GENETIC MECHANISMS OF VIRUS EVOLUTION AND EMERGENCE: RECOMBINATION, REASSORTMENT, OVERPRINTING AND MUTATION

by

ANDREW BROWNELL ALLISON

(Under the Direction of David E. Stallknecht)

ABSTRACT

Although the emergence of a novel pathogen from an existing virus (or viruses) may be a multi-faceted process involving many interdependent viral, host, and/or environmental factors, the fundamental catalyst is genetic variation of the virus. For recombination and reassortment, both mechanisms involve the direct exchange of genetic material from two (or more) parental viruses, such that the novel virus is a genetic chimera of the parental viruses. For the recombinant or reassortant to replicate efficiently, be packaged and assembled correctly, and subsequently be competently transmitted in nature, the genetic elements derived from each parental virus must be structurally and functionally compatible in order for the novel virus to be viable. In the research on recombination in alphaviruses and reassortment in orbiviruses presented here, findings suggest this to be case, as both examples reinforce the notion that the genetic and/or structural compatibility between the parental viruses was likely a prerequisite for the emergence of the novel virus(es). In the two following examples of viral evolution involving overprinting and mutation, the emergence of the novel virus or variant was not dictated by the exchange of genes from different parental viruses, but rather by a change

in existing sequence. For overprinting in rhabdoviruses, this involved the *de novo* synthesis of a novel protein through the use of an overlapping reading frame, thus increasing the coding capacity of the virus without incorporating any additional genetic elements. In the case of mutation in parvoviruses, the change in existing sequence allowed the new virus variant to jump species, thereby expanding the host range of the virus. In the description of the viruses presented within, genetic variability, whether through the process of change (overprinting or mutation) or exchange (recombination or reassortment), resulted in the emergence of a novel virus that was not only altered genetically, but also resulted in changes in the pathogenicity, antigenicity, ecology, and/or epidemiology of the virus relative to the parental virus(es), thereby shaping the evolutionary pathway of each novel virus.

INDEX WORDS: virus evolution, viral genetics, recombination, reassortment, overprinting, overlapping reading frame, mutation, selection, alphavirus, Highlands J virus, Fort Morgan virus, orbivirus, epizootic hemorrhagic disease virus, serotype, rhabdovirus, Durham virus, canine parvovirus.

GENETIC MECHANISMS OF VIRUS EVOLUTION AND EMERGENCE: RECOMBINATION, REASSORTMENT, OVERPRINTING AND MUTATION

by

ANDREW B. ALLISON

B.S., Virginia Tech, 1996

M.S., University of Georgia, 2000

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2010

GENETIC MECHANISMS OF VIRUS EVOLUTION AND EMERGENCE: RECOMBINATION, REASSORTMENT, OVERPRINTING AND MUTATION

by

ANDREW B. ALLISON

Major Professor: David E. Stallknecht

Committee: Claudio L. Afonso
Mark W. Jackwood
Liliana Jaso-Friedmann
Daniel G. Mead

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2010

DEDICATION

To my parents.

ACKNOWLEDGEMENTS

I would like to thank my boss and advisor, David Stallknecht, for giving me the freedom to pursue this research. I would also like to thank my committee members, Daniel Mead, Mark Jackwood, Claudio Afonso, and Liliana Jaso-Friedmann for their support of this project and to those at SCWDS who made me laugh over the (many, many) years.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

The concept of an 'emerging' virus is an arbitrary term that can be defined as either 1) a previously unknown virus that is newly recognized as a plant, animal, or human pathogen, or 2) a previously described virus in which the incidence or severity of disease has increased relative to all previous accounts. In either case, the reasons for their emergence are usually linked to the genetics and/or natural history (i.e., transmission dynamics) of the specific virus under question. As such, the factors the precipitate the emergence of any given virus can be highly varied and/or multi-factorial. Oftentimes, these factors might not be inherently obvious or easily defined.

