University of Veterinary Medicine, Hungary

National Food Chain Safety Office,
Veterinary Diagnostic Directorate

Comparative analysis of West Nile Virus genomes

By
Randi Fivelstad

Supervisors:

Dr. Károly Erdélyi DVM, PhD, Dipl. ECZM

Dr. Ádám Dán DVM, PhD

Budapest, Hungary
2016

Table of content

1. Introduction

In 2003 and 2004 an incident of two different West Nile virus strains caused lethal encephalitis in a flock of geese and a goshawk in south-eastern Hungary, and during the summers of 2004 and 2005 a sparrow hawk and several goshawk fledglings succumbed to encephalitis. West Nile virus (WNV) which was originally classified as a group B arbovirus is now known to belong to positive sense single stranded RNA viruses of the Flaviviridae family. More precisely, within the Flavivirus genus, West Nile virus belongs to the Japanese encephalitis virus (JEV) complex, and is now the most widespread member of this group. The virus was first isolated a human patient in the West Nile district of Uganda in 1937, and later it was detected in mosquitoes, humans, horses and other hosts on most continents. The clinical impact of West Nile virus (WNV) varies in different regions and causes a wide range of symptoms, from mild influenza-like infections to fatal encephalitis, mainly in humans, horses and birds. The virus shows high levels of sequence diversity, and has in the recent years been classified into 4 lineages based on phylogenetic relatedness. West Nile Virus has been detected in several bird species and some mammals in Hungary by PCR from 2003 till today's date. However, there were so far only three samples from Hungary which have been fully sequenced for complete West Nile Virus genomes. The first complete genome sequence was obtained from a Goshawk in Hungary in 2004 (DQ116961), the second sample came from a horse in 2010 (KC496015), and the last one was obtained from a human in 2014 (KT3599349). This study has focused on the sequencing of the genomes of two West Nile virus genomes and their comparative genetic analysis in the context of earlier isolates from Hungary and Europe.

2. Literature review

2.1 Structure of West Nile virus

West Nile virus (WNV) belongs to the Flavivirus genus of the Flaviviridae family, and it's the most widespread mosquito-borne group among the Flaviviruses (Bakonyi, et al., 2013) belonging to the Japanese encephalitis virus (JEV) complex (Bakonyi, et al., 2006).

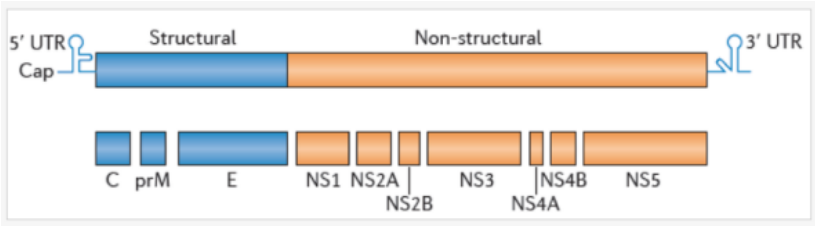


Figure 1 - Organisation of Flavivirus Genome (Beck, et al., 2013)

It was originally classified as a group B arbovirus based on epidemiological and antigenic criteria and as a Togavirus on an ultra-structural basis. This classification has been changed due to highly analytical molecular technologies that recognize a distinct group of positive-stranded RNA viruses belonging to the Flavivirus genus (Reid, et al., 2012). The Flaviviridae is usually divided into three main clusters; tick-borne viruses, mosquito-borne viruses and viruses with no arthropod vector (Reid, et al., 2012), where the West Nile Virus belong to the mosquito-borne group. The strains of the West Nile virus has been classified into different genetic lineages based on their differences in phylogeny (Szentpàli-Gavallèr, et al., 2014). There has been a suggestion of a subdivision into seven distinct genetic lineages, where lineage 1 has been detected on all continents (Bakonyi, et al., 2013). Lineages differ significantly from each other, and have wide genomic diversity (only 76-77% nucleotide identity). Lineage 1 can further be divided into three clades, where clade 1a comprises European, African and North American strains, clades 1b represents the Australian strains, and 1c comprises the Indian isolates (Bakonyi, et al., 2006). Lineage 2 West Nile virus strains used to be detected in Africa and Madagascar, but during the last decade they became endemic in Central-Eastern Europe and the Mediterranean. Lineage 3 consists of Rabensburg virus strains, and was isolated in the Czech Republic from Culex pipiens mosquitoes. In Russia in 1998 they isolated lineage 4 as a single virus (LEIV-Krnd88-190), and in India they isolated recently a lineage 5 of West Nile virus (Reid, et al., 2012).

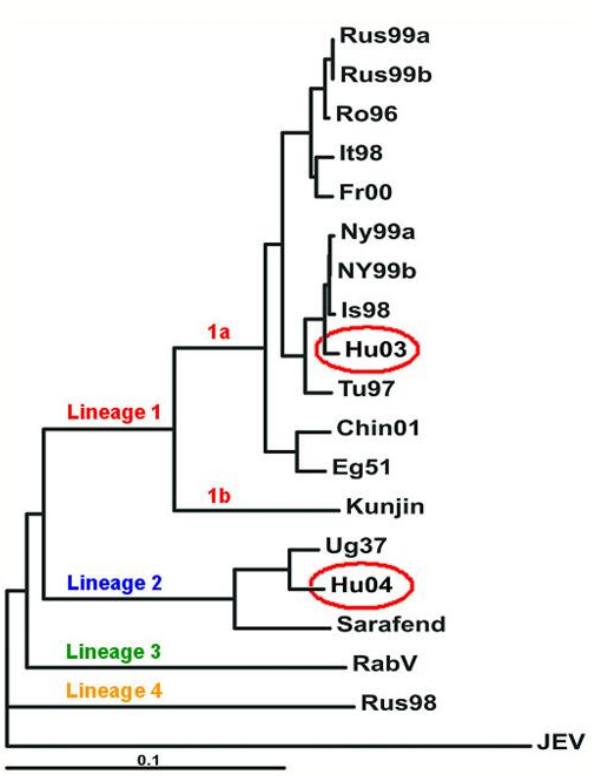


Figure 2 – Overview of West Nile Virus Lineages (National Center for Biotechnology Information, U.S. National Library of Medicine, 2006)

West Nile virus is an enveloped virus with a small virion (an entire virus particle, consisting of an outer protein shell called a capsid, and an inner core of nucleic acid. The core confers infectivity and the capsid provides specificity to the virus. In some virions the capsid is further enveloped by a fatty membrane, in which case the virion can be inactivated by exposure to fat solvents such as ether and chlorophorm) of about 40-50nm in diameter. The round nucleocapsid shows icosahedral symmetry, and is surrounded by a host-derived lipid envelope. The lipid envelope is formed of two major viral proteins: the E (envelope protein), and the M (membrane) protein. The E protein is the major viral determinant (Reid, et al., 2012)

The genome of West Nile virus (WNV) consist of an approximately 11.000 base-long linear positive-sense single RNA strand lacking the polyadenylated tail at the 3' terminus, but containing a methylated nucleotide cap at the 5' end of the genome and several conserved secondary motifs (a sequence pattern of nucleotides in a DNA sequence or amino acids in a protein) regulating viral replication (Reid, et al., 2012).

2.2 Transmission cycle of West Nile virus

The circulation of West Nile virus in Europe is confined to two basic types of cycles; the rural (sylvatic) cycle and urban cycle. The rural (sylvatic) cycle usually consist of wetland birds and ornithophilic mosquitoes (Hubàlek & Halouzka, 1999), and involves the transmission of the virus among wild birds and infected Culex mosquitos(Trevejo & Eidson, 2008), while the urban cycle consist of synantrophic or domestic birds and mosquitoes feeding on both birds and humans (mainly Cx. Pipiens/Molestus) (Hubàlek & Halouzka, 1999).

Bird feeding mosquito species, predominantly of the genus Culex, are the principal vectors of West Nile virus (Hubàlek & Halouzka, 1999), and the virus is maintained in nature in a mosquito-bird-mosquito transmission cycle (Campbell, et al., 2002).The main vector species in Africa and Middle East is Culex univittatus, while in Europe the principal vectors are Culex pipiens, Culex modestus and Coquillettidia richiardii (Hubàlek & Halouzka, 1999). The different mosquito species vary in their vector competence, based on their ability to become infected after feeding on infected hosts, and to further transmit the infection to susceptible hosts. It's also important to note that the host preferences of the individual mosquito species are an important determinant of the transmission dynamics. For example, the Culex species prefer to feed on birds, while Aedes and Ochlerotatus species typically feed on mammals (Trevejo & Eidson, 2008). Virus isolations of West Nile virus (WNV) have occasionally been reported from other hematophagous arthropods, both from bird-feeding argasid (soft) ticks or amblyommine (hard) ticks (Hubàlek & Halouzka, 1999).

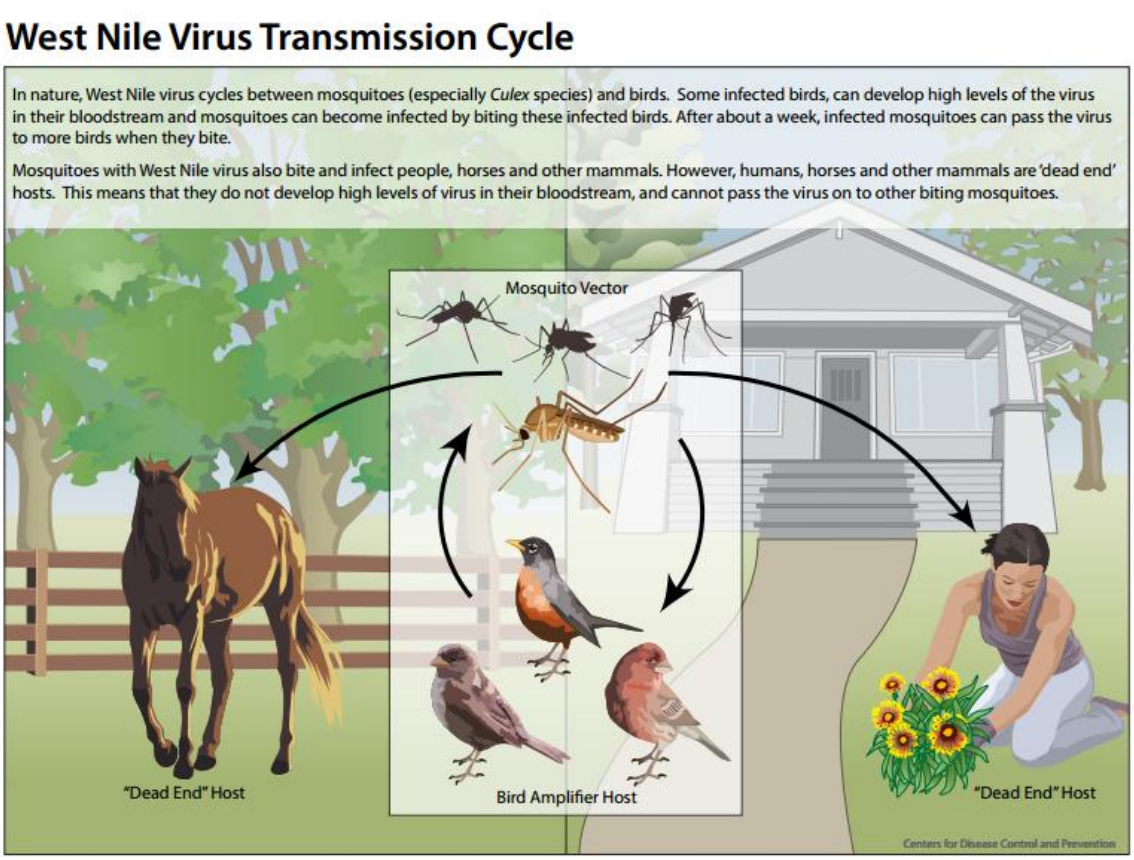


Figure 3 – West Nile Virus Transmission Cycle (Center for Disease Control and Prevention, 2015)

The principal/amplifying hosts of West Nile virus are wild birds, and the virus has been isolated from a number of both wetland and terrestrial avian species. In infected birds with high and long-term viraemia sufficient to infect vector mosquitoes, the virus persists in the organs for 20-100 days. Therefore migratory birds are able to spread the disease over larger geographic distances (Hubàlek & Halouzka, 1999). However, wild bird species vary in their ability to spread the West Nile virus. Both the duration and magnitude of viraemia in affected birds plays a part in their ability to transmit the infection further to the mosquitoes. Through surveillance efforts only a small proportion of West Nile virus positive birds have been detected to be likely competent amplifying hosts for West Nile virus. Under experimental conditions birds like the passerines (e.g. jays, finches, grackles, sparrows and crows) appear to be important reservoirs, and they can act as a source of infection for mosquitoes for a long period of time. The geographic distributions of many of those species are also wide. Other bird species like Psittaciformes, Galliformes and Columbiformes however, appear to be incompetent reservoirs because the West Nile virus associated viraemia in these birds is usually insufficient to provide a source of infection for

mosquitoes (Trevejo & Eidson, 2008). Birds that have become infected usually survive their infections and develop permanent immunity, however other individuals of some species become ill and die (Campbell, et al., 2002).

Most mammals do not appear to maintain the enzootic transmission of West Nile virus, mainly because the viraemia, which develops following the infection, is inadequate for further transmission to the mosquitoes during feeding. These species are therefore usually referred to as dead-end hosts, and they may or may not develop clinical signs of the disease. The most important dead-end hosts, the horse and the humans, are notable for their susceptibility to clinical disease resulting from West Nile virus infection. There is also evidence that the extent of viraemia, which develops in tree squirrels (Sciurus spp), eastern chipmunks (Tamias striatus), and eastern cottontail rabbits (Sylvilagus floridanus) may be a sufficient source of infection for mosquitoes (Trevejo & Eidson, 2008). Locally, only horses and lemurs have shown moderate West Nile Virus viraemia, and they seem to support West Nile virus transmission (Hubàlek & Halouzka, 1999). In Russia, the lake frog also seems to be a competent reservoir (Trevejo & Eidson, 2008).

Other routes of transmission (non-mosquito-borne transmission) have also been reported and hypothesized. In 2002 a transplacental infection was diagnosed in a human infant born with severe brain anomalies. However the results from subsequent surveillance suggest that the transplacental route of infection is extremely rare. The transplacental transmission of West Nile virus has been also proven in experimentally infected mice. There have also been reports of West Nile virus associated disease in humans following blood transfusions, organ transplantation and possibly dialysis (Trevejo & Eidson, 2008). Birds of prey can get infected by consuming WNV infected prey through the oral route. The potential of this infection route was also discussed in the article of “Explosive spread of a neuroinvasive lineage 2 West Nile virus in central Europe” (Bakonyi, et al., 2013), when the event of the long-distance, westward spread of mosquitoes seemed unlikely to be the main driver of the spread of the infection. This hypothesis is strengthened when bird species like the goshawk and the falcons more often get infected and sick from the virus than some other birds, and these birds are also well known pigeon hunters.

