV.LD...R.PT.-..V.STPTDD--
FTL.A..LGL.HR.KTM.P.Y..IK..KVWDD.D..V.RI.VSA.
Ockelbovirus                                                    R.---
.T..TR..SRA.DI.EE..NHEA.DT.LN.I.R.G.SG.SK.-SVTDD--
FTL.S..LGT.SY.HHT.P.F..IK..QVWD..D.N.IRI.TSA.

                                          410         420         430
440        450        460        470        480        490        500

....|....|....|....|....|....|....|....|....|....|....|....|....|...
.|....|....|....|....|....|....|
gi|28193970|gb|AAO33345.1|   str     VGIKKDHSTAPTYLRYMEG---HEVKDALNN-
LRVYTSAQCEVTATMGHFALARCPPGDSLSVTFMDASK---KNRMCRVPFVHKLPFLGREKNSARRYH
AMH_001346                                  ..................---.........-
.....................................---.........................H..
BAR_S2_3526                                 ..................---.........-
.....................................---.........................H..
BSA_S4_2268                                 ..................---.........-
.....................................---.........................H..
BSA_S4_2265                                 ..................---.........-
.....................................---.........................H..
GSA_S1_936                                  ..................---.........-
.....................................---.........................H..
BAR_S2_3527                                 ..................---.........-
.....................................---.........................H..
GSA_S5_4278                                 ..................---.........-
.....................................---.........................H..
AMH_001056                                        I..D.SDNHDY.KI..AD.---
.AIEN.VRSS.K.A..GD.F.HG.....I..T....EF.Q.SIQ.TRN---
AVHA..IQYH.DPQPV....FTI.PHY
O'nyong-nyongvirus                                I...T.D.HDW.K....DS---
.TPV..DRSG.F.R...P.TI.G.....I.....K.ET.T.G.V.SRR---
ISHT.MH..R.EP.LI....FHS.PQ.
Chikungunyavirus_1                                I...T.D.HDW.K....DN---
.MPA..ERAG.F.R...P.TI.G.....I.....K.ET.T.G.T.GR.---
ISHS.TH..H.DP.VI....FHS.PQ.
Chikungunyavirus_2                                I...T.D.HDW.K....DN---
.MPA..ERAG.F.R...P.TI.G.....I.....K.ET.T.G.T.GR.---
ISHS.TH..H.DP.VI....FHS.PQ.
```

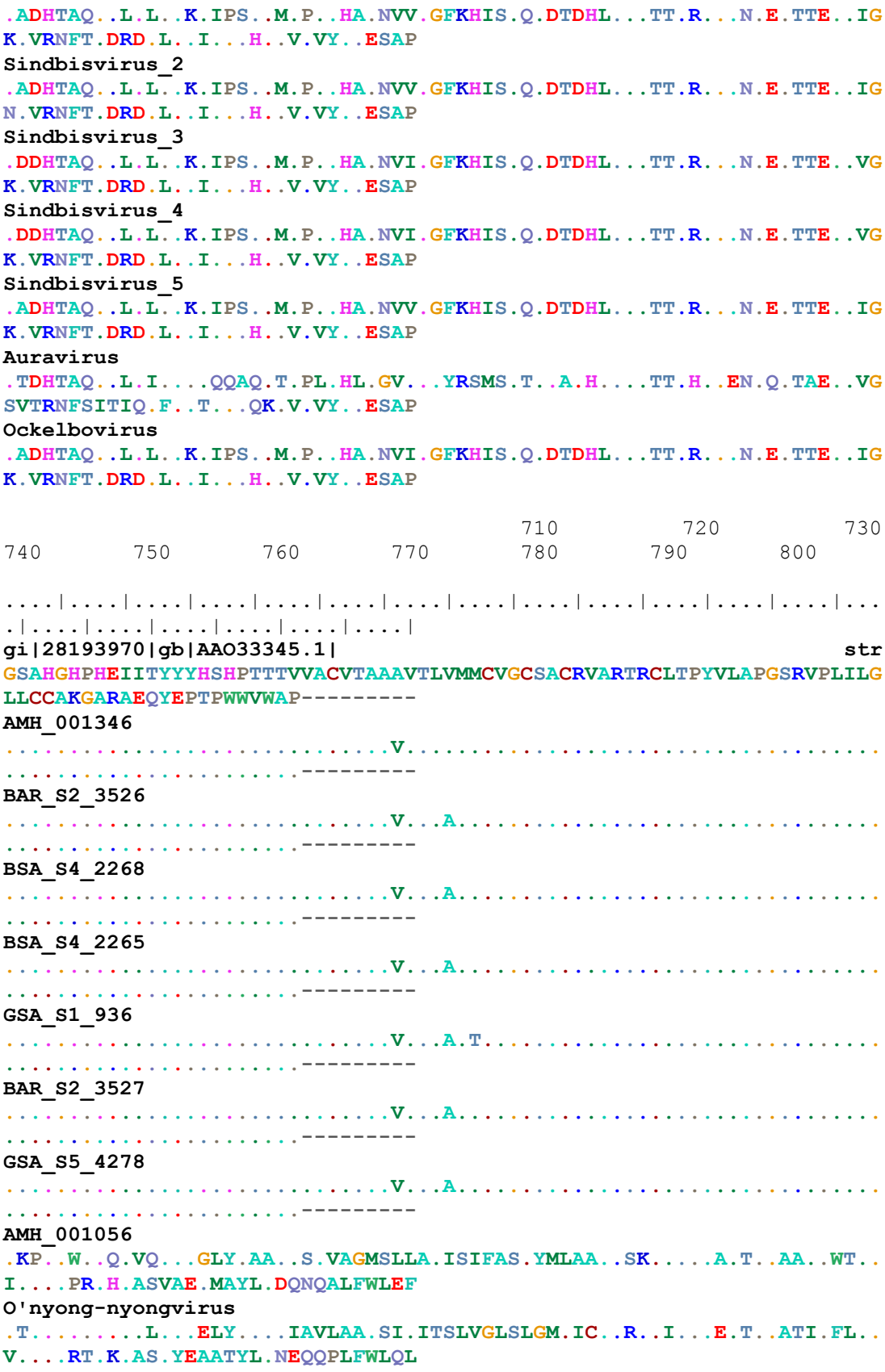
```
Chikungunyavirus_3                                I...T.D.HDW.K....DN---
.MPA..ERAG.F.R...P.TI.G.....I.....K.ET.T.G.T.GR.---
ISHS.TH..H.DP.VI....FHS.PQ.
RossRivervirus_1                                  I.LD.AGTH.H.KI...A.---
.D.QESKRDS.......A.SIHG.....IV.H.....Y.K.S.E..DS---
HVKA.K.QYK.DPLPV....FVV.PHF
RossRivervirus_2                                  I.LD.AGTH.H.K....A.---
.D.QESKRDS.......A.SIHG.....IV.H.....Y.K.S.E..DS---
HVKA.K.QYK.NPLPV....FVV.PHF
RossRivervirus_3                                  I.LD.AGTH.H.KI...A.---
.D.QESKRDS.......A.SIHG.....IV.H.....Y.K.S.E..DS---
HVKA.K.QYK.DPLPV....FVV.PHF
RossRivervirus_4                                  I.LD.AGTH.H.K....A.---
.D.QESKRDS.......A.SIHG.....IV.H.....Y.K.S.E..DS---
HVKA.K.QYK.DPLPV....FVV.PHF
Getahvirus_1                                      I..N.GGTHEHNKI..IA.---
.DM.E.NRDS.Q.H..GV.AIRG.....IV.Y.....E.K.Q.Q..ES---
HTQA.K.QYK.APAPV....FTV.PHF
getahvirus_2                                      I..N.GGTHEHNKI..IA.---
.DM.E.NRDS.Q.H..GT.AIRG.....IA.Y.....E.K.Q.Q..ES---
HIQA.K.QYK.APAPV....FTV.PHF
Getahvirus_3                                      I..N.GGTHEHNKI..IA.---
.DM.E.NRDS.Q.H..GV.AIRG.....IV.Y.....E.K.Q.Q..ES---
HTQA.K.QYK.APAPV....FTV.PHF
Getahvirus_4                                      I..N.GGTHEHNKI..IA.---
.DM.E.NRDS.Q.H..GV.AIRG.....IV.Y.....E.K.Q.Q..ES---
HTQA.K.QYK.APAPV....FTV.PHF
Getahvirus_5                                      I..N.GGTHEHNKI..IA.---
.DM.E.NRDS.Q.H..GV.AIRG.....IV.Y.....E.K.Q.Q..ES---
HTQA.K.QYK.APAPV....FTV.PHF
Getahvirus_6                                      I..N.GGTHEHNKI..IA.---
.DM.E.NRDS.Q.H..GV.AIRG.....IV.Y.....E.K.Q.Q..ES---
HTQA.K.QYK.APAPV....FTV.PHF
Semlikiforestvirus_1                              I..D.SDNHDY.KI..AD.---
.AIEN.VRSS.K.A..GD.F.HG.....I..T....EF.Q.SIQ.TRN---
AV.A..IQYH.DPQPV....FTI.PHY
Semlikiforestvirus_2                              I..D.SDNHDY.KI..AD.---
.AIEN.VRSS.K.A..GD.F.HG.....I..T....EF.Q.SIQ.TRN---
AV.A..IQYH.DPQPV....FTI.PHY
Semlikiforestvirus_3                              I..D.SDNHDY.KI..AD.---
.AIEN.VRSS.K.A..GD.F.HG.....I..T....EF.Q.SIQ.TRN---
AV.A..IQYH.DPQPV....FTI.PHY
Semlikiforestvirus_4                              I..D.SDNHDY.KI..AD.---
.AIEN.VRSS.K.A..GD.F.HG.....I..K....EF.Q.SIQ.TRN---
AV.A..IQYH.DPQPV....FTI.PHY
Semlikiforestvirus_5                              I..D.SDNHDY.KI..AD.---
.AIEN.VRSS.K.A..GD.F.HG.....I..K....EF.Q.SIQ.TRN---
AV.A..IQYH.DPQPV....FTI.PHY
Babankivirus
F.YDQSGAASTNKY...SLEQD.T..EGTMDDIKIS..GP.RRLSYK.Y.L..K......VT.SIASS
N----SATS.TMARKI.PK.V....YDLPPV.