Prototypical, highly-profiled examples of newly emerging viruses are the henipaviruses, *Nipah virus* (NiV) and *Hendra virus* (HeV), of the family *Paramyxoviridae*. Prior to 1994, neither of these viruses was known to science, but in dramatic fashion, both emerged from unknown reservoirs to cause severe disease in humans and animals. In the case of NiV, the index outbreak occurred in Malaysia in 1998, resulting in 105 human deaths and the eventual culling of over a million swine (Wong *et al*., 2002); for HeV, it was first recognized in Australia in 1994 where it initiated a much smaller, although no less sensational, focal outbreak resulting in a single human fatality and 13 equine deaths (Murray *et al*., 1995). Since their initial discoveries,

intensive research has been conducted on NiV and HeV, including 1) the characterization of their genomes and ultrastrucure (Wang *et al*., 2000; Chan *et al*., 2001; Hyatt *et al*., 2001), 2) field studies to determine the reservoir (i.e., *Pteropus* fruit bats or flying foxes) and rates of seroprevalence and incidence of the viruses in their natural hosts (Halpin *et al*., 2000; Yob *et al*., 2001; Chua *et al*., 2002b; Field *et al*., 2001), 3) infectivity and transmissibility trials in numerous species including fruit bats and horses (Williamson *et al*., 1998; Williamson *et al*., 2001; Middleton *et al*., 2007), 4) pathogenesis studies detailing comparative host pathology and tissue tropism (Hooper *et al*., 2001; Wong *et al*., 2003; Guillaume *et al*., 2009), 5) ecological assessments to determine the human or climatic factors which may have facilitated viral emergence (Chua *et al*., 2002a; Epstein *et al*., 2006), and 6) identification of the virus-host cell interactions at both the molecular and structural level, which has spurred on the development of structure-based antiviral therapies and vaccines (Bossart *et al*., 2002; Bonaparte *et al*., 2005; Bowden *et al*., 2008; Xu *et al*., 2008). In other words, retroactively, much as been learned about HeV and NiV.

As a consequence of the recognition of HeV, surveillance studies targeted at flying foxes led to the isolation of *Australian bat lyssavirus* (ABLV), a novel genotype within the genus *Lyssavirus*, in New South Wales in 1996 (Fraser *et al*., 1996; Gould *et al*., 1998). Since its fortuitous isolation, ABLV has been identified as the cause of human deaths in Australia, partly in response to its initial characterization during surveillance for HeV (McCall *et al*., 2000). Consequential to the outbreak of NiV, another flying fox-associated paramyxovirus, *Tioman virus* (TioPV; genus *Rubulavirus*), was isolated in Malaysia in 1999 (Chua *et al*., 2001). TioPV was later shown to be related to an

additional paramyxovirus transmitted by fruit bats, *Menangle virus* (MenPV), which first emerged in Australia in 1997 during an outbreak of reproductive disease in swine (Philbey *et al*., 1998). Thus, within the span of five years, four novel flying fox-associated paramyxoviruses were discovered; two of which, HeV (Australia) and NiV (Malaysia), were responsible for human deaths, while the other two, MenPV (Australia) and TioPV (Malaysia), although apparently zoonotic (Chant *et al*., 1998; Yaiw *et al*., 2007), appeared to be relatively harmless to humans.

Unfortunately, intensive surveillance initiatives aimed at identifying and characterizing novel viruses and defining their transmission dynamics are usually only undertaken retrospectively after human or companion/food animal morbidity or mortality is reported. As TioPV, unlike HeV or NiV, has not been associated with any animal or human deaths, it is unlikely that any surveillance efforts would have ever been initiated if it had been discovered prior to the emergence of the henipaviruses. That is to say, without the emergence of HeV and NiV, the isolation of TioPV would have likely been a biological curiosity at best. However, the identification and characterization (i.e., at a molecular and ecological level) of apparently non-pathogenic viruses such as TioPV could potentially have great utility, as other related viruses of human and veterinary importance that circulate under similar ecological and epidemiological conditions (i.e., HeV, NiV, and ABLV) obviously exist. As such, less or non-pathogenic 'relatives' can possibly serve as excellent (and safer) surrogate models for understanding many aspects of virus ecology and modes of transmission (e.g., bat-to bat transmission or spillover from bats to aberrant hosts), especially as it relates to closely-related pathogenic viruses. Obviously much as been learned (and in many cases, may only be learned) about NiV