2.3 Occurrence of West Nile virus

2.3.1 West Nile virus in Europe

West Nile virus was isolated for the first time in 1937 in Uganda from a febrile human patient, and was detected for the first time in Europe in 1958 in seropositive patients from Albania (Reid, et al., 2012). In 1963 the virus was isolated simultaneously from humans and mosquitoes in France. Soon after the virus was obtained from humans, wild birds, horses, rodents, mosquitoes and ticks in Portugal, Slovakia, Moldova, Ukraine, Hungary, Romania, Czechoslovakia, Italy and other countries (Reid, et al., 2012). Although a severe West Nile virus outbreak was reported in 1996 and 1997 in Romania and seroconversion against WNV and WNV presence has been regularly demonstrated in humans, mosquitoes, migrating birds and rodents in Europe during the last 3 decades, it wasn't until 2003 that WNV infections associated with clinical symptoms were first detected in Hungary (Bakonyi, et al., 2006). The 1996 – 97 – outbreak in Romania which resulted in 500 clinical cases and a fatality rate approaching 10%, was the largest outbreak of arboviral illness in Europe since the Ockelbo-Pogosta-Karelian fever epidemic caused by Sindbis virus in northern Europe in 1980's (Hubàlek & Halouzka, 1999). There was an encephalitis outbreak in a goose flock in Hungary in 2003, which resulted in 14% death rate in 6-week-old geese. WNV was identified as the causative agent based on histopathological examination, serology and nucleic acid detection by reverse transcription-polymerase chain reaction (RT-PCR). Related to this outbreak there was a serologically confirmed WNV outbreak observed in 14 humans with mild encephalitis and meningitis (Bakonyi, et al., 2006). A year later, in August 2004, West Nile virus (WNV) was detected in a goshawk fledgling which showed neurological symptoms and died in a national park in south-eastern Hungary (Bakonyi, et al., 2006). This strain was a new, lineage 2 strain, which subsequently became a resident pathogen in Hungary (Bakonyi, et al., 2013).

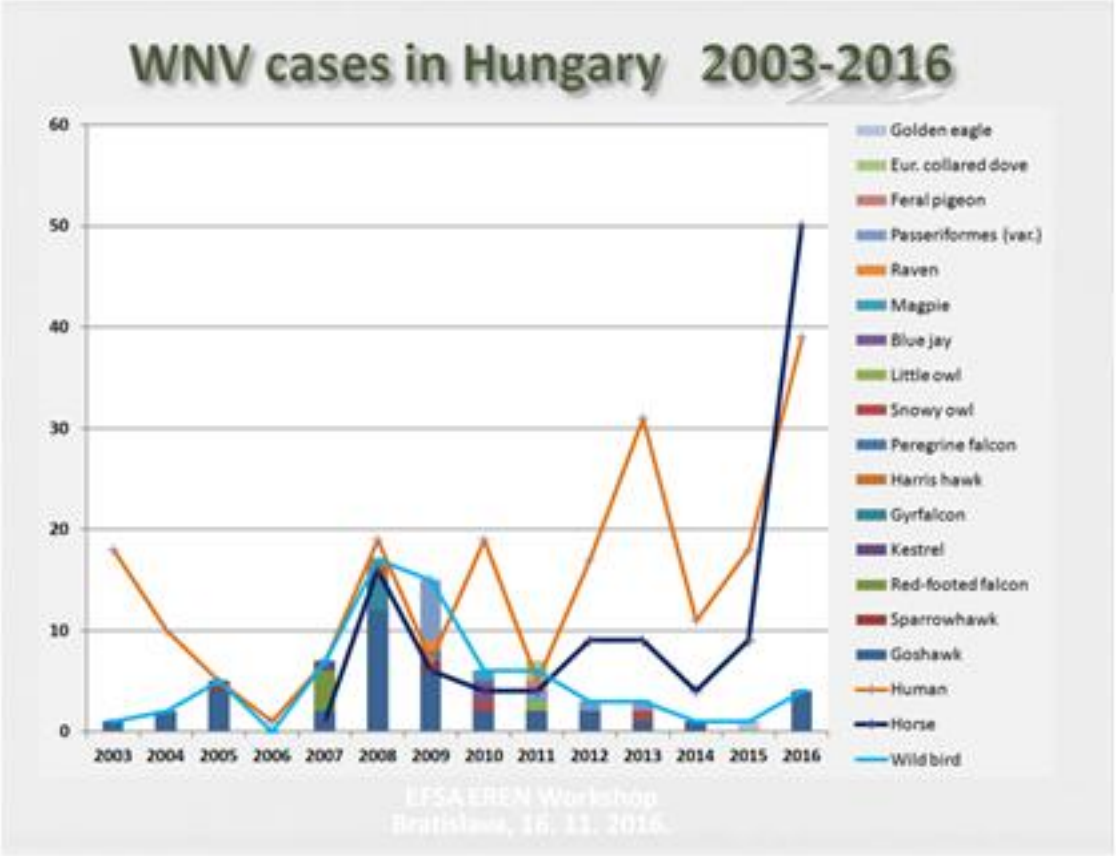


Figure 4 – West Nile Virus cases in Hungary from 2003 to 2016 (Provided by Supervisor Károly Erdélyi)

In the following years, sporadic cases were diagnosed in wild birds, sheep, horses and humans. Infections were further reported in 2008 and 2009 from Italy, Romania, France, Hungary and Austria (Bakonyi, et al., 2013). Based on a phylogeographical reconstruction, it is likely that the European sub-cluster probably started in Hungary and then spread further, first to Austria, before it spread to Italy and Greece (Szentpàli-Gavallèr, et al., 2014).

The European distribution of West Nile virus cases and the spread of the West Nile virus are visualized for the 2010 to 1016 period on the subsequent maps (Figure 5 and Figure 6). The first map showing the distribution of the West Nile virus was made by the ECDC (European Centre for Disease prevention and Control) in 2010, following the explosive expansion of the spread of the virus in 2008.

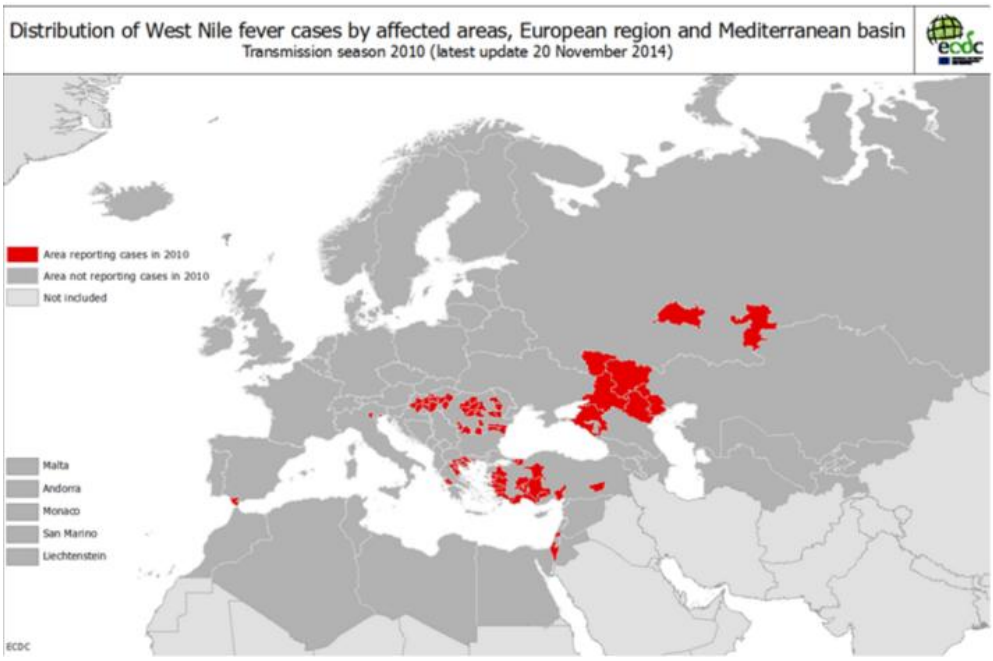


Figure 5 – West Nile Virus Cases in Hungary and beyond in 2010 (European Centre for Disease Prevention and Control, 2016)

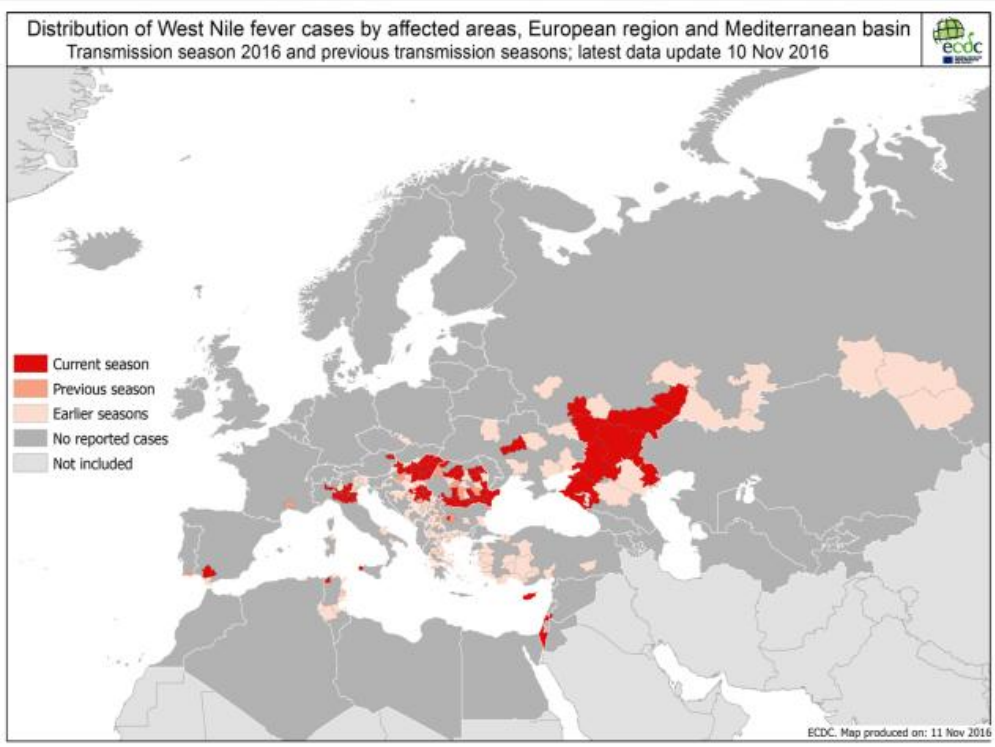


Figure 6 – West Nile Virus Cases in Hungary and beyond in 2016 (European Centre for Disease Prevention and Control, 2016)

2.3.2 West Nile virus outside of Europe

West Nile virus has been detected and reported from several countries outside of Europe, including (among others) Algeria, Asian Russia, Central African Republic, Cyprus, Egypt, India, Israel, Nigeria, Pakistan, and many more (Hubàlek & Halouzka, 1999). West Nile virus was for the first time detected in the United States of America (USA) in 1999 in New York, and it has spread and established throughout the states, causing more than 10.000 cases of severe disease, 400 human fatalities, and thousands of fatal infections in horses (Turell, et al., 2005). The outbreak in 1999 caused 62 cases of human neurological disease, 7 deaths, and thousands of crows, other birds and horses died (Graven & Roehrig, 2001). It is not clear exactly how the West Nile virus was introduced to the USA, but phylogenetic comparison of the nucleic acid sequences showed similarities with strains isolated from encephalitic geese and storks in Israel in 1998 (Bakonyi, et al., 2006). The clinical manifestation of the disease also varies between the different regions, from mild influenza-like symptoms or no apparent disease in the Old World, to increased virulence and clinical symptoms, involving the CNS, and causes death in humans and horses in the New World (Bakonyi, et al., 2006). In 2006 there were 4269 humans reported with WNV infection, of which 65% were cases of the non-neuroinvasive West Nile fever. In 2006 there had been a total of 23.500 human cases reported in the United States where 4% has been fatal (Trevejo & Eidson, 2008).

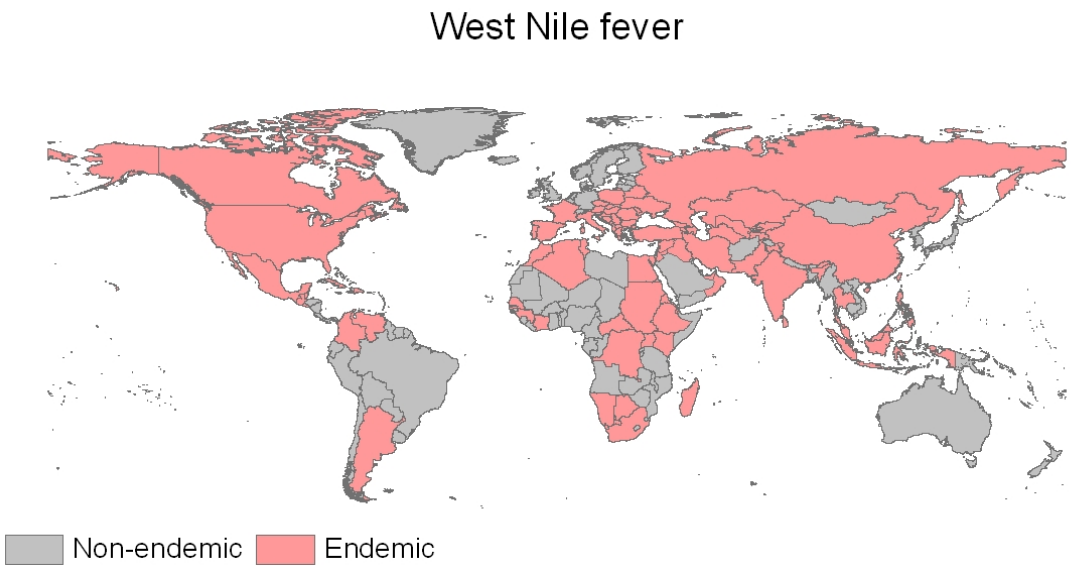


Figure 7 – World map of West Nile Virus distribution (GIDEON Informatics, Inc., 2016)

2.4 Pathogenesis and clinical signs of West Nile virus infection

2.4.1 Pathogenesis of West Nile virus

The exact location and mechanisms of West Nile virus replication following a mosquito bite are unknown, but the initial replication is thought to occur in the skin and regional lymph nodes leading to a primary viraemia, which leads to secondary replication in the reticuloendothelial system (RES). From here there can be a secondary viraemia, which spread the virus further to the central nervous system (CNS). In humans who are clinically healthy, but infected, the virus can be isolated from blood during peak viraemia that occurs from approximately 2 days before until approximately 4 days after illness onset. However, the success of virus isolation decreases significantly after the first day of illness. The most likely reason for this is the increased macrophage clearance and development of IgM (immunoglobulin M) antibody. The clinical manifestations and disease outcome are affected by the viraemia level, which is a result of virus-specific and host-specific factors (Campbell, et al., 2002).

2.4.2 West Nile virus in birds

The clinical features of West Nile virus in birds vary considerably among the different avian species, from unapparent to fatal (Trevejo & Eidson, 2008). It has been reported that different bird species have been detected with a chronic, persistent West Nile virus infection where the viral RNA persistence didn't exceed one year (Bakonyi, et al., 2016). Typical neurological signs of West Nile virus - associated disease include sudden onset of recumbence, mild ataxia, abnormal head posture, circling, swimming in circles, tetraparesis, tremors, nystagmus, seizures and disorientation. In addition to the aforementioned clinical features, signs of depression, mental dullness, anorexia and rapid weight loss, impaired vision, and sudden death may be evident (Trevejo & Eidson, 2008). Pathological findings from infected birds reveal the enlargement, multifocal necrosis and haemorrhage in the liver and spleen. The heart may show evidence of myocardial degeneration, inflammation and pericardial lesions. There are also signs of pancreatitis and chronic inflammation of the adrenal glands. Histopathological findings of neural tissue include brain haemorrhages, degeneration and necrosis of Purkinje cells, and lymphoplasmocytic encephalitis with glial nodules (Trevejo & Eidson, 2008).

2.4.3 West Nile fever in mammals

In horses West Nile virus can cause fever and diffuse encephalomyelitis with a moderate to high fatality rate. It is usually biphasic and cause staggering gait, weakness and paralysis of the hind legs (Hubàlek & Halouzka, 1999). Other clinical features in infected horses include muscle tremors, depression, knuckling over at the metacarpo- or metatarsophalangeal joints, and recumbency (Trevejo & Eidson, 2008). In sheep, which have been inoculated with West Nile virus, symptoms can be fever, abortion and rarely encephalitis. Pigs and dogs are usually asymptomatic, and rabbits, adult albino rats and guinea pigs are resistant to the virus. Adult rodents, which are stressed or immunocompromised, contract fatal encephalitis (Hubàlek & Halouzka, 1999).