Sindbisvirus_1
F.YDQSGAASSNKY...SLEQD.T..EGTMDDIKIS..GP.RRLSYK.Y.L..K......VT.SIASS
N----SATS.TMARKI.PK.V....YDLPPV.
Sindbisvirus_2
F.YDQSGAASSNKY...SLEQD.T..EGTMDDIKIS..GP.RRLSYK.Y.L..K......VT.SIASS
```

```
N----SATS.TMARKI.PK.V....YDLPPV.
Sindbisvirus_3
F.YDQSGAASANKY...SLKQD.T..EGTMDDIKIS..GP.RRLSYK.Y.L..K......VT.SIVSS
N----SATS.TLARKI.PK.V....YDLPPV.
Sindbisvirus_4
F.YDQSGAASANKY...SLKQD.T..EGTMDDIKIS..GP.RRLSYK.Y.L..K......VT.SIVSS
KNGSFSATS.TLARKI.PK.V....YDLPPV.
Sindbisvirus_5
F.YDQSGAASSNKY...TLEQD.T..EGTMDDIKIS..GP.GRLSYK.Y.L..K......VT.SIASS
N----SATS.TMARKI.PK.V....YDLPPV.
Auravirus
L.YNRAGTA.SAR..F.G.GVPP.IQEGAIADFK.F..KP.LHLSHK.Y.VIVK......ITTSLKVH
G----SDQT.TI.MRVGYK.V....YTLPPM.
Ockelbovirus
F.YD.SGAASTNKY...SFEQD.T..EGTMDDIKIS..GP.RRLSYK.Y.L..K......VT.SIASS
N----SATS.TMARKI.PK.V....YDLPPV.

                                          510          520          530
540        550        560        570        580        590        600

....|....|....|....|....|....|....|....|....|....|....|....|....|...
.|....|....|....|....|....|....
gi|28193970|gb|AAO33345.1|                                            str
GKDVTCTTYARRTDVTDENLEMHVPPHIPDATLVKVQLNNVTITPPSGATVRYNCSCKDGNRQSETSR
EVVLSGCAEAKCHAAVVDGKVWQYSSRFVPR-
AMH_001346
.....................................I.SH.............................
.AL.............................-
BAR_S2_3526
.....................................I.S..............................
.A.............................-
BSA_S4_2268
.....................................I.S..............................
.A.............................-
BSA_S4_2265
.....................................I.S..............................
.A.............................-
GSA_S1_936
.....................................I.S..............................
.A.............................-
BAR_S2_3527
.....................................I.S..............................
.A.............................-
GSA_S5_4278
.....................................I.S..............................
.A.............................-
AMH_001056
..EIP....QQT.AE.V.EID..M..DT..R..LSQ.SG..K..-VG.KK.K...T.GT..-
VGT.NSDKTINT.LIEQ..VS.T.H.K..FN.P....-
O'nyong-nyongvirus
..ELP.S..VHT.AA.A.EI.V.M..DT..Y..MTQ.AG..K..-VD.Q....K.K.DGS.-
EGLITADK.INN.KVDQ..T..TNH.K...N.PLT..N
Chikungunyavirus_1
.RELP.S...QS.AA.A.EI.V.M..DT..R..MSQ.SG..K..-VNSQ....K.N.G.SS-
EGL.TTDK.INN.KVDQ.....TNH.K...N.PL...N
Chikungunyavirus_2
```

.RELP.S...QS.AA.A.EI.V.M..DT..R..MSQ.SG..K..-VNSQ....K.N.G.SS-
EGL.TTDK.INN.KVDQ.....TNH.K...N.PL...N
Chikungunyavirus_3
.RELP.S...QS.AA.A.EI.V.M..DT..R..MSQ.SG..K..-VNSQ....K.N.G.SS-
EGL.TTDK.INN.KVDQ.....TNH.K...N.PL...N
RossRivervirus_1
.VELP..S.QLT.AP...EID..T..D...R..LSQTAG..K..-AG.R.I....T.GRD.-
VGT..TDKTINT.KIDQ.....TSHDK..FT.P....-
RossRivervirus_2
.VELP..S.QLT.AP...EID..T..D...R..LSQTAG..K..-AG.R.I....T.GRD.-
VGT..TDKTINT.KIDQ.....TSHDK..FT.P....-
RossRivervirus_3
.VELP..S.QLT.AP...EID..T..D...R..LSQTAG..K..-AG.R.I....T.GRD.-
VGT..TDKTINT.KIDQ.....TSHDK..FT.P....-
RossRivervirus_4
.VELP..S.QLT.AP...EID..T..D...R..LSQTAG..K..-AG.R.I....T.GRD.-
VGT..TDKTINT.KIDQ.....TSHDK..FT.P....-
Getahvirus_1
.IE.P....QLT.AP.E.EID..T..D...T..LSQ.SG..K..-AG.K.I....T.GS..-
VGT..SDKTINS.KI.Q.....TNHDK...T.S....-
getahvirus_2
.IE.P....QLT.AP.E.EID..T..D...I..LSQ.SG..K..-AG.K.I....T.GS..-
VGT...DKTINS.KI.Q.....TNHDK...T.S....-
Getahvirus_3
.IE.P....QLT.AP.E.EID..T..D...I..LSQ.SG..K..-AG.K.I....T.GS..-
VGT..SDKTINS.KI.Q.....TNHDK...T.S....-
Getahvirus_4
.IE.P....QLT.AP.E.EID..T..D...I..LSQ.SG..K..-AG.K.I....T.GS..-
VGT..SDKTINS.KI.Q.....TNHDK...T.S....-
Getahvirus_5
.IE.P....QLT.AP.E.EID..T..D...I..LSQ.SG..K..-AG.K.I....T.GS..-
VGT..SDKTINS.KI.Q.....TNHDK...T.S....-
Getahvirus_6
.IE.P....QLT.AP.E.EID..T..D...I..LSQ.SG..K..-AG.K.I....T.GS..-
VGT..SDKTINS.KI.Q.....TNHDK...T.S....-
Semlikiforestvirus_1
..EIP....QQT.AE.V.EID..M..DT..R..LSQ.SG..K..-VG.KK.K...T.GT..-
VGT.NSDKTINT.LIEQ..VS.T.H.K..FN.P....-
Semlikiforestvirus_2
..EIP....QQT.AE.V.EID..M..DT..R..LSQ.SG..K..-VG.KK.K...T.GT..-
VGT..SDKTINT.LIEQ..VS.T.H.K..FN.P....-
Semlikiforestvirus_3
..EIP....QQT.AE.V.EID..M..DT..R..LSQ.SG..K..-VG.KK.K...T.GT..-
VGT.NSDMTINT.LIEQ..VS.T.H.K..FN.P....-
Semlikiforestvirus_4
..EIP....QQT.AE.V.EID..M..DT..R..LSQ.SG..K..-VG.KK.K...T.GT..-
VGT.NSDMTINT.LIEQ..VS.T.H.K..FN.P....-
Semlikiforestvirus_5
..EIP....QQT.AE.V.EID.RM..DT..R..LSQ.SG..Q..-VG.KK.K...T.GT..-
VGT.NSDMTINT.LIEQ..VS.T.H.K..FN.P....-
Babankivirus
..KIP..V.D.LKET.AGYIT..R.GPHAYTSYLEESSGK.YAK....KNIT.E.Q.G.YKTGTV.T.
TEITGCT.IKQ.V.YKS.QTK.VFN.PDLI.-
Sindbisvirus_1
..KIP..V.D.LKET.AGYIT..R.GPHAYTSYLEESSGK.YAK....KNIT.E.K.G.YKTGTV.T.
TEITGCT.IKQ.V.YKS.QTK.VFN.PDLI.-


Sindbisvirus_2
..KIP..V.D.LKET.AGYIT..R.GPHAYTSYLEESSGK.YAK....KNIT.E.K.G.YKTGTV.T.
TEITGCT.IKQ.V.YKS.QTK.VFN.PDSI.-
Sindbisvirus_3
..KIP..V.D.LKET.AGYIT..R.RPHAYTSYLEESSGK.YAK....KNIT.E.K.G.YKTGTVST.
TEITGCT.IKQ.V.YKS.QTK.VFN.PDLI.-
Sindbisvirus_4
..KIP..V.D.LKET.AGYIT..R.RPHAYTSYLEESSGK.YAK....KNIT.E.K.G.YKTGTVST.
TEITGCT.IKQ.V.YKS.QTK.VFN.PDLI.-
Sindbisvirus_5
..KIP..V.D.LKET.AGYIT..R.GPHAYTSYL.ESSGK.YAK....KNIT.E.K.G.YKTGTV.T.
TEITGCT.IKQ.V.YKS.QTK.VFN.PDSI.-
Auravirus
.TQIP.L..E.TREKSAGYVT..R.GQQSITM.MEESGGE.YVQ.T..RN.T.E.K.G.FKTGTV.A.
TKIDGCTERKQ.I.ISA.HVK.VFN.PDLI.-
Ockelbovirus
..KIP..V.D.LKET.AGYIT..R.GPHAYTSYLEESSGK.YAK....KNIT.E.K.G.YKTGTV.T.
TEITGCT.IKQ.V.YKS.QTK.VFN.PDLI.-
610 620 630
640 650 660 670 680 690 700
gi|28193970|gb|AAO33345.1| str
HEKTELKGKVHVPFPLTNTTCRVSVATKPIEKHARGLLTLVLHPEYPTLLSYRYLGAEPRPVNRWISE
PTEVQIPVSKEGIEYRWGNNEPERRWAQHTTT
AMH_001346
A
BAR_S2_3526
BSA_S4_2268
BSA_S4_2265
GSA_S1_936
BAR_S2_3527
GSA_S5_4278
AMH_001056
ADEPAR....I....D.I....PM.RE.TVI.GKREV..H...DH...F...T..ED.QYHEE.VTA
AV.RT...PV..M..H....D.V.L.S.L..E
O'nyong-nyongvirus
S.QGDR...I.I....V......PK.RN.TVTYGKNRV..L...DH.......AM.RI.DYHEE..TN
KK.IS.T.PA..L.VT....D.YKY.P.LS.N
Chikungunyavirus_1
A.SGDR....I....A.V....PK.RN.TVTYGKNQVIML.Y.DH.......NM.E..NYQEE.VTH

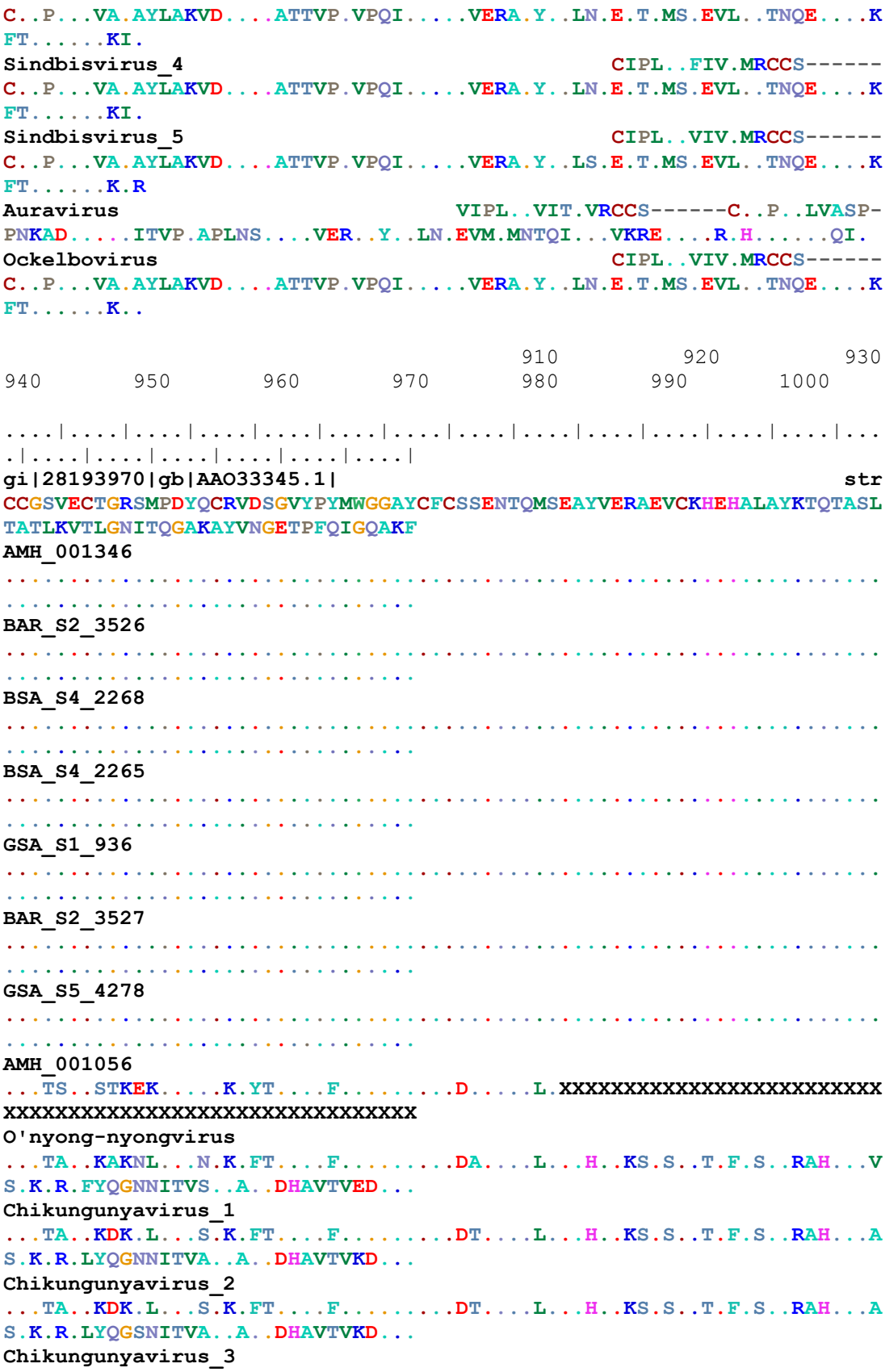
KK.IRLT.PT..L.VT.....YKY.P.LS.N
Chikungunyavirus_2
A.FGDR.....I....A.V....PK.RN.TVTYGKNQVIML.Y.DH.......NM.E..NYQEE.VTH
KK.IRLT.PT..L.VT.....YKY.P.LS.N
Chikungunyavirus_3
A.SGDR.....I....A.V....PK.RN.TVTYGKNQVIML.Y.DH.......NM.E..NYQEE.VTH
KK.IRLT.PT..L.VT.....YKY.P.LS.N
RossRivervirus_1
ADQ.ARR..........V....PL.RA.DVTYGKKEV..R...DH...F...S.....H.YEE.VDK
FS.RI...TE.....Q....P.V.L...L..E
RossRivervirus_2
ADQ.AR..........V....PL.RA.DVTYGKKEV..R...DH...F...S.....H.YEE.VDK
FS.RI...TE.....Q....P.V.L...L..E
RossRivervirus_3
ADQ.ARR..........V....PL.RA.DVTYGKKEV..R...DH...F...S.....H.YEE.VDK
FS.RI...TE.....Q....P.V.L...L..E
RossRivervirus_4
ADQ.ARR..........V....PL.RA.DVTYGKKEV..R...DH...F...S.....H.YEE.VDK
FS.RI...TE.....Q....P.V.L...L..E
Getahvirus_1
ADQLSR........L...S....P..RA.GVTYGKRE..VK...DH....T..S...D...YEE..DR
YV.RT...TED.........P.V.L...L..E
getahvirus_2
ADQLSR..........DS....PL.RA.GVTYGKRE..VK...DH....T..S...D...YEE..DR
YA.RT...TE.........P.V.L...L..E
Getahvirus_3
ADQLSR...........S....P..RA.GVTYGKRE..VK...DH....T..S...D...YEE..DR
YV.RT...TED.........P.V.L...L..E
Getahvirus_4
ADQLSR...........S....PL.RA.GVTYGKRE..VK...DH....T..S...D...YEE..DR
YV.RT...TED.........P.V.L...L..E
Getahvirus_5
ADQLSR...........S....P..RA.GVTYGKRE..VK...DH....T..S...D...YEE..DR
YV.RT...TED.........P.V.L...L..E
Getahvirus_6
ADQLSR...........S....P..RA.GVTYGKRE..VK...DH....T..S...D...YEE..DR
YV.RT...TED.........P.V.L...L..E
Semlikiforestvirus_1
ADEPAR.....I....D.I....PM.RE.TVI.GKRKV..H...DH...F...T..ED.QYHEE.VTA
AV.RT...PVD.M..H....D.V.L.S.L..E
Semlikiforestvirus_2
ADEPAR.....I....D.I....PM.RE.TVI.GKREV..H...DH...F...T..ED.QYHEE.VTA
AV.RT...PVD.M..H....D.V.L.S.L..E
Semlikiforestvirus_3
ADEPAR.....I....D.I....PM.RE.TVI.GKREV..H...DH...F...T..ED.QYHEE.VTA
AV.RT...PVD.M..H....D.V.L.S.L..E
Semlikiforestvirus_4
ADEPAR.....I....D.I....PM.RE.TVI.GKREV..H...DH...F...T..ED.QYHEE.VTA
AV.RT...PVD.M..H....D.V.L.S.L..E
Semlikiforestvirus_5
ADEPAR.....I....D.I....PM.RE.TVI.GKREV..H...DH...F...T..ED.QYHEE.VTA
AV.RT...PVD.M..H....D.M.L.S.L..E
Babankivirus
.ADHTAQ..L.L..K.IPS..M.P..HA.NVI.GFKHIS.Q.DTDHL...TT.R...N.E.TTE..IG
K.VRNFT.DRD.L..I...H..V.VY..ESAP