2.4.4 West Nile fever in humans

The majority of the Flavivirus species known as human pathogens cause fever, encephalitis and haemorrhagic fever (Reid, et al., 2012). The West Nile virus infection in humans can have either a neuroinvasive or non-neuroinvasive form. Among the patients with the neuroinvasive form it is estimated that approximately 40% have meningitis, while 60% have encephalitis (Trevejo & Eidson, 2008). West Nile fever usually presents as a febrile, influenza-like illness in human (Hubàlek & Halouzka, 1999). The incubation period is short (3-5 days, infrequently biphasic, and sometimes with chills), and the onset is therefore abrupt. Other symptoms can be headache, sore throat, backache, myalgia, arthralgia, fatigue etc. (Hubàlek & Halouzka, 1999). The infection is mainly mosquito-borne, and although the distribution is global, the highest incidence is usually in the tropics (due to the vectors) (Reid, et al., 2012). In Europe however, there is only a small number of mosquito-borne Flaviviruses which have been isolated, and West Nile virus is one of them (Reid, et al., 2012). The severity of the disease varies from asymptomatic infection, to fatal neurological disease, and depends greatly on the virulence of the strain (Reid, et al., 2012). In less than 15% of the clinical cases the infection can cause acute aseptic meningitis or encephalitis, which is associated with neck stiffness, vomiting, confusion, disturbed consciousness, somnolence, tremor of extremities, abnormal reflexes etc. (Hubàlek & Halouzka, 1999). Both lineage 1 and lineage 2 strains may exhibit a neuroinvasive phenotype which determines the neuroinvasiveness and neurovirulence of the virus which is influenced by different genetic markers (Bakonyi, et al., 2013). It has also been reported that in humans, who has been diagnosed with persistent renal infections, the viral RNA could still be detected in the urine of the convalescent patients up to 6 years later (Bakonyi, et al., 2016).

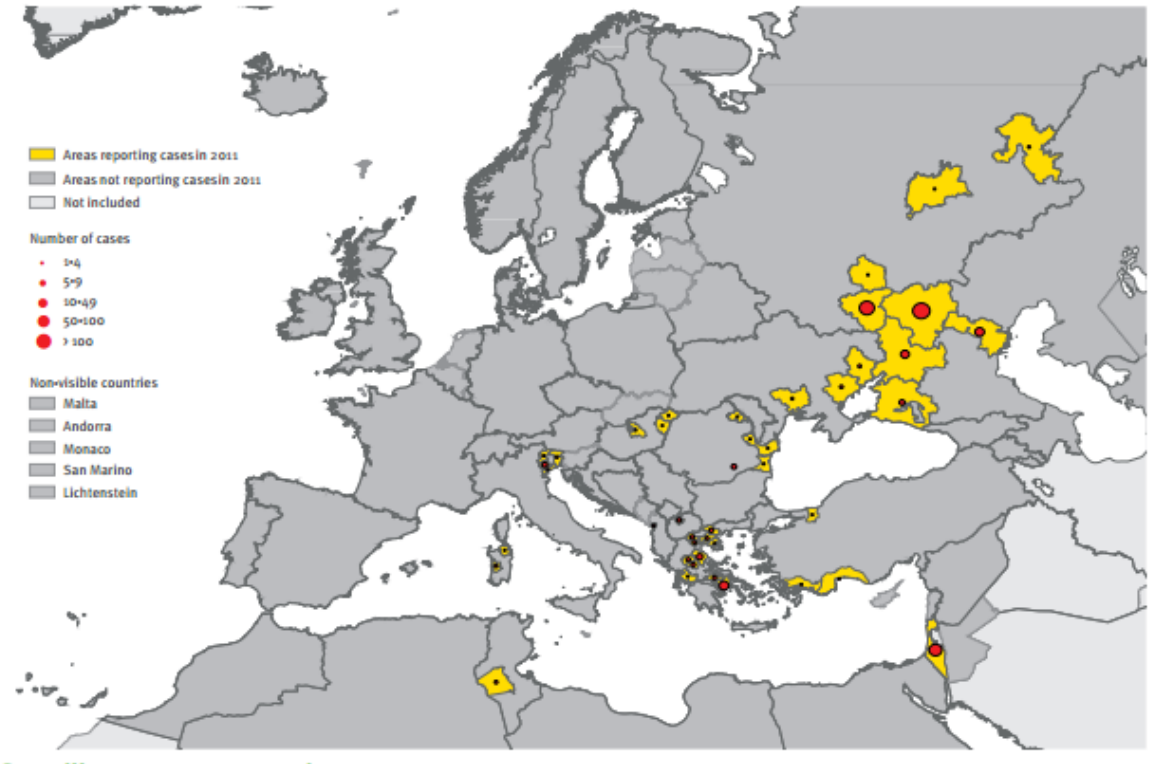


Figure 8 – Human West Nile Virus cases in Europe 2011 (European Centre for Disease Prevention and Control, 2012)

2.5 Diagnosis of West Nile viral disease

In both animals and in humans, there are several potential causes to meningitis and encephalitis including bacterial, viral and non-infectious causes. Based on clinical signs and history, the West Nile virus may be suspected, but only through laboratory methods can the diagnosis be made and other important differential diagnosis be ruled out (Trevejo & Eidson, 2008). The specimens used to test for West Nile virus is cerebrospinal fluid (CSF), serum, blood, tissues, oral and cloacal swabs, and feather pulp. Serological methods are most commonly used for diagnosis of West Nile virus infections in humans and other animals. The West Nile virus-specific immunoglobulin M (IgM) can be detected in the serum 8 days post infection in humans and chickens. Sera can be tested with IgM- and IgG-capture ELISA, and the results are further classified as positive, negative, or equivocal/inconclusive. Through the IgM-capture enzyme-linked immunosorbent assay (ELISA) the anti-WNV antibodies can be detected in the cerebrospinal fluid (CSF) already on day 1 of infection (Trevejo & Eidson, 2008). The virus can also be detected in acute-

phase CSF samples through RT-PCR assay or through viral isolation methods. Different tissues, including the brain, spinal cord, heart, liver and kidneys from dead or euthanized animals can also be submitted for West Nile virus testing. The viral detection methods include cell culture, RT-PCR assay and immunohistochemistry (Trevejo & Eidson, 2008).

2.6 Prevention and control of West Nile virus

The main focus of preventing West Nile virus infection is the environmental control of the mosquito vectors. Such measures include the removal of mosquito breeding sites. Mosquito repellents and avoidance are another line of defence against infection. People can remain indoors during dusk and dawn to avoid the Culex mosquito species' exposure during their peak feeding periods. For horses and dogs topical application of insect repellents, which contains permethrin, pyrethrins, or butoxypolypropylene glycol, can be used. In North America the use of vaccines for horses are highly recommended. There are several vaccines available, including both killed virus or modified-live recombinant virus (Trevejo & Eidson, 2008).

3. Goals and questions

West Nile Virus has been detected in several bird species and in some mammals in Hungary from 2003 till today's date, and based on sequencing results from other countries and regions the virus shows high levels of sequence diversity. In Hungary, however, only three samples have been sequenced so far; the first from 2004, the second from 2010 and the last one from 2014. This study will therefore try to fill in the gap between 2004 and 2010. Two samples have been collected from two goshawks which died of West Nile virus in 2009, and will make a contribution to add to the sequence data of WNV in Hungary. The study will look at where these sequences fit into the phylogenetic context related to the previous complete genomes from 2004. It will also investigate how the viruses relate to the sequences obtained from other geographical areas in the following years.

4. Materials and methods

4.1 Samples

Both of the samples used in the study were isolated from goshawks (Accipiter gentilis) in 2009. The first sample 09/53 was isolated in August 2009 in Szeged, Hungary. This goshawk was found in the area in very weak and in poor body condition. As a try to save the bird, it was submitted to the local zoo, but died shortly after. The other sample 09/67 was isolated in September 2009 from Hortobágy, Hungary. This bird was also found in a bad condition, and was submitted to Hortobágy National park, where it died a few days later. In both birds West Nile virus infection was diagnosed by PCR, histology and immunohistochemistry.

4.2 Reverse Transcription Polymerase Chain reaction (RT-PCR)

Genome segments were amplified directly from the template RNA by a one-step Reverse Transcriptase – Polymerase Chain Reaction (RT-PCR) using a set of 22 primer pairs (Bakonyi, et al., 2016) With this method it is possible to produce multiple copies of a particular DNA or RNA sequence through an amplification process. In RT-PCR the RNA is transcribed into its complement DNA (cDNA) by utilizing reverse transcriptase enzyme (Differencebetween.net, 2011). A set of primers were added to the mix, which is needed for the polymerase enzyme to add the first nucleotide, which further contributes to the synthesis of new RNA strands complementary to the offered template. From there on it is possible to delineate a specific region of the desired template sequence for further amplification. The primers used are listed in the Table 1 below.

Primer pairs	Primer	Size
1	WNVII 1f	250
	WNVII 200r	250
2	WNVII 81f	900
	WNVII 954r	900
3	WNVII 870f	800
	WNVII 1630r	800
4	WNVII 1584f	950
	WNVII 2491r	950
5	WNVII 2397f	900
	WNVII 3189r	900
6	WNVII 3027f	700
	WNVII 3699r	700
7	Usu 3606f	1200
	Usu 4759r	1200
8	WNVII 3506f	500
	WNVII 3971r	500
9	WNVII 3861f	600
	WNVII 4454r	600
10	Usu 4251f	1200
	Usu 5392r	1200
11	WNVII 4378f	1000
	WNVII 5378r	1000

Primer pairs	Primer	Size
12	Usu 5454f	400
	Usu 5790r	400
13	WNVII 5629f	700
	WNVII 6245r	700
14	WNVII 6127f	1000
	WNVII 7058r	1000
15	WNVII 6936f	700
	WNVII 7606r	700
16	WNVII 7505f	700
	WNVII 8158r	700
17	WNV 8078f	800
	WNV 8821r	800
18	WNV 8723f	800
	WNV 9461r	800
19	WNV 9368f	800
	WNVII 10137r	800
20	WNVII 10023f	600
	WNVII 10541r	600
21	Usu 10596f	500
	Usu 11014r	500
22	WNV 3f	800
	WNV 2r	800

Table 1 – Primer pairs

This table has been copied from the “Chronic West Nile virus infection in Kea” paper (Bakonyi, et al., 2016).

Mix:	Ad 25µl/reaction	1x(µl)	(n+1) x (µl)
	Number of samples (n)		25
	H_2O	14.25	370.5
	5xBuffer (Qiagen, 25mM $MgCl_2$)	5	130
	dNTPs (10nM):	1	26
	RNasin (Promega):	0.25	6.5
	Enzyme mix (Qiagen):	1	26
	Primer F+R mix (40 pmol/µl):	1	1-1
	Template RNA	2.5	65

Table 2 – PCR mix

The PCR mix containing individual primer pairs, nucleotides, and the RNA template, was assembled from the ingredients by the automated system QI agility (QIAGEN, u.d.). The template used is the sample RNA which contains the target sequence. The QI Agility robot could fit 16 samples at a time to mix all the ingredients, so the process was divided in two, with first 16 tubes, then the last 6 after. After mixing the ingredients with sample RNA and primers, the samples were ready to be placed in the thermal cycler.

The thermal cycler used to perform the PCR was an Applied Biosystems Rotor Gene Thermocycler W. The eppendorf tubes containing the RNA, primers and the other PCR ingredients were placed into the thermal cycler and the cycler was programmed to perform the RT-PCR. The samples were first heated to 50^{o}C for 40 minutes and thereafter at 95^{o}C for 15 minutes. A cyclic program was set at a 94^{o}C denaturation for 30 seconds, which separates the strands from each other, followed by annealing at 57^{o}C for 50 seconds and an elongation phase at 72^{o}C for 1.5 minute. The annealing phase consists of primer binding, and is where the polymerase enzyme synthesizes new strands of RNA complementary to the target sequence. From here the elongation process synthesizes new RNA from the end of the primer using deoxynucleotide triphosphates (dNTPs) as building blocks for further amplification of the RNA. This cycle was repeated 45 times, before the final elongation step at 72^{o}C for 7 minutes. The amplification of the RNA was now complete and could be used for further investigation.

4.3 Gel electrophoresis

Based on size and charge, gel electrophoresis is a method to separate mixtures of DNA, RNA or proteins. The gel used was 1.5% agarose gel to which the PCR samples would be added. Into 130 ml TAE buffer solution (a mixture of Tris base, acetic acid and EDTA) 1.95g of agarose powder was added. The solution was micro waved until boiling, and then cooled to 50-60^{o}C. After cooling the solution to appropriate temperature, 13µl stain was added. Earlier etidium bromide stain was used, but this compound has been proven to be carcinogenic, so the Midori Green advance stain was used instead. The gel plate used was 20x10cm (100ml) and contained in total 30 wells. The solution was poured carefully into the gel plate. The pouring was performed with caution to avoid the formation of air bubbles in the gel. The consequence of this can be a non-valid result. After waiting for 20-30 minutes for the solution to settle and become firm, a loading stain, Gene Ruler^{TM} 100 bp DNA ladder Plus (MBI Fermentas) (ThermoFisher Scientific, u.d.), was added to the sample before placing it into the wells. The loading stain helps the sample to stay in the well, and not spread through the gel. A marker or lead stain was applied between every 10th sample, so it was placed between samples 10 and 11, and between samples 20 and 21. The electric current which is applied across the gel was turned on, and the molecules based on their size and charge started to migrate. The RNA molecules are negatively charged, and will travel towards the positively charged electrode. The shorter strands of the RNA travel faster and further through the gel than the larger strands, and from there fragments arranged in order of size are visible. The result is a series of bands, where each band contains RNA molecules of a particular size.

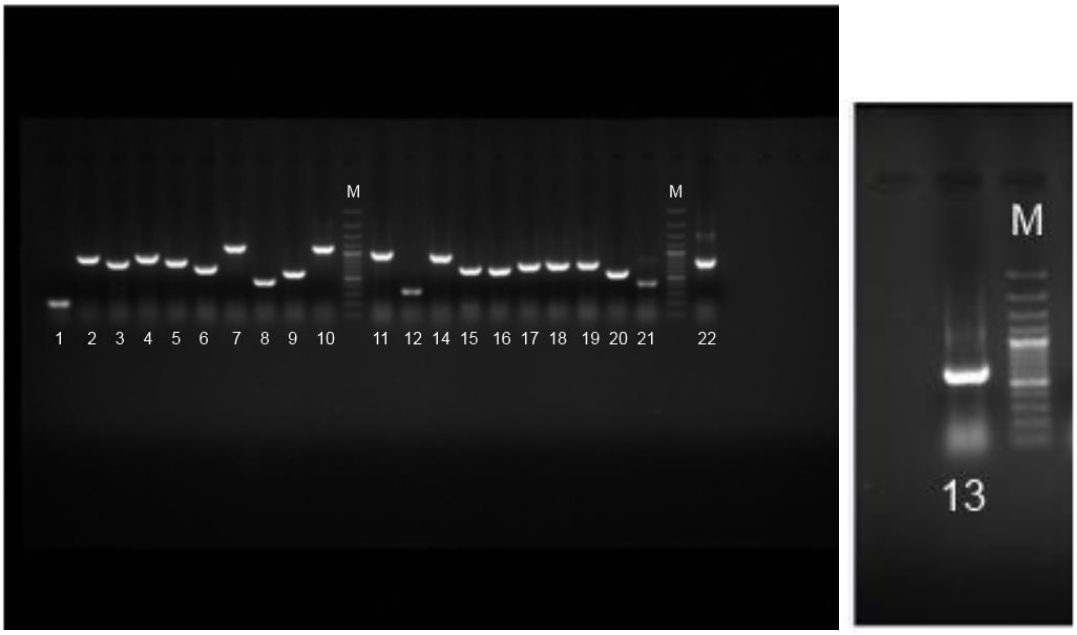


Figure 9 – Gel electrophoresis results

The figure above illustrates the distance the different amplified PCR products have travelled. The numbers (1-22) correlates with the primer pair numbers, which can be seen in the table in Table 1. The 'M' represents the marker stain wells, which was applied between every 10th well.

4.4 Quantum Prep PCR Kleen Spin Columns

In order to send the amplified PCR products to a laboratory for sequencing it is necessary to purify/extract RNA from the PCR samples. Through the Quantum Prep PCR Kleen Spin Columns method the PCR products are separated from smaller impurities, which include unincorporated dNTPs, polymerases, primers, and small primer-dimers etc. (Bio-Rad Laboratories, 2000). The method consists of a column and a collection tube. The column contains resin where all by-products from the PCR will be gathered, and the collection tube is where the RNA product will be collected in the end. The resin in the column was resuspended by vortexing it for 5 seconds. Further on the cap was removed and the tip snapped off, before the column was placed in a 2.0ml tube. The column was thereafter pre-spun for 1 minute at 735G before being placed in a clean 1.5ml collection tube. From there the sample was placed in the centre and on top of the resin layer. The application was done with caution in order to not disturb the resin layer. The column was then placed in a centrifuge and spun for 2 minutes at 735G. After centrifuging the column the purified

RNA could be collected at the bottom of the 1,5ml collection tube. The unincorporated dNTP's, primers and short primer-dimers remained in the column.

4.5 Sequencing and phylogeny

After purification and RNA extraction, the samples were sent to a commercial lab in Netherlands (Macrogen Europe) for direct sequencing with primers used in the PCR reactions. The chromatogram files (obtained from the sequencing of individual PCR products) received from the lab were evaluated and assembled in SeqMan software (Lasergene software package). The identity of the partial and the assembled complete genomic sequences was checked by a BLAST search (Basic Local Alignement Search Tool) of the GenBank at the NCBI (National Center for Biotechnology Information). In order to compare our sequences with other West Nile virus genomes we have retrieved a set of 43 lineage 2 WNV sequences from GenBank for the phylogenetic analysis (Appendix A). The editing and alignment of the sequences was performed by the ClustalW method in BioEdit software (Hall, 1999), where the alignment was reduced to the 10305 bp (base pairs) long nucleotide sequence of the polyprotein coding region, and used in the subsequent analyses. The most suitable nucleotide substitution model applying to the alignment was determined using the tool provided in the Mega7 software (Kumar, et al., 2016). Phylogenetic trees were constructed using three methods: Neighbour Joining (NJ), Maximum Likelihood (ML), and the Bayesian method. The Neighbour Joining phylogeny was produced in Mega7 software, while the Maximum Likelihood and Bayesian analysis were obtained with the help of TOPALi v2.5 software (Milne, et al., 2009), where the RAxML and MrBayes analysis was used. Evolutionary distances of the analysed genomes were also computed in MEGA7. All free model parameters were estimated by RAxML: GAMMA model of rate heterogeneity, ML estimate of alpha-parameter, 4 discrete rate categories. GTR+GAMMA Model parameters were estimated up to an accuracy of 0.1000000000 Log Likelihood units. The Bayesian analysis was run in MrBayes environment with the following parameters: mcmc nruns = 2 ngen = 100000, sample frequency = 10; sump burning = 2500; sumt burnin = 2500.

5. Results and Discussion

5.1 Results

The results from the PCR were successful for all 22 primer products for both samples. As an example, the PCR products obtained from the WNV0953 sample visualised through gel electrophoresis can be seen in Figure 9. Primer pair number 13, which had been misplaced in the thermocycler, ran alone after the rest of the primer pairs had gone through the gel. From the Figure 9 above it is possible to see that the size of the PCR products correlates with the distance they have travelled. The marker (M) stain, which was applied between every 10^{th} samples, makes it possible to read how many base pairs the primers consist of (molecular weight). Each line represent 100 bp, and the stronger lines, seen in the marker stain line, represents 500 and 1000 base pairs. This result was compared with the length of the predicted product sequences listed in Appendix A, and showed high correlation, which makes the results from the PCR reliable.

The results from the Macrogen Europe laboratory in The Netherlands showed that the sequencing was successful. See Appendix A for the sequences which were obtained from the laboratory. The two sequences were assembled in the Seqman software, which is illustrated in the Figure 10. The partial constructs and the completely assembled WNV sequences were submitted to a BLAST search in GenBank at the NCBI website, and both sequences of the samples were recognized as lineage 2 West Nile virus by comparison to the other WNV genomes deposited in GenBank. The local alignment of the different lineage 2 strains was performed in BioEdit (ClustalW). The sequences of both samples are presented in Appendix A. The first was designated “WNV 0953 – Goshawk – 2009 – HU”, and the second one “WNV 0967 – Goshawk – 2009 – HU”.

We have retrieved the other, already registered, lineage 2 West Nile Viruses from the NCBI GenBank to perform the further comparison analysis in Appendix A. There was 100% coverage by at least two sequences on the whole length, with the exception for a single region (product of PCR No4.). The Seqman coverage is illustrated in the Figure 10 below.

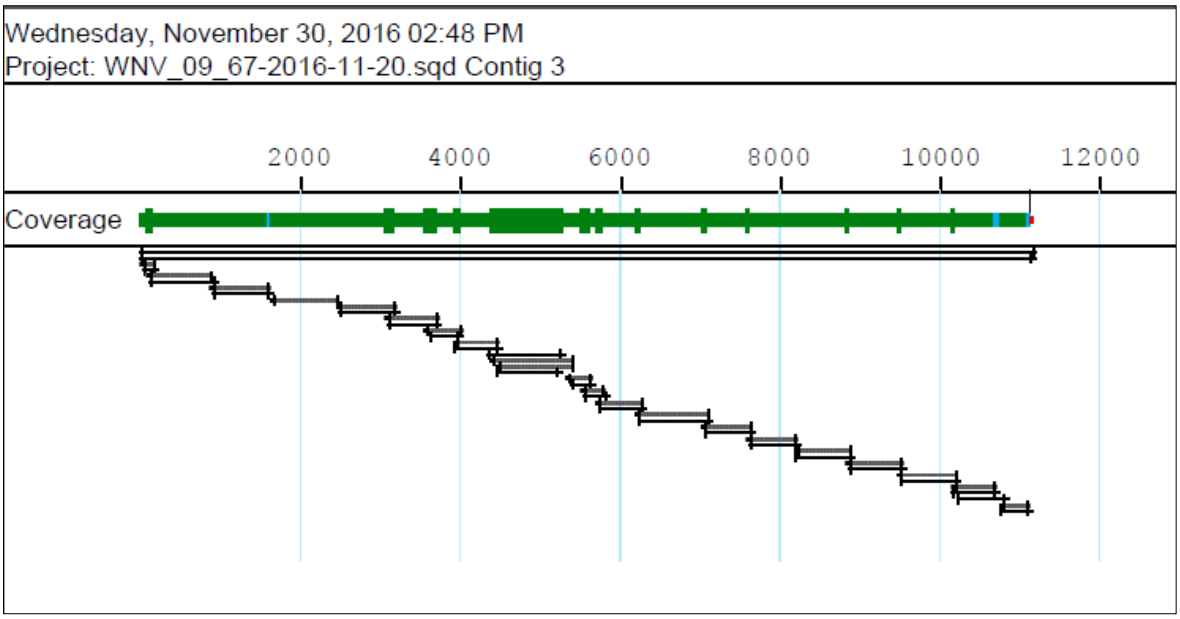


Figure 10 – Seqman coverage

Nucleotide substitution models were tested for the subsequent phylogenetic analysis on the alignment of 45 sequences in total, and the best fitting models were chosen in the following order: 1. TN93+G, 2. TN93+G+I, 3. GTR+G. These nucleotide substitution models are showed in Appendix B. All the phylogenetic tree models were analysed by using the Tamura-Nei model, and conducted with standard error estimated through 500 bootstrap replicates, and the evolutionary distances were computed. The Bayesian model was chosen as the best fitting model for the interpretation the results for the two sequences, and showed reliable bootstrap numbers. The Bayesian tree can be seen in the subsequent Figure 11. The other two trees (Neighbour Joining tree and Maximum Likelihood tree) are available in Appendix and Appendix D.

All three phylogenetic methods (Neighbour Joining (JN), Maximum Likelihood (ML) and the Bayesian method (MB)) produced virtually the same topology in the resulting phylogenetic trees. The analyses showed that the both investigated genomic sequences clustered together with other Hungarian isolates from the respective time period. The regional subclustering of West Nile virus isolates is also evident from our analysis.

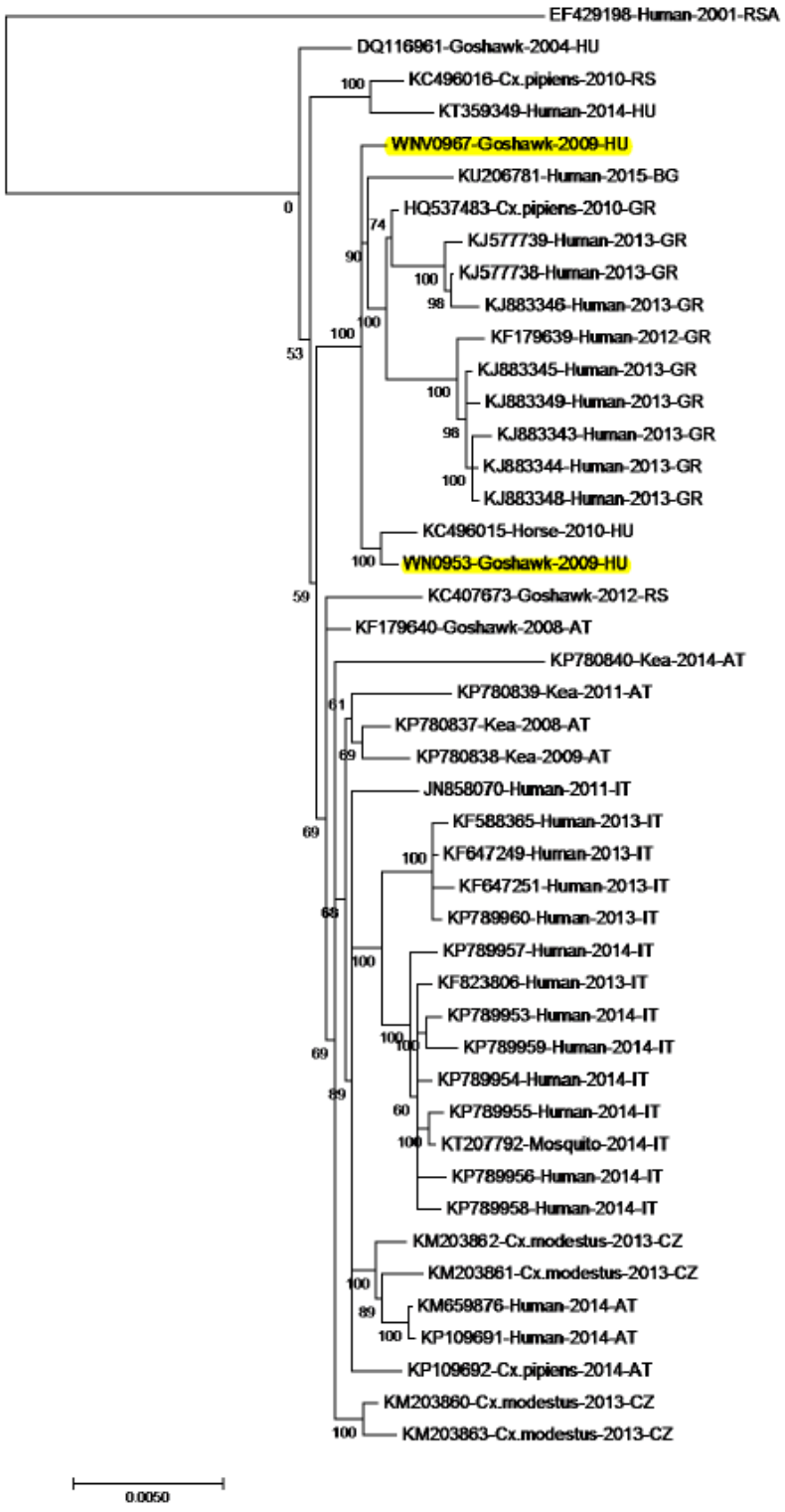


Figure 11 – Bayesian analysis phylogram demonstrating the genetic relatedness of the viruses

Virus codes are composed of GenBank accession numbers and details of detection source; country code (HU=Hungary, AT = Austria, CZ = Czech Republic, IT = Italy, RS = Serbia, BG = Bulgaria, GR = Greece and RSA = South Africa), hosts that the virus was isolated from and the year they were isolated. The samples, which were sequenced in this study, are highlighted in yellow.

5.2 Discussion

By looking at the sequences investigated in the Bayesian tree above, there are a lot of conclusions and assumptions to be drawn from it. By looking at the investigated sample sequences, both of the sequences share a common ancestor with the 'KU206781' WNV strain discovered in Bulgaria, and the Greek strains: 'HQ537483', 'KJ 577739', 'KJ577738', 'KJ883346', 'KF179639', 'KJ883345', 'KJ883349', 'KJ883343', 'KJ883344' and 'KJ883348'. The sample sequence 'WNV0953' forms a distinct cluster with the 2010 isolate 'KC496050' derived from a horse which evolved in a different direction than the other above mentioned strains detected in Bulgaria and Greece. The 'KC496015' was isolated from a horse in Hungary in 2010, so neither of these two strains, nor their descendants, has been detected outside of Hungary yet. All of the other relatives of the 'WNV0967' and 'WNV0953' strains have expanded from Hungary to Bulgaria, Greece and perhaps further, but still not detected or sequenced. The latest strain of a related virus was isolated from a human in Bulgaria in 2015.

All the lineage 2 viruses listed in Figure 11 originate from the same ancestor. The 2004 'DQ116961' goshawk strain was the first detected lineage 2 strain in Europe, and has given rise to the epidemical spread of the virus in Europe ever since. From 2008 there was an explosive spread of the West Nile virus, and the reason for this is yet unknown, but it is believed that vertebrate hosts are the main cause for the dissemination of the virus (Bakonyi, et al., 2013). The goshawks have proven to be quite susceptible to the virus, and many have succumbed during the last decade, however, studies show that pigeons seem to be suitable reservoirs for the virus (Bakonyi, et al., 2013), and it has been emphasized that they play an important part in the spread of the virus as well. The outbreaks usually follow a seasonal pattern, and according to the article of the "Explosive spread of a neuroinvasive Lineage 2 West Nile viruses in Central Europe" ((Bakonyi, et al., 2013)), the first cases usually were discovered in July and August, and the last cases detected in late October, which gives an indication that the mosquito vectors also play an important part in the spread.

We should note that the genetic material of the virus changes over time. We should also take in to consideration the role and contribution the individual species and permanently infected hosts play in the diversification of the virus strains.

For each branching there is a change in the viral sequence genome which have caused the viruses to change. From the tree, it is possible to see how the different strains relate to each other and how they have evolved. Though the Greek strains and the Italian strains both share a common ancestor with the Hungarian strain ('DQ116961'), they seem to have evolved separately, and in the end became distinctly different from each other. The outbreaks in Italy and Greece happened around the same time in 2013-2014. From this we can make an assumption, based on the branching in the phylogenetic tree, that the viruses have evolved separately over time, and not been spread from one of the countries to the other, or from a third country to both Italy and Greece. If this had been the case, then the sequences collected from these countries, in this time period, would have been more closely related to each other than they are.