Sindbisvirus_1


.ADHTAQ..L.L..K.IPS..M.P..HA.NVV.GFKHIS.Q.DTDHL...TT.R...N.E.TTE..IG
K.VRNFT.DRD.L..I...H..V.VY..ESAP
Sindbisvirus_2
.ADHTAQ..L.L..K.IPS..M.P..HA.NVV.GFKHIS.Q.DTDHL...TT.R...N.E.TTE..IG
N.VRNFT.DRD.L..I...H..V.VY..ESAP
Sindbisvirus_3
.DDHTAQ..L.L..K.IPS..M.P..HA.NVI.GFKHIS.Q.DTDHL...TT.R...N.E.TTE..VG
K.VRNFT.DRD.L..I...H..V.VY..ESAP
Sindbisvirus_4
.DDHTAQ..L.L..K.IPS..M.P..HA.NVI.GFKHIS.Q.DTDHL...TT.R...N.E.TTE..VG
K.VRNFT.DRD.L..I...H..V.VY..ESAP
Sindbisvirus_5
.ADHTAQ..L.L..K.IPS..M.P..HA.NVV.GFKHIS.Q.DTDHL...TT.R...N.E.TTE..IG
K.VRNFT.DRD.L..I...H..V.VY..ESAP
Auravirus
.TDHTAQ..L.I....QQAQ.T.PL.HL.GV...YRSMS.T..A.H....TT.H..EN.Q.TAE..VG
SVTRNFSITIQ.F..T...QK.V.VY..ESAP
Ockelbovirus
.ADHTAQ..L.L..K.IPS..M.P..HA.NVI.GFKHIS.Q.DTDHL...TT.R...N.E.TTE..IG
K.VRNFT.DRD.L..I...H..V.VY..ESAP
710
720
730
740
750
760
770
780
790
800
....|....|....|....|....|....|....|....|....|....|....|....|....|...
.|....|....|....|....|....|....
gi|28193970|gb|AAO33345.1|
str
GSAHGHPHEIITYYYHSHPTTTVVACVTAAAVTLVMMCVGCSACRVARTRCLTPYVLAPGSRVPLILG
LLCCAKGARAEQYEPTPWWVWAP---------
AMH_001346
.............................V.....................................
.........................---------
BAR_S2_3526
.............................V...A.................................
.........................---------
BSA_S4_2268
.............................V...A.................................
.........................---------
BSA_S4_2265
.............................V...A.................................
.........................---------
GSA_S1_936
.............................V...A.T...............................
.........................---------
BAR_S2_3527
.............................V...A.................................
.........................---------
GSA_S5_4278
.............................V...A.................................
.........................---------
AMH_001056
.KP..W..Q.VQ...GLY.AA..S.VAGMSLLA.ISIFAS.YMLAA..SK.....A.T..AA..WT..
I....PR.H.ASVAE.MAYL.DQNQALFWLEF
O'nyong-nyongvirus
.T.........L...ELY....IAVLAA.SI.ITSLVGLSLGM.IC..R..I...E.T..ATI.FL..
V....RT.K.AS.YEAATYL.NEQQPLFWLQL

Chikungunyavirus_1
.T.........L...ELY..M...VVSV.SFIL.S.VG.AVGM.MC..R..I...E.T..AT..FL.S
.I..IRT.K.AT.QEAAVYL.NEQQPLFWLQA
Chikungunyavirus_2
.T.........L...ELY..M...VLSV.SFIL.S.VG.AVGM.MC..R..I...E.T..AT..FL.S
.I..IRT.K.AT.QEAAVYL.NEQQPLFWLQA
Chikungunyavirus_3
.T.........L...ELY..M...VVSV.SFIL.S.VG.AVGM.MC..R..I...E.T..AT..FL.S
.I..IRT.K.AT.QEAAVYL.NEQQPLFWLQA
RossRivervirus_1
.KP..W.....Q...GLY.AA.IA.XSG.SLMA.LTLAAT.CMLAT..RK.....A.T..AV...T..
.....PR.N.ASFAE.MAYL.DENKTLFWMEF
RossRivervirus_2
.KP..W.....Q...GLY.AA.IA.VSG.SLMA.LTLAAT.CMLAT..RK.....A.T..AV...T..
.....PR.N.ASFAE.MAYL.DENKTLFWMEF
RossRivervirus_3
.KP..W.....Q...GLY.AA.IA.VSG.SLMA.LTLAAT.CMLAT..RK.....A.T..AV...T..
.....PR.N.ASFAE.MAYL.DENKTLFWMEF
RossRivervirus_4
.KP..W.....Q...GLY.AA.IA.VSG.SLMA.LTLAAT.CMLAT..RK.....A.T..AV...T..
.....PR.N.ASFAE.MAYL.DENKTLFWMEF
Getahvirus_1
.KP..W.....L...GLY.AA.IT.VSA.GLAVVLSLLAS.YMFAT..RK.....A.T..AV..VT..
V....PR.H.ASFAESMAYL.DENQTLFWLEL
getahvirus_2
.KP..W.....L...GLY.AA.IA.VSA.GLAVVLSLLAS.YMFAT..RK.....A.T..AI..VT..
V....PR.H.ASFAESMAYL.DENQTLFWLEL
Getahvirus_3
.KP..W.....L...GLY.AA.IA.VSA.GLAVVLSLLAS.YMFAT..RK.....A.T..AV..VT..
V....PR.H.ASFAESMAYL.DENQTLFWLEL
Getahvirus_4
.KP..W.....L...GLY.AA.IA.VSA.GLAVVLSLLAS.YMFAT..RK.....A.T..AV..VT..
V....PR.H.ASFAESMAYL.DENQTLFWLEL
Getahvirus_5
.KP..W.....L...GLY.AA.IA.VSA.GLAVVLSLLAS.YMFAT..RK.....A.T..AV..VT..
V....PR.H.ASFAESMAYL.DENQTLFWLEL
Getahvirus_6
.KP..W.....L...GLY.AA.IA.VSA.GLAVVLSLLAS.YMFAT..RK.....A.T..AV..VT..