For the future it will be interesting to investigate to which extent the individual species, and permanently infected hosts, contribute to the diversification of the West Nile virus. From Figure 11 there are two virus strains collected from keas (Nestor notabilis) in Austria. Number one ('KP780839) was isolated in 2011, while the other ('KP 780840') was isolated in 2014. The level of genetic distance of both strains, but especially the distinctness of the 2014 strain, can indicate that hosts which are permanently infected contribute for the virus to change and evolve drastically. According to a study on chronically infected kea in Austria (Bakonyi, et al., 2016), the increase in genetic distance was proportional to the length of the chronic period of disease in the different birds investigated. All of the birds were infected simultaneously, but the genetic divergence of the virus of the bird ('KP780839'), which died in 2011, is a lot higher than the virus isolated from bird ('KP780837'), who died in 2008.

It will also be interesting for future research to investigate both the direct and indirect contribution of migratory bird dispersal to the spread of West Nile Virus. The possibility of indirect spread of West Nile Virus from pigeons to birds of prey has been hypothesized by (Bakonyi, et al., 2013). Pigeons seem like likely reservoir hosts, which may travel far distances, and are a frequent prey for predatory birds like the goshawk and the falcons. The migratory birds also play an important role in the spread through countries and continents. The infected birds can infect mosquito species in the area, which will spread the virus further to other hosts. Further analyses and additional WNV genomic sequences may be necessary for the clarification of the above issues.

6. Summary (abstract)

The lineage 2 strain of the West Nile virus was detected for the first time, outside of the sub-Saharan Africa, in a goshawk (Accipiter gentilis) in Hungary. From there the virus has spread through Europe through reservoir hosts (migratory birds) and vectors (mosquitoes). West Nile virus has neuroinvasive properties and can be fatal for several host species, including humans. Several mammals and birds have after the first outbreak in 2004 been infected, and the spread of the virus exploded in 2008. West Nile virus has from then on been detected in several different species; mainly birds, human and horses, but other species can also be affected. There was missing sequencing data of the evolution and the diversity of the viral strains, and in order to understand and follow the virus further as it spread, it was necessary to fill in the gaps in the time period for the expansion of the Lineage 2 strains of the virus in Europe. By sequencing the virus, relationships can be discerned between the strains, and the evolution and the spread of the virus can be mapped. For this reason, it was important to sequence two viral genomes, obtained from two goshawks, which died of the virus in 2009. This was solved through several methods like PCR, gel electrophoresis, sequencing and in the end the phylogenetic analysis. The phylogenetic methods used were the Neighbour-Joining (NJ), Maximum Likelihood (ML) and Bayesian model (MB), and with the help of them it was possible to draw some conclusions of the relationship between the strains and their ancestors. These methods also help in interpreting how the ancestral strains have affected later generations of the virus and where it has spread to. Based on this study, it is reasonable to believe that the ancestral pool containing the investigated viruses gave rise to the spread of the West Nile virus to Greece and Italy, where it caused devastating neuroinvasive West Nile fever in humans the following years. These two sequences will be important for future studies of this virus, both to explain the viral diversity and evolution, and for research of further spread of the virus.

7. Bibliography

GIDEON Informatics, Inc., 2016. *West Nile fever.* [Online]
Available at: http://web.gideononline.com/web/epidemiology
[Accessed 11 2016].

Bakonyi, T. et al., 2013. Explosive spread of a neuroinvasive lineage 2 West Nile virus in Central Europe. In: *Veterinary Microbiology 165.* s.l.:s.n., pp. 61-70.

Bakonyi, T. et al., 2016. Chronic West Nile virus infection in kea (Nestor notabilis). *Veterinary Microbiology 183*, pp. 135-139.

Bakonyi, T. et al., 2006. Lineage 1 and 2 strains of Encephalitic West Nile Virus, Central Europe. *Emerging Infectious Diseases, Vol. 12, No. 4*, April, pp. 618-623.

Beck, C. et al., 2013. Flaviviruses in Europe: Complex Circulation Patterns and Their Consequences for the Diagnosis and Control of West Nile Disease. *International Journal of Environmental Research and Public Health — Open Access Journal*, 11, pp. 6049-6083.

Bio-Rad Laboratories, 2000. *Quantum Prep: PCR Kleen Spin Columns.* Hecules, CA: s.n.

Campbell, G. L., Marfin, A. A., Lanciotti, R. S. & Gubler, D. J., 2002. West Nile Virus. *The Lancet Infectious Diseases Vol. 2*, 9, pp. 519-529.

Center for Disease Control and Prevention, 2015. *West Nile Virus.* [Online]
Available at: http://www.cdc.gov/westnile/transmission/
[Accessed 11 2016].

Differencebetween.net, 2011. *Difference Between RT-PCR and QPCR.* [Online]
Available at: http://www.differencebetween.net/science/difference-between-rt-pcr-and-qpcr/
[Accessed 11 2016].

European Centre for Disease Prevention and Control, 2012. *Annual epidemiological report 2012.* [Online]
Available at: http://ecdc.europa.eu/en/publications/publications/annual-epidemiological-report-2012.pdf
[Accessed 11 2016].

European Centre for Disease Prevention and Control, 2016. *West Nile fever maps.* [Online]
Available at: http://ecdc.europa.eu/en/healthtopics/west_nile_fever/west-nile-fever-maps/pages/index.aspx
[Accessed 11 2016].

Graven, R. B. & Roehrig, J. T., 2001. West Nile Virus. *JAMA*, 8 8, pp. 651-653.

Hall, T., 1999. *BioEdit: a user-friendly biological sequence alignment editor and analysis, Nucl. Acids. Symp. Ser. 41:95-98..* s.l.:s.n.

Hubàlek, Z. & Halouzka, J., 1999. West Nile Fever - a Reemerging Mosquito-Borne Viral Disease in Europe. *Emerging Infectious Diseases Vol. 5*, 9-10, pp. 643-650.

Kumar, S., Stecher, G. & Tamura, K., 2016. *MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.* s.l.:Department of Biological Sciences, Tokyo Metropolitan University.

Milne, I. et al., 2009. *TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. Bioinformatics.* s.l.:s.n.

National Center for Biotechnology Information, U.S. National Library of Medicine, 2006. *Lineage 1 and 2 Strains of Encephalitic West Nile Virus, Central Europe.* [Online]
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3294705/
[Accessed 11 2016].

QIAGEN, n.d. *QIAGEN.* [Online]
Available at: https://www.qiagen.com/us/

Reid, H. W., Weissenböck, H. & Erdèlyi, K., 2012. Flavivirus infections. *Infectious Diseases of Wild Mammals and Birds in Europe*, pp. 128-145.

Szentpàli-Gavallèr, K. et al., 2014. Monitoring of West Nile Virus in Mosquitoes Between 2011-2012 in Hungary. *Vector-Borne And Zoonotic Diseases, Vol. 14*, pp. 648-655.

ThermoFisher Scientific, n.d. *Thermo Scientific.* [Online]
Available at: https://www.thermofisher.com/no/en/home/brands/thermo-scientific.html#/legacy=www.fermentas.com
[Accessed 2016].

Trevejo, R. T. & Eidson, M., 2008. Zoonosis Update: West Nile Virus. *Vet Med today: Zoonosis Update*, 1 5, pp. 1302-1309.

Turell, M. J. et al., 2005. An update on the potential of North American mosquitoes (Diptera: Culicidae) to transmit West Nile Virus. *Journal of Medical Entomology, Vol 42, no 1*, pp. 57-62.

8. Acknowledgements

I would like to place a special thanks to my supervisor Dr. Erdélyi, Károly for giving me the opportunity to learn and write about a subject, which is getting more and more important for each year, and for being patient when guiding me through the ocean of articles and literature. Thank you to Dr. Dán, Ádám who kindly showed and explained the PCR and Gel electrophoresis procedures. A big thank you to my brother, Arne Fivelstad for correcting my English. And at last I would like to thank my boyfriend, Jan Hovden Eide, for the best help in editing and formatting my thesis to a great standard.

9. Appendices

A. The West Nile Virus sequences retrieved of the two samples from the lab
B. Maximum Likelihood fits of 24 different nucleotide substitution models
C. Neighbour joining tree
D. Maximum likelihood

Appendix A

The West Nile Virus sequences retrieved of the two samples from the laboratory

Sample 1: WNV 0953 – Goshawk – 2009 - HU

[ATGTCTAAGAAACCAGGAGGGCCCGGTAAAAACCGGGCTGTCAATATGCTAAAACGCGGTATGCCCCGCGGATTGTCCT
TGATAGGACTGAAGAGGGCTATGCTGAGTCTGATTGACGGGAAAGGCCCAATACGTTTCGTGTTGGCTCTTTTGGCGTTTT
TCAGATTCACTGCAATCGCTCCGACTCGTGCGGTGCTGGACAGATGGAGAGGCGTCAACAAACAAACAGCGATGAAGCA
TCTCTTGAGTTTCAAGAAAGAACTAGGAACTCTGACCAGTGCCATCAACCGCCGGAGCACAAAACAAAAGAAAAGAGGA
GGCACAGCGGGCTTTACTATCTTGCTTGGGCTGATCGCTTGTGCTGGAGCTGTGACCCTCTCGAACTTCCAGGGCAAAGTG
ATGATGACAGTCAATGCAACCGATGTTACTGACGTGATTACCATCCCAACAGCTGCTGGGAAAAACCTGTGCATTGTGAG
GGCTATGGACGTGGGATACCTTTGTGAGGATACTATCACTTATGAATGTCCGGTCCTAGCTGCTGGAAATGACCCTGAAG
ACATTGACTGCTGGTGCACGAAATCATCTGTTTACGTGCGCTATGGAAGATGCACAAAAACTCGGCATTCCCGTCGAAGC
AGAAGGTCTCTGACTGTCCAGACACATGGAGAAAGTACACTGGCCAACAAGAAAGGAGCTTGGTTGGACAGCACAAAAG
CCACGAGATATCTGGTGAAGACAGAATCATGGATACTGAGAAACCCGGGCTACGCCCTCGTTGCAGCGGTCATTGGATG
GATGCTAGGAAGCAACACAATGCAACGCGTCGTGTTTGCCATCCTATTGCTCCTGGTGGCACCAGCATACAGCTTCAATT
GCTTGGGAATGAGTAACAGAGACTTCCTGGAGGGAGTGTCTGGAGCTACATGGGTTGATCTGGTACTGGAAGGCGATAG
TTGTGTGACCATAATGTCAAAGGACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCCAACCTCGCAGAT
GTACGCAGCTACTGTTACCTAGCTTCGGTCAGCGACTTGTCAACGAGAGCCGCGTGTCCAACCATGGGTGAAGCCCACAA
CGAGAAAAGAGCTGACCCCGCCTTCGTTTGCAAGCAAGGCGTTGTGGACAGAGGATGGGGAAATGGCTGCGGACTGTTT
GGAAAGGGAAGCATTGACACATGCGCGAAGTTTGCCTGTACAACCAAAGCAACTGGATGGATCATCCAGAAGGAGAACA
TCAAGTATGAGGTCGCCATATTTGTGCATGGCCCGACGACTGTTGAATCTCATGGCAATTATTCAACACAGATAGGGGCC
ACCCAGGCTGGAAGATTCAGTATAACTCCATCGGCGCCATCTTACACGCTAAAGTTGGGTGAGTATGGTGAGGTCACGGT
TGACTGTGAGCCACGGTCAGGAATAGACACTAGCGCCTATTACGTTATGTCAGTTGGTGCGAAGTCCTTCCTGGTTCACC
GAGAATGGTTTATGGATCTGAACCTGCCATGGAGCAGTGCTGGAAGCACCACGTGGAGGAATCGGGAAACACTGATGGA
GTTTGAAGAACCTCATGCCACCAAACAATCTGTTGTGGCTCTAGGGTCGCAGGAAGGTGCATTGCACCAAGCTCTGGCCG
GAGCGATTCCTGTTGAGTTCTCAAGCAATACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAA
GTTGCAGCTGAAGGGAACAACATATGGAGTATGTTCAAAAGCGTTTAAATTCGCTGGGACTCCCGCTGACACTGGCCATG
GAACGGTGGTGTTGGAACTGCAATACACCGGAACAGACGGTCCCTGCAAAGTGCCCATTTCTTCCGTAGCTTCCCTGAAT
GACCTCACACCTGTTGGAAGACTGGTGACCGTGAATCCATTTGTGTCTGTGGCCACAGCCAACTCGAAGGTCTTGATTGA
ACTCGAACCCCCGTTTGGTGACTCTTACATCGTGGTGGGAAGAGGAGAACAGCAGATAAACCACCACTGGCACAAATCT
GGGAGCAGCATCGGAAAGGCCTTTACCACTACACTCAGAGGAGCTCAACGACTCGCAGCTCTTGGAGACACTGCTTGGG
ATTTTGGGTCAGTTGGAGGGGTTTTCACCTCAGTGGGGAAAGCCATACACCAAGTCTTTGGAGGAGCTTTTAGATCACTCT
TTGGAGGGATGTCCTGGATCACACAGGGACTTCTAGGAGCCCTTCTGTTGTGGATGGGAATCAATGCCCGTGACAGGTCA
ATTGCCATGACGTTTCTTGCGGTTGGAGGAGTTTTGCTCTTCCTTTCGGTCAACGTCCATGCTGACACAGGCTGTGCCATT
GATATTGGCAGGCAAGAGCTCCGGTGTGGAAGTGGAGTGTTTATCCACAACGATGTGGAAGCCTGGATGGATCGTTACA
AGTTCTACCCGGAGACACCACAAGGCCTAGCAAGAATTATTCAGAAAGCACATGCAGAAGGAGTCTGCGGCTTGCGTTC
CGTTTCCAGACTCGAGCACCAAATGTGGGAAGCCATTAAGGATGAGCTGAACACCCTGTTGAAAGAGAATGGAGTCGAC
TTGAGTGTCGTGGTGGAAAAACAAAATGGGATGTATAAAGCAGCACCAAAACGTTTGGCTGCCACCACCGAAAAACTGG
AGATGGGTTGGAAGGCTTGGGGCAAGAGCATCATCTTTGCTCCAGAACTAGCTAACAACACCTTTGTCATCGACGGTCCT
GAGACTGAGGAATGCCCAACGGCCAACCGAGCATGGAACAGTATGGAGGTAGAAGACTTTGGATTTGGATTGACAAGCA
CTCGCATGTTCCTGAGGATTCGGGAAACGAACACAACTGAATGCGACTCGAAGATCATAGGAACCGCCGTCAAGAACAA
CATGGCTGTGCATAGTGATCTGTCGTACTGGATAGAGAGCGGACTCAACGACACCTGGAAGCTTGAGAGGGCGGTTCTAG
GAGAAGTCAAATCATGCACCTGGCCAGAAACCCACACTTTGTGGGGTGATGGAGTTCTGGAAAGTGATCTCATCATACCC
ATCACCTTGGCAGGACCCAGAAGCAATCATAACAGGAGACCAGGGTACAAAACTCAGAACCAAGGTCCATGGGATGAGG

GGCGTGTCGAGATTGACTTTGACTATTGCCCAGGAACAACAGTAACTATAAGTGATAGTTGCGGACACCGTGGACCTGCG
GCACGCACAACCACTGAGAGTGGGAAGCTCATTACAGACTGGTGTTGCAGAAGCTGCACCCTCCCTCCACTGCGCTTCCA
GACTGAGAATGGCTGTTGGTATGGAATGGAAATTCGACCCACGCGGCACGACGAAAAGACCCTCGTGCAATCGAGAGTG
AATGCATACAACGCCGACATGATTGATCCTTTTCAGTTGGGCCTTCTGGTCGTGTTCTTGGCCACCCAGGAGGTCCTTCGC
AAGAGGTGGACGGCCAAGATCAGCATTCCAGCTATCATGCTTGCACTCCTAGTCCTAGTGTTTGGGGGTATTACGTACAC
TGATGTCCTGCGATATGTCATTCTCGTCGGCGCCGCGTTTGCTGAAGCAAACTCAGGAGGAGACGTTGTGCATTTGGCACT
CATGGCTACATTCAAGATTCAACCAGTCTTTCTGGTGGCTTCCTTTTTGAAGGCAAGGTGGACCAACCAAGAGAGTATTTT
GCTCATGCTTGCAGCTGCTTTCTTTCAAATGGCTTACTATGACGCCAAGAATGTTTTGTCATGGGAAGTGCCTGACGTTTT
GAACTCTCTCTCCGTTGCGTGGATGATTCTCAGAGCTATAAGCTTCACCAACACTTCAAATGTGGTGGTGCCGCTGCTGGC
CCTTCTGACACCTGGACTAAAATGCCTAAACCTTGATGTGTACAGAATTTTGCTACTCATGGTTGGAGTTGGAAGCCTCAT
CAAAGAAAAAAGGAGCTCTGCAGCAAAAAAGAAAGGGGCTTGCCTCATCTGCCTAGCGCTGGCGTCTACAGGAGTGTTC
AATCCAATGATACTTGCAGCTGGGCTGATGGCTTGCGACCCCAACCGCAAGCGGGGCTGGCCTGCCACAGAAGTGATGA
CTGCAGTTGGACTCATGTTTGCCATCGTTGGGGGTCTGGCAGAACTTGACATAGACTCTATGGCTGTCCCCATGACCATTG
CCGGACTCATGTTCGTGGCATTTGTCATCTCTGGAAAGTCAACAGACATGTGGATTGAGAGGACGGCTGATATTACTTGG
GAAAGTGACGCTGAAATCACAGGCTCTAGTGAAAGAGTAGATGTTAGGCTGGATGATGATGGAAATTTTCAATTGATAA
ATGACCCCGGGGCGCCATGGAAAATTTGGATGCTTAGGATGGCCTGCCTGGCAATAAGTGCCTACACACCTTGGGCGATT
CTTCCCTCGATCATTGGATTTTGGATAACCCTTCAGTACACAAAGAGAGGAGGTGTCCTTTGGGACACACCATCACCTAG
GGAGTACAAGAAGGGTGATACCACCACTGGCGTTTACAGAATCATGACTCGAGGTCTGCTTGGCAGTTACCAAGCTGGA
GCCGGAGTGATGGTAGAGGGAGTGTTCCACACACTATGGCACACCACTAAGGGAGCTGCTCTCATGAGTGGTGAAGGAC
GTCTGGATCCCTATTGGGGGAGCGTGAAAGAAGACCGGCTTTGCTATGGGGGGCCATGGAAACTCCAACATAAATGGAA
TGGACATGATGAGGTCCAAATGATTGTTGTGGAGCCAGGGAAAAATGTGAAAAACGTCCAGACTAAGCCCGGAGTGTTT
AAGACACCAGAAGGAGAAATTGGGGCAGTTACGCTAGACTATCCCACTGGAACGTCAGGTTCCCCCATTGTGGATAAAA
ATGGAGATGTGATTGGATTGTATGGGAACGGCGTCATCATGCCTAATGGTTCATACATAAGCGCCATTGTGCAAGGAGAG
AGAATGGAAGAACCGGCACCAGCTGGCTTCGAACCTGAAATGTTGAGGAAGAAACAGATCACTGTCCTTGATCTGCACC
CCGGAGCAGGAAAGACGCGCAAGATACTTCCCCAAATCATCAAGGAGGCCATCAACAAAAGATTGAGGACGGCTGTACT
GGCACCCACCAGGGTCGTTGCTGCTGAGATGTCTGAGGCCCTGAGAGGACTTCCCATCCGGTACCAAACCTCAGCAGTGC
CCAGAGAGCACAGTGGAAATGAGATCGTTGATGTCATGTGCCATGCCACTCTCACACACAGGCTGATGTCTCCACACAGA
GTCCCCAATTACAATCTGTTCATAATGGATGAAGCCCATTTCACGGATCCAGCGAGCATCGCAGCCAGAGGATACATAGC
AACCAAGGTTGAATTGGGCGAAGCCGCCGCGATTTTCATGACGGCAACGCCACCCGGGACTTCTGACCCCTTTCCAGAGT
CCAATGCTCCTATCTCGGACATGCAAACAGAGATCCCAGACAGAGCCTGGAACACTGGATATGAATGGATAACTGAGTA
TGTTGGAAAGACCGTTTGGTTTGTTCCAAGTGTGAAAATGGGAAATGAGATTGCCCTCTGTCTGCAACGGGCGGGGAAGA
AGGTTATCCAGCTGAACAGAAAGTCCTATGAGACAGAGTACCCCAAGTGCAAGAACGATGATTGGGATTTTGTCATTACC
ACAGACATATCAGAAATGGGAGCAAACTTCAAGGCAAGCAGAGTGATCGACAGCCGCAAAAGCGTGAAACCCACCATC
ATTGAGGAAGGTGATGGAAGAGTCATCCTGGGGGAACCCTCAGCCATCACGGCTGCCAGCGCTGCTCAGCGGAGAGGAC
GCATAGGAAGAAACCCATCACAAGTTGGTGATGAGTATTGCTATGGAGGGCACACAAATGAGGATGATTCCAACTTTGC
CCACTGGACAGAGGCTCGCATCATGCTAGACAACATCAACATGCCGAATGGTCTGGTGGCCCAACTATATCAGCCTGAGC
GCGAGAAGGTGTACACCATGGACGGGGAATATAGGCTCAGAGGGGAAGAACGGAAGAACTTCCTTGAATTCTTGAGAAC
AGCTGATTTACCAGTCTGGCTCGCTTACAAAGTGGCAGCAGCAGGAATATCATACCATGACCGGAAGTGGTGCTTTGATG
GACCTCGAACCAACACGATTCTTGAAGACAACAATGAAGTTGAAGTCATCACGAAGTTGGGTGAGAGAAAGATCCTAAG
ACCCAGGTGGGCGGATGCCAGAGTGTACTCAGACCACCAAGCTCTAAAGTCCTTCAAAGATTTTGCATCAGGGAAACGAT
CACAAATCGGGCTCGTTGAGGTGCTCGGGAGAATGCCTGAGCACTTCATGGGGAAAACTTGGGAGGCGTTGGACACGAT
GTATGTGGTGGCAACCGCTGAAAAAGGAGGCCGAGCTCACAGGATGGCTCTTGAGGAGCTACCGGACGCCCTTCAGACA
ATAGCTTTGATTGCACTATTGAGTGTGATGTCCTTAGGTGTGTTTTTTCTACTCATGCAGAGGAAAGGCATTGGTAAGATT
GGCTTGGGAGGAGTGATCTTAGGAGCTGCCACATTCTTCTGCTGGATGGCTGAAGTCCCAGGAACGAAAATAGCAGGCAT
GCTCCTGCTTTCCCTGCTGCTCATGATTGTTTTGATTCCGGAGCCGGAAAAGCAGCGCTCACAGACTGACAACCAGCTTGC
CGTGTTTTTGATTTGTGTGCTCACACTGGTTAGCGCCGTGGCTGCCAATGAAATGGGTTGGCTGGACAAGACCAAGAATG
ACATTGGCAGCCTGTTGGGACACAGGCCAGAAACTAGAGAGACGACCCTGGGAGTTGAGAACTTCTTGCTTGATCTGCGG
CCGGCTACAGCATGGTCTCTCTATGCCGTAGCGACAGCCGTTCTTACCCCTTTGCTGAAACATCTAATCACGTCAGACTAC
ATCAACACTTCGCTGACCTCAATAAACGTCCAAGCCAGTGCGTTGTTCACCTTGGCCAGAGGCTTCCCTTTTGTGGACGTT
GGTGTGTCAGCTCTTTTGCTGGCGGCCGGGTGCTGGGGCCAAGTGACCCTGACTGTGACTGTGACTGCAGCTGCCCTGCTC

TTTTGCCACTATGCTTACATGGTACCAGGCTGGCAAGCGGAAGCCATGCGATCCGCCCAGCGGCGGACAGCTGCTGGCAT
CATGAAAAATGCAGTGGTGGATGGGATCGTGGCTACTGATGTACCTGAACTTGAGCGAACAACTCCAGTCATGCAGAAA
AAAGTTGGACAGATCATGCTGATCTTGGTGTCAATGGCCGCGGTTGTCGTCAATCCATCAGTGAGAACTGTCAGAGAGGC
TGGAATTCTGACTACAGCAGCAGCAGTCACCTTATGGGAGAATGGTGCTAGTTCAGTGTGGAATGCAACGACAGCTATTG
GGCTTTGTCACATCATGCGAGGAGGATGGCTCTCGTGTCTCTCCATCACGTGGACTCTCATCAAAAACATGGAGAAACCA
GGCCTCAAGAGGGGTGGAGCCAAGGGACGCACGCTAGGGGAAGTTTGGAAGGAGAGACTCAACCACATGACGAAGGAA
GAATTTACCAGATACAGAAAAGAAGCCATCACTGAAGTCGACCGCTCCGCGGCAAAACATGCCAGGAGAGAGGGAAAC
ATCACTGGAGGCCATCCAGTCTCACGGGGAACCGCGAAATTACGGTGGTTAGTGGAAAGGCGTTTCCTCGAGCCAGTGG
GAAAGGTTGTGGATCTCGGATGTGGTAGAGGCGGCTGGTGTTATTACATGGCCACCCAGAAGAGGGTACAGGAAGTGAA
AGGGTACACGAAAGGAGGACCTGGCCATGAAGAGCCACAACTGGTGCAGAGCTATGGTTGGAATATTGTTACCATGAAG
AGTGGAGTCGACGTCTTCTACAGACCATCAGAAGCGAGCGACACACTGCTCTGTGACATTGGAGAGTCATCGTCAAGTGC
TGAGGTAGAAGAACACCGCACCGTCCGTGTCTTGGAGATGGTGGAAGACTGGCTGCACAGAGGACCTAAGGAATTCTGC
ATCAAAGTGCTATGCCCTTACATGCCCAGAGTGATTGAGAAGATGGAAACACTCCAAAGGCGATATGGAGGTGGCCTTGT
GAGAAACCCCCTTTCACGCAACTCTACCCATGAGATGTACTGGGTGAGCCACGCTTCAGGCAACATTGTCCACTCCGTGA
ACATGACAAGCCAGGTGCTTCTGGGGAGGATGGAAAAGAAAACATGGAAGGGACCCCAGTTTGAGGAAGATGTCAACTT
GGGAAGTGGAACGCGGGCAGTAGGGAAGCCTCTCCTCAATTCTGATACTAGCAAGATCAAGAACCGAATTGAGAGGCTG
AAGAAAGAATACAGCTCCACATGGCACCAGGATGCGAATCACCCCTACAGGACCTGGAACTACCACGGAAGCTATGAAG
TGAAACCAACCGGCTCAGCCAGCTCCCTTGTGAATGGGGTAGTTAGATTACTCTCAAAACCATGGGACACTATCACCAAT
GTGACCACGATGGCCATGACAGACACCACTCCTTTCGGTCAACAACGAGTGTTCAAGGAAAAGGTGGATACAAAGGCTC
CAGAGCCTCCAGAAGGGGTCAAATACGTCCTCAATGAGACCACGAACTGGCTGTGGGCTTTTCTAGCCCGCGATAAGAA
ACCCAGGATGTGTTCCCGGGAGGAATTCATTGGAAAAGTCAACAGTAATGCCGCCCTAGGAGCGATGTTTGAAGAACAG
AACCAATGGAAGAATGCCCGGGAAGCCGTGGAGGATCCAAAGTTTTGGGAGATGGTGGATGAGGAGCGTGAGGCGCATC
TCCGTGGAGAATGCAACACCTGCATCTACAACATGATGGGAAAGAGAGAGAAGAAGCCTGGAGAGTTCGGCAAAGCTAA
AGGCAGCAGAGCCATTTGGTTCATGTGGCTAGGGGCCCGCTTCCTGGAGTTTGAAGCTCTCGGATTCCTCAATGAAGACC
ACTGGCTGGGTAGGAAGAACTCAGGAGGAGGAGTTGAAGGCTTAGGACTGCAGAAGCTTGGGTACATCTTGAAGGAAGT
CGGAACAAAGCCTGGAGGAAAGATCTACGCCGATGATACCGCAGGCTGGGACACACGCATCACCAAAGCTGACCTCGAG
AACGAAGCGAAGGTTCTTGAACTGCTGGACGGAGAACACCGACGCCTAGCGCGGTCCATCATTGAGCTCACATACCGAC
ACAAAGTCGTGAAAGTGATGAGGCCAGCGGCCGACGGGAAAACTGTGATGGATGTCATCTCCAGAGAGGATCAGAGAG
GAAGCGGGCAGGTAGTGACTTACGCCCTGAACACCTTCACTAATCTAGCAGTTCAGCTGGTCAGAATGATGGAGGGGGA
GGGGGTCATTGGACCTGACGATGTTGAAAAACTGGGAAAAGGAAAAGGCCCTAAGGTCAGAACCTGGCTGTTTGAGAAT
GGCGAGGAGCGTCTCAGTCGCATGGCCGTCAGCGGTGATGACTGCGTGGTGAAACCTTTGGACGACCGCTTCGCCACATC
ACTACACTTCCTAAATGCCATGTCAAAGGTCCGCAAAGACATTCAGGAATGGAAACCCTCGACAGGGTGGTATGACTGGC
AGCAGGTCCCATTCTGTTCAAACCATTTCACGGAACTGATCATGAAGGACGGCAGGACGTTGGTGGTCCCGTGTCGTGGA
CAAGACGAGTTGATTGGACGGGCCAGAATCTCTCCAGGGGCTGGATGGAATGTGCGCGACACCGCCTGCCTGGCGAAGT
CATACGCGCAGATGTGGCTGTTGCTTTATTTTCACCGTAGAGACCTGAGATTGATGGCTAACGCCATCTGTTCCGCTGTGC
CTGTCAACTGGGTTCCCACAGGGCGTACCACCTGGTCGATCCACGCAAAAGGAGAATGGATGACGACAGAAGACATGCT
CGCAGTCTGGAACAGAGTGTGGATTGAGGAGAATGAGTGGATGGAAGACAAAACACCAGTTGAGAGGTGGAGTGATGTT
CCATACTCTGGAAAGAGAGAAGACATTTGGTGTGGCAGTTTGATTGGCACACGAACCCGCGCTACTTGGGCTGAAAATAT
CCATGTGGCAATCAATCAGGTCCGTTCGGTGATTGGAGAAGAGAAGTATGTGGATTACATGAGCTCCCTGAGGAGGTATG
AAGACACCATTGTAGTTGAGGACACTGTTTTGTAA]

Sample 2: WNV 0967 – Goshawk – 2009 - HU

[ATGTCTAAGAAACCAGGAGGGCCCGGTAAAAACCGGGCTGTCAATATGCTAAAACGCGGTATGCCCCGCGGATTGTCCT
TGATAGGACTGAAGAGGGCTATGCTGAGTCTGATTGACGGGAAAGGCCCAATACGTTTCGTGTTGGCTCTTTTGGCGTTTT
TCAGATTCACTGCAATCGCTCCGACTCGTGCGGTGCTGGACAGATGGAGAGGCGTCAACAAACAAACAGCGATGAAGCA
TCTCTTGAGTTTCAAGAAAGAACTAGGAACTCTGACCAGTGCCATCAACCGCCGGAGCACAAAACAAAAGAAAAGAGGA
GGCACAGCGGGCTTTACTATCTTGCTTGGGCTGATCGCTTGTGCTGGAGCTGTGACCCTCTCGAACTTCCAGGGCAAAGTG
ATGATGACAGTCAATGCAACCGATGTTACTGACGTGATTACCATCCCAACAGCTGCTGGGAAAAACCTGTGCATTGTGAG
GGCTATGGACGTGGGATACCTTTGTGAGGATACTATCACTTATGAATGTCCGGTCCTAGCTGCTGGAAATGACCCTGAAG