V....PR.H.ASFAESMAYL.DENQTLFWLEL
Semlikiforestvirus_1
.KP..W..Q.VQ...GLY.AA..S.V.GMSLLA.ISIFAS.YMLAA..SK.....A.T..AA..WT..
I....PR.H.ASVAE.MAYL.DQNQALFWLEF
Semlikiforestvirus_2
.KP..W..Q.VQ...GLY.AA.AS.VAGMSLLA.ISIFAS.YMLAA..SK.....A.T..AA..WT..
I....PR.H.ASVAE.MAYL.DQNQALFWLEF
Semlikiforestvirus_3
.KP..W..Q.VQ...GLY.AA..S.V.GMSLLA.ISIFAS.YMLVA..SK.....A.T..AA..WT..
I....PR.H.ASVAE.MAYL.DQNQALFWLEF
Semlikiforestvirus_4
.KP..W..Q.VQ...GLY.AA..S.V.GMSLLA.ISIFAS.YMLVA..SK.....A.T..AA..WT..
I....PR.H.ASVAE.MAYL.DQNQALFWLEF
Semlikiforestvirus_5
.KP..W..Q.VQ...GLY.AA..S.V.GMSLLA.ISIFAS.YMLVA..SK.....A.T..AA..WT..
I....PR.H.ASVAE.MAYL.DQNQALFWLEF
Babankivirus
.DP..W....VQH...R..VY.IL.VAS.VVAMMIGVT.AALCACK..RE.....A...NAVI.TS.A

```
....VRS.N..TFTE.MSYL.SNSQPFFWVQL
Sindbisvirus_1
.DP..W....VQH...R..VY.IL.VAS..VAMMIGVT.AALCACK..RE.....A...NAVI.TS.A
....VRS.N..TFTE.MSYL.SNSQPFFWVQL
Sindbisvirus_2
.DP..W....VQH...R..VY.IL.VAS..VAMMIGVT.AALCACK..RE.....A...NAVI.TS.A
....VRS.N..TFTE.MSYL.SNSQPFFWVQL
Sindbisvirus_3
.DP..W....VQH...R..VY.IL.VAS.TVAMMIGVT.AVLCACK..RE.....A...NAVI.TS.A
....VRS.N..TFTE.MSYL.SNSQPFFWVQL
Sindbisvirus_4
.DP..W....VQH...R..VY.IL.VAS.TVAMMIGVT.AVLCACK..RE.....A...NAVI.TS.A
....VRS.N..TFTE.MSYL.SNSQPFFWVQL
Sindbisvirus_5
.DP..W....VQH...R..VY.IL.VAS..VAMMIGVT.AALCACK..RE.....A...NAVI.TS.A
....VRS.N..TFTE.MSYL.SNSQPFFWVQL
Auravirus
.NP..W....VRH...LY.FY..TVLSGMGLAICAGLVISILC.CK..RD.....Q...NAT..FLVT
.C..FQRTS.DEFTD.MGYL.QHSQTMFWIQL
Ockelbovirus
.DP..W....VQH...R..VY.IL.VAS..VAMMIGVT.AALCACK..RE.....A...NAVI.TS.A
....VRS.N..TFTE.MSYF.SNSQPFFWVQL

                                        810          820          830
840        850        860        870        880        890        900

....|....|....|....|....|....|....|....|....|....|....|....|....|...
.|....|....|....|....|....|....
gi|28193970|gb|AAO33345.1|       str              ----AALVLLCGLRK-----
CLCLTFLVILGLASPPTQAYEHTAVMSNQVGVPYKALINKPGFAPMILQIEVLQSSLIPSLELDYITC
EYKTVVPSPFVK
AMH_001346                                       ----..........-----
..........................................................................
...........
BAR_S2_3526                                      ----..........-----
..........................................................................
...........
BSA_S4_2268                                      ----..........-----
..........................................................................
...........
BSA_S4_2265                                      ----..........-----
..........................................................................
...........
GSA_S1_936                                       ----..........-----
..........................................................................
...........
BAR_S2_3527                                      ----..........-----
..........................................................................
...........
GSA_S5_4278                                      ----..........-----
..........................................................................
...........
AMH_001056                          AAPV.CILIITYCLRNVLCC.KS.S...L.S.G-
ATAR....ST..P.V..F....H.ER..YS.LT..MQ.VET..E.T.N.E.............Y..
O'nyong-nyongvirus
LIPLS.AIVV.NCL.LLPCC.KT....AVMSIGARTVT....AT.IP.T....C.T.VSR..YS..V.
```

```
EM.LQSVT.E.A.S..........IT...Y..
Chikungunyavirus_1
LIPL...IV..NCLRLLPCC.KM....AV.SVGAHTVS....VT.IP.T......T.V.R..YS..V.
EM.L.SVT.E.T.S...........I...Y..
Chikungunyavirus_2
LIPL...IV..NCLRLLPCC.KM....AV.SVGAHTVS....VT.IP.T......T.V.R..YS..A.
EM.L.SVT.E.T.S...........I...Y..
Chikungunyavirus_3
LIPL...IV..NCLRLLPCC.KM....AV.SVGAHTVS....VT.IP.T......T.V.R..YS..V.
EM.L.SVT.E.T.S...........I...Y..
RossRivervirus_1                    AAPA...A..ACCI.SLICC.KPFS...L.S.G-
ASAK......TIP.V..F....H.ERN..S..T..L..VET..E.T.N.E.............I.
RossRivervirus_2                    AAPA...A..ACCI.SLICC.KPFS...L.S.G-
ASVK......TIP.V..F....H.ERN..S..T..L..VET..E.T.N.E.............I.
RossRivervirus_3                    AAPA...A..ACCI.SLICC.KPFS...L.S.G-
ASAK......TIP.V..F....H.ERN..S..T..L..VET..E.T.N.E.............I.
RossRivervirus_4                    AAPA...A..ACCI.SLICC.KPFS...L.S.G-
ASAK......TIP.V..F....H.ERN..S..T..L..VET..E.T.N.E.............I.
Getahvirus_1
ATPL..III.VCCL.NLLCC.KP.S...LVS.GT.VVKS.....TIP.V..F....H.ERN..S..T.
.L...GT..E.T.N.E.............YI.
getahvirus_2
ATPL..III.VCCL.NLLCC.KP.S...LVS.GT.VVKS.....TIP.V..F....H.ERN..S..T.
.L...GT..E.T.N.E.............YI.
Getahvirus_3
ATPL..III.VCCL.NLLCC.KP.S...LVS.GT.VVKS.....TIP.V..F....H.ERN..S..T.
.L...GT..E.T.N.E.............YI.
Getahvirus_4
ATPL..III.VCCL.NLLCC.KP.S...LVS.GT.VVKS.....TIP.V..F....H.ERN..S..T.
.L...GT..E.T.N.E.............YI.
Getahvirus_5
ATPL..III.VCCL.NLLCC.KP.S...LVS.GT.VVKS.....TIP.V..F....H.ERN..S..T.
.L...GT..E.T.N.E.............YI.
Getahvirus_6
ATPL..III.VCCL.NLLCC.KP.S...LVS.GT.VVKS.....TIP.V..F....H.ERN..S..T.
.L...GT..E.T.N.E.............YI.
Semlikiforestvirus_1                AAPV.CILIITYCLRNVLCC.KS.S...L.S.G-
ATAR....ST..P.V..F....H.ER..YS.LT..MQ.VET..E.T.N.E.............Y..
Semlikiforestvirus_2                AAPV.CILIITYCLRNVLCC.KS.S...L.S.G-
ATAR....ST..P.V..F....H.ER..YS.LT..MQ.VET..E.T.N.E.............Y..
Semlikiforestvirus_3                AAPV.CILIITYCLRNVLCC.KS.S...L.S.G-
ATAR....ST..P.V..F....H.ER..YS.LT..MQ.VET..E.T.N.E.............Y..
Semlikiforestvirus_4                AAPV.CILIITYCLRNVLCC.KS.S...L.S.G-
ATAR....ST..P.V..F....H.ER..YS.LT..MQ.VET..E.T.N.E.............Y..
Semlikiforestvirus_5                AAPV.CILIITYCLRNVLCC.KS.S...L.S.G-
ATAR....ST..P.V..F....H.ER..YS.LT..MQ.VET..E.T.N.E.............Y..
Babankivirus                                  CIPL..VIV.MRCCS------
C..P...VA.AYLAKVD....ATTVP.VPQI.....VERA.Y..LN.E.T.MS.EVL..TNQE....K
FT......K..