ACATTGACTGCTGGTGCACGAAATCATCTGTTTACGTGCGCTATGGAAGATGCACAAAAACCCGGCATTCCCGTCGAAGC
AGAAGGTCTCTGACTGTCCAGACACATGGAGAAAGTACACTGGCCAACAAGAAAGGAGCTTGGTTGGACAGCACAAAAG
CCACGAGATATCTGGTGAAGACAGAATCATGGATACTGAGAAACCCGGGCTACGCCCTCGTTGCAGCGGTCATTGGATG
GATGCTAGGAAGCAACACAATGCAACGCGTCGTGTTTGCCATCCTATTGCTCCTGGTGGCACCAGCATACAGCTTCAATT
GCTTGGGAATGAGTAACAGAGACTTCCTGGAGGGAGTGTCTGGAGCTACATGGGTTGATCTGGTACTGGAAGGCGATAG
TTGTGTGACCATAATGTCAAAGGACAAGCCAACCATTGATGTCAAAATGATGAACATGGAAGCAGCCAACCTCGCAGAT
GTGCGCAGCTACTGTTACCTAGCTTCGGTCAGCGACTTGTCAACGAGAGCCGCGTGTCCAACCATGGGTGAAGCCCACAA
CGAGAAAAGAGCTGACCCCGCCTTCGTTTGCAAGCAAGGCGTTGTGGACAGAGGATGGGGAAATGGCTGCGGACTGTTT
GGAAAGGGAAGCATTGACACATGCGCGAAGTTTGCCTGTACAACCAAAGCAACTGGATGGATCATCCAGAAGGAGAACA
TCAAGTATGAGGTCGCCATATTTGTGCATGGCCCGACGACTGTTGAATCTCATGGCAATTATTCAACACAGATAGGGGCC
ACCCAGGCTGGAAGATTCAGTATAACTCCATCGGCGCCATCTTACACGCTAAAGTTGGGTGAGTATGGTGAGGTCACGGT
TGACTGTGAGCCACGGTCAGGAATAGACACTAGCGCCTATTACGTTATGTCAGTTGGTGCGAAGTCCTTCCTGGTTCACC
GAGAATGGTTTATGGATCTGAACCTGCCATGGAGCAGTGCTGGAAGCACCACGTGGAGGAATCGGGAAACACTGATGGA
GTTTGAAGAACCTCATGCCACCAAACAATCTGTTGTGGCTCTAGGGTCGCAGGAAGGTGCATTGCACCAAGCTCTGGCCG
GAGCGATTCCTGTTGAGTTCTCAAGCAATACTGTGAAGTTGACATCAGGACATCTGAAGTGTAGGGTGAAGATGGAGAA
GTTGCAGCTGAAGGGAACAACATATGGAGTATGTTCAAAAGCGTTTAAATTCGCTGGGACTCCCGCTGACACTGGCCATG
GAACGGTGGTGTTGGAACTGCAATACACCGGAACAGACGGTCCCTGCAAAGTGCCCATTTCTTCCGTAGCTTCCCTGAAT
GACCTCACACCTGTTGGAAGACTGGTGACCGTGAATCCATTTGTGTCTGTGGCCACAGCCAACTCGAAGGTCTTGATTGA
ACTCGAACCCCCGTTTGGTGACTCTTACATCGTGGTGGGAAGAGGAGAACAGCAGATAAACCACCACTGGCACAAATCT
GGGAGCAGCATCGGAAAGGCCTTTACCACTACACTCAGAGGAGCTCAACGACTCGCAGCTCTTGGAGACACTGCTTGGG
ATTTTGGGTCAGTTGGAGGGGTTTTCACCTCAGTGGGGAAAGCCATACACCAAGTCTTTGGAGGAGCTTTTAGATCACTCT
TTGGAGGGATGTCCTGGATCACACAGGGACTTCTAGGAGCCCTTCTGTTGTGGATGGGAATCAATGCCCGTGACAGGTCA
ATTGCCATGACGTTTCTTGCGGTTGGAGGAGTTTTGCTCTTCCTTTCGGTCAACGTCCATGCTGACACAGGCTGTGCCATT
GATATTGGCAGGCAAGAGCTCCGGTGTGGAAGTGGAGTGTTTATTCACAACGATGTGGAAGCCTGGATGGATCGTTACAA
GTTCTACCCGGAGACACCACAAGGCCTAGCAAGAATTATTCAGAAAGCACATGCAGAAGGAGTCTGCGGCTTGCGTTCC
GTTTCCAGACTCGAGCACCAAATGTGGGAAGCCATTAAGGATGAGCTGAACACCCTGTTGAAAGAGAATGGAGTCGACT
TGAGTGTCGTGGTGGAAAAACAAAATGGGATGTATAAAGCAGCACCAAAACGTTTGGCTGCCACCACCGAAAAACTGGA
GATGGGTTGGAAGGCTTGGGGCAAGAGCATCATCTTTGCTCCAGAACTAGCTAACAACACCTTTGTCATCGACGGTCCTG
AGACTGAGGAATGCCCAACGGCCAACCGAGCATGGAACAGTATGGAGGTAGAAGACTTTGGATTTGGATTGACAAGCAC
TCGCATGTTCCTGAGGATTCGGGAAATGAACACAACTGAATGCGACTCGAAGATCATAGGAACCGCCGTCAAGAACAAC
ATGGCTGTGCATAGTGATCTGTCGTACTGGATAGAGAGCGGACTCAACGACACCTGGAAGCTTGAGAGGGCGGTTCTAG
GAGAAGTCAAATCATGCACCTGGCCAGAAACCCACACTTTGTGGGGTGATGGAGTTCTGGAAAGTGATCTCATCATACCC
ATCACCTTGGCAGGACCCAGAAGCAATCATAACAGGAGACCAGGGTACAAAACTCAGAACCAAGGTCCATGGGATGAGG
GGCGTGTCGAGATTGACTTTGACTATTGCCCAGGAACAACAGTAACTATAAGTGATAGTTGCGGACACCGTGGACCTGCG
GCACGCACAACCACTGAGAGTGGGAAGCTCATTACAGACTGGTGTTGCAGAAGCTGCACCCTCCCTCCACTGCGCTTCCA
GACTGAGAATGGCTGTTGGTATGGAATGGAAATTCGACCCACGCGGCACGACGAAAAGACCCTCGTGCAATCGAGAGTG
AATGCATACAACGCCGACATGATTGATCCTTTTCAGTTGGGCCTTCTGGTCGTGTTCTTGGCCACCCAGGAGGTTCTTCGC
AAGAGGTGGACGGCCAAGATCAGCATTCCAGCTATCATGCTTGCACTCCTAGTCCTAGTGTTTGGGGGTATTACGTACAC
TGATGTCCTGCGATATGTCATTCTCGTCGGCGCCGCGTTTGCTGAAGCAAACTCAGGAGGAGACGTTGTGCATTTGGCACT
CATGGCTACATTCAAGATTCAACCAGTCTTTCTGGTGGCTTCCTTTTTGAAGGCAAGGTGGACCAACCAAGAGAGTATTTT
GCTCATGCTTGCAGCTGCTTTCTTTCAAATGGCTTACTATGACGCCAAGAATGTTTTGTCATGGGAAGTGCCTGACGTTTT
GAACTCTCTCTCCGTTGCGTGGATGATTCTCAGAGCTATAAGCTTCACCAACACTTCAAATGTGGTGGTGCCGCTGCTGGC
CCTTCTGACACCTGGACTAAAATGCCTAAACCTTGATGTGTACAGAATTTTGCTACTCATGGTTGGAGTTGGAAGCCTTAT
CAAAGAAAAAAGGAGCTCTGCAGCAAAAAAGAAAGGGGCTTGCCTCATCTGCCTAGCGCTGGCGTCTACAGGAGTGTTC
AATCCAATGATACTTGCAGCTGGGCTGATGGCTTGCGACCCCAACCGCAAGCGGGGCTGGCCTGCCACAGAAGTGATGA
CTGCAGTTGGACTCATGTTTGCCATCGTTGGGGGTCTGGCAGAACTTGACATAGACTCTATGGCTATCCCCATGACCATTG
CCGGACTCATGTTCGTGGCATTTGTCATCTCTGGAAAGTCAACAGACATGTGGATTGAGAGGACGGCTGATATTACTTGG
GAAAGTGACGCTGAAATCACAGGCTCTAGTGAAAGAGTAGATGTTAGGCTGGATGATGATGGAAATTTTCAATTGATAA
ATGACCCCGGGGCGCCATGGAAAATTTGGATGCTTAGGATGGCCTGCCTGGCAATAAGTGCCTACACACCTTGGGCGATT
CTTCCCTCGATCATTGGATTCTGGATAACCCTTCAGTACACAAAGAGAGGAGGTGTCCTTTGGGACACACCATCACCTAG
GGAGTACAAGAAGGGTGATACCACCACTGGCGTTTACAGAATCATGACTCGAGGTCTGCTTGGCAGTTACCAAGCTGGA
GCCGGAGTGATGGTAGAGGGAGTGTTCCACACACTATGGCACACCACTAAGGGAGCTGCTCTCATGAGTGGTGAAGGAC
GTCTGGATCCCTATTGGGGGAGCGTGAAAGAAGACCGGCTTTGCTATGGGGGGCCATGGAAACTCCAACATAAATGGAA
TGGACATGATGAGGTCCAAATGATTGTTGTGGAGCCAGGAAAAAATGTGAAAAACGTCCAGACTAAGCCCGGAGTGTTT
AAGACACCAGAAGGAGAAATTGGGGCAGTTACGCTAGACTATCCCACTGGAACGTCAGGTTCCCCCATTGTGGATAAAA
ATGGAGATGTGATTGGATTGTATGGGAACGGCGTCATCATGCCTAATGGTTCATACATAAGCGCCATTGTGCAAGGAGAG
AGAATGGAAGAACCGGCACCAGCTGGCTTCGAACCTGAAATGTTGAGGAAGAAACAGATCACTGTCCTTGATCTGCACC
CCGGAGCAGGAAAGACGCGCAAGATACTTCCCCAAATCATCAAGGAGGCCATCAACAAAAGATTGAGGACGGCTGTACT
GGCACCCACCAGGGTCGTTGCTGCTGAGATGTCTGAGGCCCTGAGAGGACTTCCCATCCGGTACCAAACCTCAGCAGTGC
CCAGAGAGCACAGTGGAAATGAGATCGTTGATGTCATGTGCCATGCCACTCTCACACACAGGCTGATGTCTCCACACAGA
GTCCCCAATTACAATCTGTTCATAATGGATGAAGCCCATTTCACGGATCCAGCGAGCATCGCAGCCAGAGGATACATAGC
AACCAAGGTTGAATTGGGCGAAGCCGCCGCGATTTTCATGACGGCAACGCCACCCGGGACTTCTGACCCCTTTCCAGAGT
CCAATGCTCCTATCTCGGACATGCAAACAGAGATCCCAGACAGAGCCTGGAACACTGGATATGAATGGATAACTGAGTA
TGTCGGAAAGACCGTTTGGTTTGTTCCAAGTGTGAAAATGGGAAATGAGATTGCCCTCTGTCTGCAACGGGCGGGGAAGA
AGGTTATCCAGCTGAACAGAAAGTCCTATGAGACAGAGTACCCCAAGTGCAAGAACGATGATTGGGATTTTGTCATCACC
ACAGACATATCAGAAATGGGAGCAAACTTCAAGGCAAGCAGAGTGATCGACAGCCGCAAAAGCGTGAAACCCACCATC