Sindbisvirus_1                                CIPL..VIV.MRCCS------
C..P...VA.AYLAKVD....ATTVP.VPQI.....VERA.Y..LN.E.T.MS.EVL..TNQE....K
FT......K..
Sindbisvirus_2                                CIPL..V.V.MRCCS------
C..P...VA.AYLAKVD....ATTVP.VPQI.....VERA.Y..LN.E.T.MS.EVL..TNQE....K
FT......K.R
Sindbisvirus_3                                CIPL..FIV.MRCCS------
```


C..P...VA.AYLAKVD....ATTVP.VPQI.....VERA.Y..LN.E.T.MS.EVL..TNQE....K
FT......KI.
Sindbisvirus_4 CIPL..FIV.MRCCS------
C..P...VA.AYLAKVD....ATTVP.VPQI.....VERA.Y..LN.E.T.MS.EVL..TNQE....K
FT......KI.
Sindbisvirus_5 CIPL..VIV.MRCCS------
C..P...VA.AYLAKVD....ATTVP.VPQI.....VERA.Y..LS.E.T.MS.EVL..TNQE....K
FT......K.R
Auravirus VIPL..VIT.VRCCS------C..P..LVASP-
PNKAD.....ITVP.APLNS....VER..Y..LN.EVM.MNTQI...VKRE....R.H......QI.
Ockelbovirus CIPL..VIV.MRCCS------
C..P...VA.AYLAKVD....ATTVP.VPQI.....VERA.Y..LN.E.T.MS.EVL..TNQE....K
FT......K..
910 920 930
940 950 960 970 980 990 1000
gi|28193970|gb|AAO33345.1| str
CCGSVECTGRSMPDYQCRVDSGVYPYMWGGAYCFCSSENTQMSEAYVERAEVCKHEHALAYKTQTASL
TATLKVTLGNITQGAKAYVNGETPFQIGQAKF
AMH_001346
BAR_S2_3526
BSA_S4_2268
BSA_S4_2265
GSA_S1_936
BAR_S2_3527
GSA_S5_4278
AMH_001056
...TS..STKEK.....K.YT....F.........D.....L.XXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
O'nyong-nyongvirus
...TA..KAKNL...N.K.FT....F.........DA....L...H..KS.S..T.F.S..RAH...V
S.K.R.FYQGNNITVS..A..DHAVTVED...
Chikungunyavirus_1
...TA..KDK.L...S.K.FT....F.........DT....L...H..KS.S..T.F.S..RAH...A
S.K.R.LYQGNNITVA..A..DHAVTVKD...
Chikungunyavirus_2
...TA..KDK.L...S.K.FT....F.........DT....L...H..KS.S..T.F.S..RAH...A
S.K.R.LYQGSNITVA..A..DHAVTVKD...
Chikungunyavirus_3

...TA..KDK.L...S.K.FT....F.........DT....L...H..KS.S..T.F.S..RAH...A
S.K.R.LYQGNNITVA..A..DHAVTVKD...
RossRivervirus_1
...TS..SSKEQ.....K.YT....F.........D.....L.....D.SD....D..S...AH....
K..IRISY.T.N.TTE.F....HAVNV.GS..
RossRivervirus_2
...TS..SSKEQ.....K.YT....F.........D.....L.....D.SDI...D......AH....
K..IRISY.T.N.TTE.F....HAVNV.GS..
RossRivervirus_3
...TS..SSKEQ.....K.YT....F.........D.....L.....D.SD....D..S...AH....
K..IRISY.T.N.TTE.F....HAVNV.GS..
RossRivervirus_4
...TS..SSKEQ.....K.YT....F.........D.....L.....D.SD....D..S...AH....
K..IRISY.T.N.TTE.F....HAVNV.GS..
Getahvirus_1
...TS..RSMER.....Q.YT....F.........DT....L.....D.SD....D..A...AH..AV
K..IRISY..LN.TTT.F....HTVTV.GSR.
getahvirus_2
...TS..RSKER.....Q.YT....F.........DT....L.....D.SD....D..A...AH..AM
K..IRISY..LN.TTT.F....HTVNV.GSR.
Getahvirus_3
...TS..RSMER.....Q.YT....F.........DT....L.....D.SD....D..A...AH..AM
K..IRISY..LN.TTT.F....HTVTV.GSR.
Getahvirus_4
...TS..RSMER.....Q.YT....F.........DT....L.....D.SD....D..A...AH..AM
K..IRISY..LN.T.T.F....HTVTV.GSR.
Getahvirus_5
...TS..RSMER.....Q.YT....F.........DT....L.....D.SD....D..A...AH..AM
K..IRISY..LN.TTT.F....HTVTV.GSR.
Getahvirus_6
...TS..RSMER.....Q.YT....F.........DT....L.....D.SD....D..A...AH..AM
K..IRISY..LN.TTT.F....HTVTV.GSR.
Semlikiforestvirus_1
...AS..STKEK.....K.YT....F.........D.....L.....D.SD..R.D..S...AH....
K.KVR.MY..VN.TVDV....DHAVT..GTQ.
Semlikiforestvirus_2
....S..STKEK.....K.YT....F.........D.....L.....D.SD....D..S...AH....
K.KVR.MY..VN.TVDV....DHAVT..GTQ.
Semlikiforestvirus_3
...AS..STKEK.....K.YT....F.........D.....L.....D.SD..R.D..S...AH....
K.KVR.MY..VN.TVDV....DHAVT..GTQ.
Semlikiforestvirus_4
...AS..STKEK.....K.YT....F.........D.....L.....D.SD..R.D..S...AH....
K.KVR.MY..VN.TVDV....DHAVT..GTQ.
Semlikiforestvirus_5
...AS..STKEK.....K.YT....F.........D.....L.....D.SD..R.D..S...AH....
K.KVR.MY..VN.TVDV....DHAVT..GTQ.
Babankivirus
....L..QPAAHA..T.K.FG....F.....Q...D...S........LSAD.ATD..Q.V.VH..AM
KVG.RIVY..T.SFLDV....V..GTSKDL.V
Sindbisvirus_1
....L..QPAAHA..T.K.FG....F.....Q...D...S........LSAD.ATD..Q.I.VH..AM
KVG.RIVY..T.SFLDV....V..GTSKDL.V
Sindbisvirus_2
....L..QPAAHA..T.K.FG....F.....Q...D...S........LSVD.ATD..Q.I.VH..AM
KVG.RIVY..T.SFLDV....V..GTSKDL.V


Sindbisvirus_3
....L..QPAAHA..T.K.FG....F.....Q...D...S........LSA..ASD..Q.I.VH..AM
KVG.RIVY..T.SFLDV....V..GTSKDL.V
Sindbisvirus_4
....L..QPAAHA..T.K.FG....F.....Q...D...S........LSAD.ASD..Q.I.VH..AM
KVG.RIVY..T.SFLDV....V..GTSKDL.V
Sindbisvirus_5
....L..QPAAHA..T.K.FG....F.....Q...D...S........LSAD.ATD..Q.I.VH..AM
KVG.RIVY..TISFLDV....V..GTSKDL.V
Auravirus
...T...PKGEKA..T.K.FT....FL....Q...D...S.L.DK...LSTD.ATD..E.VRVH...V
KSQ.RI.Y..S.AQVDVF...V..ARSKDM.L
Ockelbovirus
....L..QPAAHA..T.K.FG....F.....Q...D...S........LSAD.ATD..Q.I.VH..AM
KVG.RIVY..T.SFLDV....V..GTSKDL.V
1010
1020
1030
1040
1050
1060
1070
1080
1090
1100
gi|28193970|gb|AAO33345.1|
str
VLGPISAAWSPFNPKIVVYKDDVYNYDFPAYGAGQPGRFGDIQSRTVDSKDLYARTHLRLDRPASGNI
HVPYTQIPSGFKYWMQEKGEPLNRQAAFGCVV
AMH_001346
BAR_S2_3526
BSA_S4_2268
BSA_S4_2265
GSA_S1_936
BAR_S2_3527
GSA_S5_4278
AMH_001056
XXXXXXXXXXXXXXX.I....E.F.Q...P..S..............E.N....N.A.K.A..SP.TV
......T.......LK...TA..TK.P...QI
O'nyong-nyongvirus
.I..L.S.....DN......GE...M.Y.PF...R..Q........P....V..N.Q.I.Q...A.A.
....S.A.......LK...AS.QHT.P...QI
Chikungunyavirus_1
IV..M.S..T..DN......G....M.Y.PF...R..Q........PE.E.F..N.Q.V.Q..SA.TV
....S.A.......LK.R.AS.QHT.P...QI
Chikungunyavirus_2
IV..M.S..T..DN......G....M.Y.PF...R..Q........PE.E.V..N.Q.V.Q..SA.TV

....S.A.......LK.R.AS.QHT.P...QI
Chikungunyavirus_3
IV..M.S..T..DN......G....M.Y.PF...R..Q........PE.E.V..N.Q.V.Q..SA.TV
....S.A.......LK.R.AS.QHT.P...QI
RossRivervirus_1
IF....T.....DN............Q...P..S.............E......N.A.K.S..SP.VV
......T.......LK...SS..TK.P...KI
RossRivervirus_2
IF....T.....DN............Q...P..S.............E......N.A.K.S..SP.VV
......T.......LK...SS..TK.P...KI
RossRivervirus_3
IF....T.....DN............Q...P..S.............E......N.A.K.S..SP.VV
......T.......LK...SS..TK.P...KI
RossRivervirus_4
IF....T.....DN............Q...P..S.............E......N.A.K.S..SP.VV
......T.......LK...SS..TK.P...KI
Getahvirus_1
TF....T..T..DN......N....H...P..S.............E......N.A.K.S..S..TV
......T.......IK.R.TS..DK.P....I
getahvirus_2
TF....T..T..DN..A...N....Q...P..S.............E......N.A.K.S..S..TV
......T.......IK.R.TS..DK.P...II
Getahvirus_3
TF....T..T..DN......N....Q...P..S.............E......N.A.K.S..S..TV
......T.......IK.R.TS..DK.P....I
Getahvirus_4
TF....T..T..DN......N....Q...P..S.............E......N.A.K.S..S..TV
......T.......IR.R.TS..DK.P....I
Getahvirus_5
TF....T..T..DN......N....Q...P..S.............E......N.A.K.S..S..TV
......T.......IK.R.TS..DK.P....I
Getahvirus_6
TF....T..T..DN......N....Q...P..S.............E......N.A.K.S..S..TV
......T.......IK.R.TS..DK.P....I
Semlikiforestvirus_1
IF..L.S..T..DN.......E.F.Q...P..S.............E.N....N.A.K.A..SP.TV
......T.......LK...TA..TK.P...QI
Semlikiforestvirus_2
IF..L.S..T..DN.......E.F.Q...P..S.............E.N....N.A.K.A..SP.TV
......T.......LK...TA..TK.P...QI
Semlikiforestvirus_3
IF..L.S..T..DN.......E.F.Q...P..S.............E.N....N.A.K.A..SP.MV
......T.......LK...TA..TK.P...QI
Semlikiforestvirus_4
IF..L.S..T..DN.......E.F.Q...P..S.............E.N....N.A.K.A..SP.MV
......T.......LK...TA..TK.P...QI
Semlikiforestvirus_5
IF..L.S..T..DN.......E.F.Q...P..S.............E.N....N.A.K.A..SP.MV
......T.......LK...TA..TK.P...QI
Babankivirus
IA.....SFT..DH.V.IHRGL.......E...MK..A.....ATSLT....I.S.DI..LK.SAK.V
......AA...EM.KNNS.R..QET.P...KI
Sindbisvirus_1
IA.....SFT..DH.V.IHRGL.......E...MK..A.....ATSLT....I.S.DI..LK.SAK.V
......AA...EM.KNNS.R..QET.P...KI
Sindbisvirus_2


IA.....LFT..DH.V.INRGL.......E...MK..A.....ATSLT....I.S.DI..LK.SAK.V
......AA...EM.KNNS.R..QET.P...KI
Sindbisvirus_3
IA.....SFT..DH.V.IHRGL.......E...MK..A.....ATSLT....I.S.DI..LK.SAK.V
......AS...EM.KNNS.R..QET.P...KI
Sindbisvirus_4
IA.....SFT..DH.V.IHRGL.......E...MK..A.....ATSLT....I.S.DI..LK.SAK.V
......AS...EM.KNNS.R..QET.P...KI
Sindbisvirus_5
IA.....LFT..DH.V.IHRGL.......E...MK..A.....ATSLT....I.S.DI..LK.SAK.V
......AA...EM.KNNS.R..QET.P..RKI
Auravirus
IA..L.TTF...DN.VII.HGK.......EF...T..A...V.ASSTTGS..L.N.AIH.Q..EAR..
......A....EF.KNNS.Q..SDT.P...K.
Ockelbovirus
IA.....SFT..DH.V.IHRGL.......E...MK..V.....ATSLT....I.S.DI..LK.SAK.V
......AA...EM.KNNS.R..QET.P...KI
1110
1120
1130
1140
1150
1160
1170
1180
1190
1200
gi|28193970|gb|AAO33345.1|
str
KTNPVRADNCAYGNIPISVDIPDAMFTRVVDIPAVSHLQCIDVSCAHSAGFGGVAKLRFMADRTGKCA
IHTASKIVAIQEATVDLPEIAVGAGGGTAEIT
AMH_001346
BAR_S2_3526
BSA_S4_2268
BSA_S4_2265
GSA_S1_936
BAR_S2_3527
GSA_S5_4278
AMH_001056
.......M...V....V.MNL..SA...I.EA.TITD.T.TVAT.T..SD....LT.TYKT.KN.D.S
V.SH.NVATL....AKVK--------TAGKV.
O'nyong-nyongvirus
A......V...V....V.I.....A....T.A.SITDMS.EVA..T..SD...A.VIKYT.SKK....
V.SVTNA.T.R.PN..VK--------...QLQ
Chikungunyavirus_1
A......M...V..M...I.....A.......A.SLTDMS.EVSA.T..SD.....IIKYA.SKK....
V.SMTNA.T.W..EIEVE--------.NSQLQ

```
Chikungunyavirus_2
A......M...V..M...I.....A......A.SLTDMS.EVSA.T..SD.....IIKYA.SKK....
V.SMTNA.T.R..EIEVE--------.NSQLQ
Chikungunyavirus_3
A......M...V..M...I.....A......A.SLTDMS.EVSA.T..SD.....IIKYA.SKK....
V.SMTNA.T.R..EIEVE--------.NSQLQ
RossRivervirus_1
.......MD..V.S..V.M....SA......A...TD.S.QV.V.T..SD.....T.SYKT.KP....
V.SH.NVATL......VK--------EDGKV.
RossRivervirus_2
.......MD..V.S..V.M....SA......A...TD.S.QVAV.T..SD.....T.SYKT.KP....
V.SH.NVATL......VK--------EDGKV.
RossRivervirus_3
.......MD..V.S..V.M....SA......A...TD.S.QV.V.T..SD.....T.SYKT.KP....
V.SH.NVATL......VK--------EDGKV.
RossRivervirus_4
.......MD..V.S..V.M....SA......A...TD.S.QVAV.T..SD.....T.SYKT.KP....
V.SH.NVATL......VK--------EDGKV.
Getahvirus_1
.......E...V....V.M....TA....I.A...TN.E.QVAV.T..SD...I.T.T.KT.KP....
V.SH.NVAT....A..IK--------TDGK..
getahvirus_2
.......E...V....V.M....SA....I.A...TN.E.QVAV.T..SD...I.T.T.KT.KP....
V.SH.NVAT....A..IK--------TDGK..
Getahvirus_3
.......E...V....V.M....TA....I.A...TN.E.QVAV.T..SD...I.T.T.KT.KP....
V.SH.NVAT....A..IK--------TDGK..
Getahvirus_4
.......E...V....V.M....SA....I.A...TN.E.QVAV.T..SD...I.T.T.KT.KP....
V.SH.NVAT....A..IK--------TDGK..
Getahvirus_5
.......E...V....V.M....TA....I.A...TN.E.QVAV.T..SD...I.T.T.KT.KP....
V.SH.NVAT....A..IK--------TDGK..
Getahvirus_6
.......E...V....V.M....TA....I.A...TN.E.QVAV.T..SD...I.T.T.KT.KP....
V.SH.NVAT....A..IK--------TDGK..
Semlikiforestvirus_1
.......M...V....V.MNL..SA...I.EA.TITD.T.TVAT.T..SD....LT.TYKT.KN.D.S
V.SH.NVATL....AKVK--------TARKV.
Semlikiforestvirus_2
.......M...V....V.MNL..SA...I.EA.TITD.T.TVAT.T..SD....LT.KYKT.KN.D.S
V.SH.NVATL....AKVK--------TAGKV.
Semlikiforestvirus_3
.......M...V....V.MNL..SA...I.EA.TIID.T.TVAT.T..SD....LT.TYKT.KN.D.S
V.SH.NVATL....AKVK--------TAGKV.