ATTGAGGAAGGTGATGGAAGAGTCATCCTGGGGGAACCCTCAGCCATCACGGCTGCCAGCGCTGCTCAGCGGAGAGGAC
GCATAGGAAGAAACCCATCACAAGTTGGTGATGAGTATTGCTATGGAGGGCACACAAATGAGGATGATTCCAACTTTGC
CCACTGGACAGAGGCTCGCATCATGCTAGACAACATCAACATGCCGAATGGTCTGGTGGCCCAACTATATCAGCCTGAGC
GCGAGAAGGTGTACACCATGGACGGGGAATATAGGCTCAGAGGGGAAGAACGGAAGAACTTCCTTGAATTCTTGAGAAC
AGCTGATTTACCAGTCTGGCTCGCTTACAAAGTGGCAGCAGCAGGAATATCATACCATGACCGGAAGTGGTGCTTTGATG
GACCTCGAACCAACACGATTCTTGAAGACAACAATGAAGTTGAAGTCATCACGAAGTTGGGTGAGAGAAAGATCCTAAG
ACCCAGGTGGGCGGATGCCAGAGTGTACTCAGACCACCAAGCTCTAAAGTCCTTCAAAGATTTTGCATCAGGGAAACGAT
CACAAATCGGGCTCGTTGAGGTGCTCGGGAGAATGCCTGAGCACTTCATGGGGAAAACTTGGGAGGCGTTGGACACGAT
GTATGTGGTGGCAACCGCTGAAAAAGGAGGCCGAGCTCACAGGATGGCTCTTGAGGAGCTACCGGACGCCCTTCAGACA
ATAGCTTTGATTGCACTATTGAGTGTGATGTCCTTAGGTGTGTTTTTTCTACTCATGCAAAGGAAAGGCATTGGTAAGATT
GGCTTGGGAGGAGTGATCTTAGGAGCTGCCACATTCTTCTGCTGGATGGCTGAAGTCCCAGGAACGAAAATAGCAGGCAT
GCTCCTGCTTTCCCTGCTGCTCATGATTGTTTTGATTCCGGAGCCGGAAAAGCAGCGCTCACAGACTGACAACCAGCTTGC
CGTGTTTTTGATTTGTGTGCTCACACTGGTTAGCGCCGTGGCTGCCAATGAAATGGGTTGGCTGGACAAGACCAAGAATG
ACATTGGCAGCCTGTTGGGACACAAGCCAGAAACTAGAGAGACGACCCTGGGAGTTGAGAACTTCTTGCTTGATCTGCGG
CCGGCTACAGCATGGTCGCTCTATGCCGTAGCGACAGCCGTTCTTACCCCTTTGCTGAAACATCTAATCACGTCAGACTAC
ATCAACACTTCGCTGACCTCAATAAACGTCCAAGCCAGTGCGTTGTTCACCTTGGCCAGAGGCTTCCCTTTTGTGGACGTT
GGTGTGTCAGCTCTCTTGCTGGCGGCCGGGTGCTGGGGCCAAGTGACCCTGACTGTGACTGTGACTGCAGCTGCCCTGCT
CTTTTGCCACTATGCTTACATGGTACCAGGCTGGCAAGCGGAAGCCATGCGATCCGCCCAGCGGCGGACAGCTGCTGGCA
TCATGAAAAATGCAGTGGTGGATGGGATCGTGGCTACTGATGTACCTGAACTTGAGCGAACAACTCCAGTCATGCAGAA
AAAAGTTGGACAGATCATGCTGATCTTGGTGTCAATGGCCGCGGTTGTCGTCAATCCATCAGTGAGAACTGTCAGAGAGG
CTGGAATTCTGACTACAGCAGCAGCAGTCACCTTATGGGAGAATGGTGCTAGTTCAGTGTGGAATGCAACGACAGCTATT
GGGCTTTGTCACATCATGCGAGGAGGATGGCTCTCGTGTCTCTCCATCACGTGGACTCTCATCAAAAACATGGAGAAACC
AGGCCTCAAGAGGGGTGGAGCCAAGGGACGCACGCTAGGGGAAGTTTGGAAGGAGAGACTCAACCACATGACGAAGGA
AGAATTTACCAGATACAGAAAAGAAGCCATCACTGAAGTCGACCGCTCCGCGGCAAAACATGCCAGGAGAGAGGGAAA
CATCACTGGAGGCCATCCAGTCTCACGGGGAACCGCGAAATTACGGTGGTTAGTGGAAAGGCGTTTCCTCGAGCCAGTGG
GAAAGGTTGTGGATCTCGGATGTGGTAGAGGCGGCTGGTGTTATTACATGGCCACCCAGAAGAGGGTACAGGAAGTGAA
AGGGTACACGAAAGGAGGACCTGGCCATGAAGAGCCACAACTGGTGCAGAGCTATGGTTGGAATATTGTTACCATGAAG
AGTGGAGTCGACGTCTTCTACAGACCATCAGAAGCGAGCGACACACTGCTCTGTGACATTGGAGAGTCATCGTCAAGTGC
CGAGGTAGAAGAACACCGCACCGTCCGTGTCTTGGAGATGGTGGAAGACTGGCTGCACAGAGGACCTAAGGAATTCTGC
ATCAAAGTGCTATGCCCTTACATGCCCAGAGTGATTGAGAAGATGGAAACACTCCAAAGGCGATATGGAGGTGGCCTTGT
GAGAAACCCCCTTTCACGCAACTCTACCCATGAGATGTACTGGGTGAGCCACGCTTCAGGCAACATTGTCCACTCCGTGA
ACATGACAAGCCAGGTGCTTCTGGGGAGGATGGAAAAGAAAACATGGAAGGGACCCCAGTTTGAGGAAGATGTCAACTT
GGGAAGTGGAACGCGGGCAGTAGGGAAGCCTCTCCTCAATTCTGATACTAGCAAGATCAAGAACCGAATTGAGAGGCTG
AAGAAAGAATACAGCTCCACATGGCACCAGGATGCGAATCACCCCTACAGGACCTGGAACTACCACGGAAGCTATGAAG
TGAAACCAACCGGCTCAGCCAGCTCCCTTGTGAATGGGGTAGTTAGATTACTCTCAAAACCATGGGACACTATCACCAAT
GTGACCACGATGGCCATGACAGACACCACTCCTTTCGGTCAACAACGAGTGTTCAAGGAAAAGGTGGATACAAAGGCTC
CAGAGCCTCCAGAAGGGGTCAAATACGTCCTCAATGAGACCACGAACTGGCTGTGGGCTTTTCTAGCCCGCGATAAGAA
ACCCAGGATGTGTTCCCGGGAGGAATTCATTGGAAAAGTCAACAGTAATGCCGCCCTAGGAGCGATGTTTGAAGAACAG
AACCAATGGAAGAACGCCCGGGAAGCCGTGGAGGATCCAAAGTTTTGGGAGATGGTGGATGAGGAGCGTGAGGCGCAT
CTCCGTGGAGAATGCAACACCTGCATCTACAACATGATGGGAAAGAGAGAGAAGAAGCCTGGAGAGTTCGGCAAAGCTA
AAGGCAGCAGAGCCATTTGGTTCATGTGGCTAGGGGCCCGCTTCCTGGAGTTTGAAGCTCTCGGATTCCTCAATGAAGAC
CACTGGCTGGGTAGGAAGAACTCAGGAGGAGGAGTTGAAGGCTTAGGACTGCAGAAGCTTGGGTACATCTTGAAGGAAG
TCGGAACAAAGCCTGGAGGAAAGATCTACGCCGATGATACCGCAGGCTGGGACACACGCATCACCAAAGCTGACCTCGA
GAACGAAGCGAAGGTTCTTGAACTGCTGGACGGAGAACACCGACGCCTAGCGCGGTCCATCATTGAGCTCACATACCGA
CACAAAGTCGTGAAAGTGATGAGGCCAGCGGCCGACGGGAAAACTGTGATGGATGTCATCTCCAGAGAGGATCAGAGAG
GAAGCGGGCAGGTAGTGACTTACGCCCTGAACACCTTCACTAATCTAGCAGTTCAGCTGGTCAGAATGATGGAGGGGGA
GGGGGTCATTGGACCTGACGATGTTGAAAAACTGGGAAAAGGAAAAGGCCCTAAGGTCAGAACCTGGCTGTTTGAGAAT
GGCGAGGAGCGTCTCAGTCGCATGGCCGTCAGCGGTGATGACTGCGTGGTGAAACCTTTGGACGACCGCTTCGCCACATC
ACTACACTTCCTAAATGCCATGTCAAAGGTCCGCAAAGACATTCAGGAATGGAAACCCTCGACAGGGTGGTATGACTGGC
AGCAGGTCCCATTCTGTTCAAACCATTTCACGGAGCTGATCATGAAGGACGGCAGGACGTTGGTGGTCCCGTGTCGTGGA
CAAGACGAGTTGATTGGACGGGCCAGAATCTCTCCAGGGGCTGGATGGAATGTGCGCGACACCGCCTGCCTGGCGAAGT
CATACGCGCAGATGTGGCTGTTGCTTTATTTTCACCGTAGAGACCTGAGATTGATGGCTAACGCCATCTGTTCCGCTGTGC
CTGTCAACTGGGTTCCCACAGGGCGTACCACCTGGTCGATCCACGCAAAAGGAGAATGGATGACGACAGAAGACATGCT
CGCAGTCTGGAACAGAGTGTGGATTGAGGAGAATGAGTGGATGGAAGACAAAACACCAGTTGAGAGGTGGAGTGATGTT
CCATACTCTGGAAAGAGAGAAGACATTTGGTGTGGCAGTTTGATTGGCACACGAACCCGCGCTACTTGGGCTGAAAATAT
CCATGTGGCAATCAATCAGGTCCGTTCGGTGATTGGAGAAGAGAAGTATGTGGATTACATGAGCTCCCTGAGGAGGTATG
AAGACACCATTGTAGTTGAGGACACTGTTTTGTAA]

Appendix B

Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)
TN93+G	93	41117,784	40090,627	-19952,295	n/a	0,230	12,670	0,272	0,218	0,227	0,283
TN93+G+I	94	41127,399	40089,198	-19950,580	0,470	0,830	12,710	0,272	0,218	0,227	0,283
GTR+G	96	41134,473	40074,184	-19941,072	n/a	0,220	12,570	0,272	0,218	0,227	0,283
GTR+G+I	97	41145,470	40074,136	-19940,047	0,490	0,820	12,630	0,272	0,218	0,227	0,283
TN93+I	93	41162,147	40134,990	-19974,476	0,460	n/a	12,440	0,272	0,218	0,227	0,283
GTR+I	96	41180,437	40120,147	-19964,053	0,460	n/a	12,410	0,272	0,218	0,227	0,283
TN93	92	41237,571	40221,458	-20018,711	n/a	n/a	12,310	0,272	0,218	0,227	0,283
GTR	95	41256,777	40207,531	-20008,746	n/a	n/a	12,310	0,272	0,218	0,227	0,283
HKY+G	92	41289,845	40273,732	-20044,848	n/a	0,140	12,980	0,272	0,218	0,227	0,283
HKY+G+I	93	41297,080	40269,923	-20041,943	0,570	0,840	13,050	0,272	0,218	0,227	0,283
K2+G	89	41311,455	40328,476	-20075,221	n/a	0,170	12,640	0,250	0,250	0,250	0,250
K2+G+I	90	41319,684	40325,660	-20072,812	0,540	0,830	12,680	0,250	0,250	0,250	0,250
T92+G	90	41322,530	40328,506	-20074,235	n/a	0,170	12,640	0,245	0,245	0,255	0,255
T92+G+I	91	41330,756	40325,688	-20071,826	0,540	0,850	12,680	0,245	0,245	0,255	0,255
HKY+I	92	41359,190	40343,077	-20079,520	0,460	n/a	12,470	0,272	0,218	0,227	0,283
K2+I	89	41370,849	40387,870	-20104,917	0,460	n/a	12,390	0,250	0,250	0,250	0,250
T92+I	90	41382,028	40388,004	-20103,984	0,460	n/a	12,390	0,245	0,245	0,255	0,255
HKY	91	41444,909	40439,840	-20128,902	n/a	n/a	12,280	0,272	0,218	0,227	0,283
K2K2	88	41452,325	40480,389	-20152,178	n/a	n/a	12,280	0,250	0,250	0,250	0,250
T92	89	41463,556	40480,576	-20151,271	n/a	n/a	12,280	0,245	0,245	0,255	0,255
JC+G	88	42593,208	41621,273	-20722,620	n/a	0,170	0,500	0,250	0,250	0,250	0,250
JC+G+I	89	42602,277	41619,298	-20720,632	0,53	0,84	0,5	0,25	0,25	0,25	0,25
JC+I	88	42649,927	41677,991	-20750,979	0,46	n/a	0,5	0,25	0,25	0,25	0,25
JCJC	87	42730,38	41769,489	-20797,728	n/a	n/a	0,5	0,25	0,25	0,25	0,25

Model	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(GT)	r(GC)
TN93+G	0,008	0,008	0,137	0,010	0,336	0,010	0,010	0,323	0,010	0,132	0,008	0,008
TN93+G+I	0,008	0,008	0,137	0,010	0,337	0,010	0,010	0,324	0,010	0,132	0,008	0,008
GTR+G	0,012	0,007	0,137	0,015	0,336	0,015	0,008	0,323	0,001	0,132	0,012	0,001
GTR+G+I	0,012	0,007	0,136	0,015	0,337	0,016	0,008	0,324	0,000	0,131	0,013	0,000
TN93+I	0,008	0,008	0,138	0,010	0,335	0,010	0,010	0,322	0,010	0,133	0,008	0,008
GTR+I	0,012	0,007	0,138	0,015	0,335	0,015	0,009	0,322	0,002	0,133	0,011	0,002
TN93	0,008	0,008	0,139	0,010	0,334	0,010	0,010	0,321	0,010	0,134	0,008	0,008
GTR	0,012	0,007	0,139	0,015	0,333	0,015	0,009	0,321	0,002	0,134	0,011	0,002
HKY+G	0,008	0,008	0,262	0,010	0,210	0,010	0,010	0,202	0,010	0,252	0,008	0,008
HKY+G+I	0,008	0,008	0,262	0,010	0,210	0,010	0,010	0,202	0,010	0,252	0,008	0,008
K2+G	0,009	0,009	0,232	0,009	0,232	0,009	0,009	0,232	0,009	0,232	0,009	0,009
K2+G+I	0,009	0,009	0,232	0,009	0,232	0,009	0,009	0,232	0,009	0,232	0,009	0,009
T92+G	0,009	0,009	0,236	0,009	0,236	0,009	0,009	0,227	0,009	0,227	0,009	0,009
T92+G+I	0,009	0,009	0,236	0,009	0,236	0,009	0,009	0,227	0,009	0,227	0,009	0,009
HKY+I	0,008	0,009	0,261	0,010	0,210	0,011	0,010	0,202	0,011	0,252	0,008	0,009
K2+I	0,009	0,009	0,231	0,009	0,231	0,009	0,009	0,231	0,009	0,231	0,009	0,009
T92+I	0,009	0,010	0,236	0,009	0,236	0,010	0,009	0,227	0,010	0,227	0,009	0,010
HKY	0,008	0,009	0,261	0,010	0,209	0,011	0,010	0,201	0,011	0,251	0,008	0,009
K2K2	0,009	0,009	0,231	0,009	0,231	0,009	0,009	0,231	0,009	0,231	0,009	0,009
T92	0,009	0,010	0,236	0,009	0,236	0,010	0,009	0,227	0,010	0,227	0,009	0,010
JC+G	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083
JC+G+I	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083
JC+I	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083
JCJC	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083

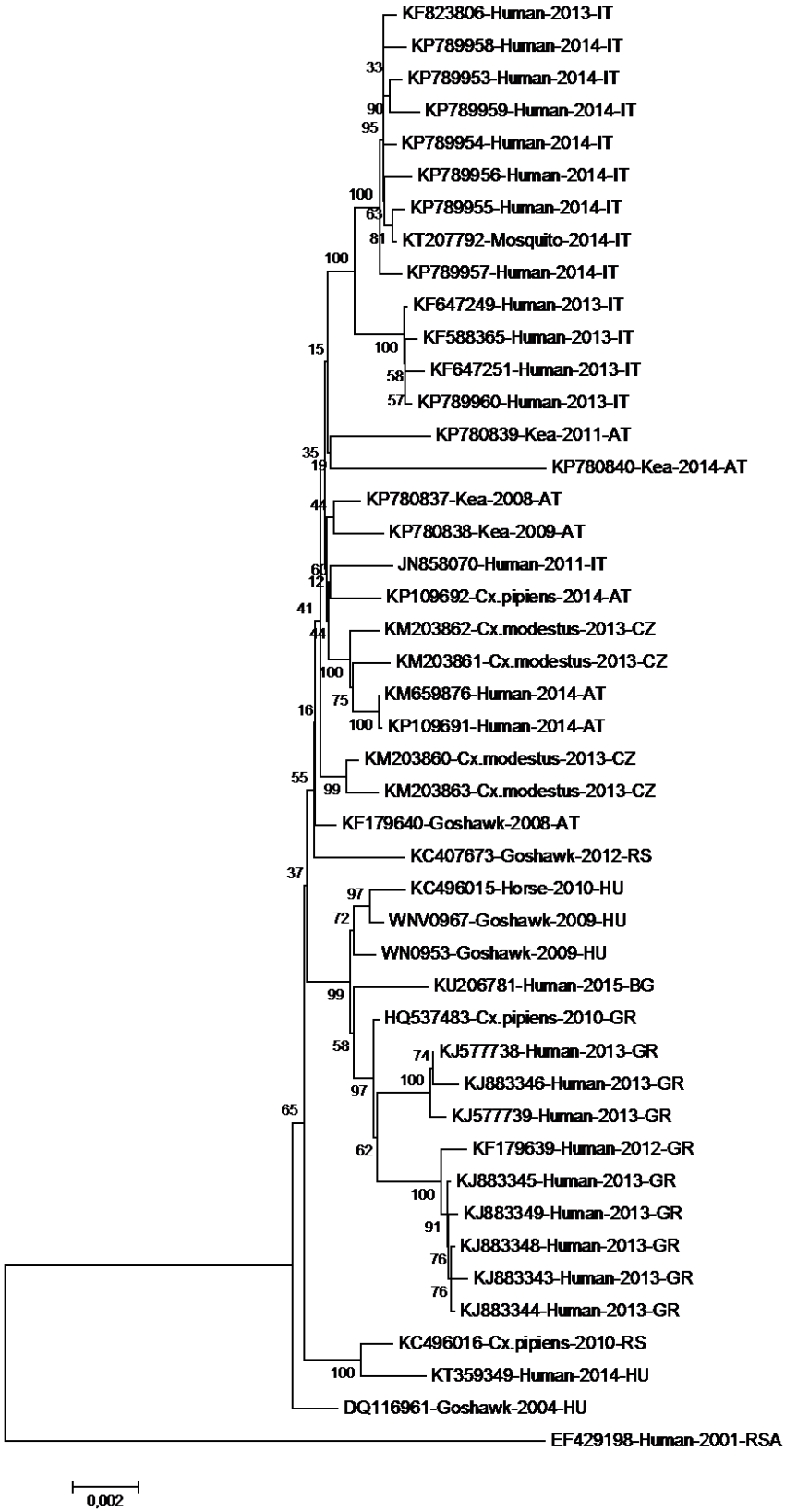
NOTE.-- Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) are also presented [1]. Non-uniformity of evolutionary rates among sites may be modelled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. Assumed or estimated values of transition/transversion bias (R) are shown for each model, as well. They are followed by n
ucleotide frequencies (f) and rates of base substitutions (r) for each nucleotide pair. Relative values of instantaneous r should be considered when evaluating them. For simplicity, sum of r values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. The analysis involved 45 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated.

There were a total of 10285 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [2].

Abbreviations: GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei;
T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

Appendix C

Neighbour joining phylogenetic tree



Appendix D

Maximum likelihood phylogenetic tree