Semlikiforestvirus_4
.......M...V....V.MNL..SA...I.EA.TIID.T.TVAT.T..SD....LT.TYKTNKN.D.S
V.SH.NVATL....AKVK--------TAGKV.
Semlikiforestvirus_5
.......M...V....V.MNL..SA...I.EA.TIID.T.TVAT.T..SD....LT.TYKTAKN.D.S
V.SH.NVATL....AKVK--------TAGKV.
Babankivirus
AV..L..VD.S.......I...N.A.I.TS.A.L..TVK.DVSE.TY..D...M.T.QYVS..E.Q.P
V.SH.STATL..S..HVL--------EKGAV.
Sindbisvirus_1
AV..L..VD.S.......I...N.A.I.TS.A.L..TVK.DVSE.TY..D...M.T.QYVS..E.Q.P
```

V.SH.STATL..S..HVL--------EKGAV.
Sindbisvirus_2
AV..L..VD.S.......I...N.A.I.TS.A.L..TVK.DVSE.TY..D...M.T.QYVS..E.Q.P
V.SH.STATL..S..HVL--------EKGAV.
Sindbisvirus_3
AV..L..VD.S.......I...N.A.I.TS.A.L..TVK.EVSE.TY..D...M.T.QYVS..E.Q.P
V.SH.STATL..S..HVL--------EKGAV.
Sindbisvirus_4
AV..L..VD.S.......I...N.A.I.TS.A.L..TVK.EVSE.TY..D...M.T.QYVS..E.Q.P
V.SH.STATL..S..HVL--------EKGAV.
Sindbisvirus_5
AV..LQ.VD.S.......I...N.A.I.TS.A.L..TVK.DVSE.TY..D...M.T.QYVS..E.Q.P
V.SH.STATL..S..HVL--------EKGAV.
Auravirus NV..L...K..V.SL.........A....S-
E.LP.L.K.TVT..TY.TDY...LV.TYES..A.Q..V.SH.STAVLRDPS.YVE--------
QKG.T.
Ockelbovirus
AV..L..VD.S.......I...N.A.I.TS.A.L..TVK.DVSE.TY..D...M.T.QYVS..E.Q.P
V.SH.STATL..S..HVL--------EKGAV.

1210 1220 1230
1240 1250 1260 1270 1280

....|....|....|....|....|....|....|....|....|....|....|....|....|...
.|....|....|....|.
gi|28193970|gb|AAO33345.1| str
FHYSTALASPSFDVQLCSAHVTCDTKCEPPKDHIVPFAAQHMSNDMPNLSATAMNWVTGLGTSIGTFV
FLFLGILLVVTIIRCFTR
AMH_001346
....
....
BAR_S2_3526
....
....
BSA_S4_2268
....
....
BSA_S4_2265
....
....
GSA_S1_936
....
....
BAR_S2_3527
....
....
GSA_S5_4278
....
....
AMH_001056
L.F...S.....V.S....KA..SAS.........Y..S.SNVVF.DM.G..LS..QKISGGL.A.A
IGAILV.V...C.GLRR-
O'nyong-nyongvirus
IAF......AE.K..I..TL.H.SAT.H.......NYPSP.TTLGVQDI.T...S..QKITGGV.LV.
AIAAL..II.LCVSFSRH
Chikungunyavirus_1

ISFL.....AE.R..V..TQ.H.AAE.H.......NYP.S.TTPGVQDI.....S..QKITGGV.LV.
AVAAL..I..LCVSFSRH
Chikungunyavirus_2
ISF......AE.R..V..TQ.H.AAE.H.......NYP.S.TTPGVQDI.....S..QKITGGV.LV.
AVAAL..I..LCVSFSRH
Chikungunyavirus_3
ISFL.....AE.R..V..TQ.H.AAE.H.......NYP.S.TTPGVQDI.....S..QKITGGV.LV.
AVAAL..I..LCVSFSRH
RossRivervirus_1
V.F...S...A.K.SV.D.KT..TAA..........YG.S.NNQVF.DM.G...T..QR.ASGL.GLA
LIAVVV.VL..C.TMRR-
RossRivervirus_2
V.F...S...A.K.SV.D.KT..TAA..........YG.S.NNQVF.DM.G...T..QRMASGL.GLA
LIAVLV.VL..C.TMRR-
RossRivervirus_3
V.F...S...A.K.SV.D.KT..TAA..........YG.S.NNQVF.DM.G...T..QR.ASGL.GLA
LIAVVV.VL..C.TMRR-
RossRivervirus_4
V.F...S...A.K.SV.D.KT..TAA..........YG.S.NNQVF.DM.G...T..QR.ASGL.GLA
LIAVVV.VL..C.TMRR-
Getahvirus_1
L.F...S...A.K.SV...KT..MAA..........YG.S.NNQVF.DM.G...T..QRVAGGL.GLT
LAAVAV.IL..CVTMRR-
getahvirus_2
L.F...S...A.K.SV...KT..TAA..........YG.S.NNQVF.DM.G...T..QRVAGGL.GLT
LAAVAV.IL..CVTMRR-
Getahvirus_3
L.F...S...A.K.SV...KT..MAA..........YG.S.NNQVF.DM.G...T..QRVAGGL.GLT
LAAVAV.IL..CVTMRR-
Getahvirus_4
L.F...S...A.K.SV...KT..MAA..........YG.S.NNQVF.DM.G...T..QRVAGGL.GLT
LAAVAV.IL..CVTMRR-
Getahvirus_5
L.F...S...A.K.SV...KT..MAA..........YG.S.NNQVF.DM.G...T..QRVAGGL.GLT
LAAVAV.IL..CVTMRR-
Getahvirus_6
L.F...S...A.K.SV..VKT..MAA..........YG.S.NNQVF.DM.G...T..QRVAGGL.GLT
LAAVAV.IL..CVTMRR-
Semlikiforestvirus_1
L.F...S.....V.S....KA..SAS..........Y..S.SNVVF.DM.G..LS..QKISGGL.A.A
IGAILV.V...C.GLRR-
Semlikiforestvirus_2
L.F...S.....V.S....KA..SAS..........Y..S.SNVVF.DM.G..LS..QKISGGL.A.A
IGAILV.V...C.GLRR-
Semlikiforestvirus_3
L.F...S.....V.S....RA..SAS..........Y..S.SNVVF.DM.G..LS..QKISGGL.A.A
IGAILV.V...C.GLRR-
Semlikiforestvirus_4
L.F...S.....V.S....RA..SAS..........Y..S.SNVVF.DM.G..LS..QKISGGL.A.A
IGAILV.V...C.GLRR-
Semlikiforestvirus_5
L.F...S.....V.S....RA..SAS..........Y..S.SNVVF.DM.G..LS..QKISGGL.A.A
IGAILV.V...C.GLR.-
Babankivirus
V.I...SPQAN.I.S..GKKT..NAE.K..A....STPHKNDQEFQAAI.K.SWS.LFA.FGGASSLL
IIG.M.FACSMMLTSTR.

Sindbisvirus_1
V.F...SPQAN.I.S..GKKT..NAE.K..A....STPHKNDQEFQAAI.K.SWS.LFA.FGGASSLL
IIG.M.FACSMMLTSTR.
Sindbisvirus_2
V.F...SPQAN.I.S..GKKT..NAE.K..A....STPHKNDQEFQAAI.K.SWS.LFA.FGGASSLL
IIG.M.FACSMMLTSTR.
Sindbisvirus_3
V.F...SPQAN.I.S..GKKT..NAE.K..A....STPHKNDQEFQAAI.K.SWS.LFA.FGGASSLL
IIG.M.FACSMMLTSTR.
Sindbisvirus_4
V.F...SPQAN.I.S..GKKT..NAE.K..A....STPHKNDQEFQAAI.K.SWS.LFA.FGGASSLL
IIG.M.FACSMMLTSTR.
Sindbisvirus_5
V.F...SPQAN.I.S..GKKT..NAE.K..A....STPHKNDQEFQAAI.K.SWS.LFA.FGGASSLL
IIG.M.FACSMMLTSTR.
Auravirus
LKF..RSLQAD.E.SM.GTRT..HAQ.Q..TE.VMNRPQKSTPDFSSAI.K.SW..I.A.MGG.SSIA
AIAAIV.VIALVFTAQH.
Ockelbovirus
V.F...SPQAN.I.S..GKKT..NAE.K..A....STPHKNDQEFQAAI.K.SWS.LFA.FGGASSLL
IIG.T.FACSMMLTSTR.