and HeV by directly studying the agents themselves; however, both of these viruses are biosafety level 4 pathogens and, therefore, a vast amount of research is primarily limited to a few facilities capable of such containment. Although the emergence of the henipaviruses was demonstrated not to be a result of a genetic change in the viruses that altered host range or pathogenicity, but rather by the increasing encroachment of the spatial interface between humans and the reservoir (*Pteropus* bats) of NiV and HeV, they serve as a reminder that potential insights into the biological and ecological aspects of pathogenic viruses, or even the detection and characterization of pathogenic viruses prior to their emergence in humans or animals, can possibly be gained by research into less pathogenic, closely-related viruses (e.g., TioPV and MenPV).

The concept of the underutilization of research into non-pathogenic viruses is exemplified by the fact that, in this study, the genomic sequences of *Highlands J virus* (HJV) and *Fort Morgan virus* (FMV) [genus *Alphavirus*, family *Togaviridae*], both of which are derived from *eastern equine encephalitis virus* (EEEV) and related to *Western equine encephalitis virus* (WEEV) (i.e., two extensively researched human and equine pathogens), are described here for the very first time; not surprisingly, neither HJV nor FMV are human or equine pathogens. A vast amount of information related to the replication, genetics, and pathogenesis of alphaviruses has been determined predominately using two Old World alphaviruses, *Sindbis virus* (SINV) and *Semliki Forest virus* (SFV), as model systems. The cumulative body of many in-depth studies using these two model systems is enormous and has not only revolutionized the field of alphavirus research, but has transformed current concepts in the understanding of the structural, cellular, and molecular biology of animal viruses as a whole. That being said,

clear differences in the pathogenesis, genetics, structure, and viral-host interactions exist between Old and New World viruses (Paredes *et al*., 2001; Powers *et al*., 2001; Garmashova *et al*., 2007; Ryman and Klimstra, 2008), and data obtained from Old World alphavirus model systems (e.g., murine models) may not always be applicable, or be relevantly extrapolated, to bird-associated pathogenic alphaviruses present in the New World (e.g., WEEV and EEEV). As such, information obtained from more closely-related North American alphaviruses like HJV or FMV may provide a unique perspective into the factors associated with the transmission dynamics and host-viral interactions of New World alphaviruses. Indeed, details of why viruses like HJV are not pathogenic in mammalian hosts may, in fact, disclose why viruses like EEEV are. Since HJV is derived from EEEV, uses the same enzootic vector as EEEV, utilizes the same amplifying hosts as EEEV, and shares the same geographic distribution as EEEV, suggests that much can be potentially learned regarding the factors which govern the maintenance and transmission of EEEV (which is classified as select agent by the U.S. Department of Health and Human Services) by using HJV as a surrogate or comparative model. However, in the past, outside of experimental infections in poultry spurred on by concomitant HJV and EEEV outbreaks in North Carolina in 1991 (Ficken *et al*., 1993; Guy *et al*., 1993; Guy *et al*., 1994), very little, if any, comparative research between HJV and EEEV has been performed. Hopefully, the data presented here may act as a starting point for future research.

Additionally, as part of this research, a novel rhabdovirus [Durham virus (DURV), genus *Vesiculovirus*, family *Rhabdoviridae*] isolated from the brain of a neurological American coot (*Fulica americana*), is described for the first time. As this is

the only isolate of DURV that has yet to be recovered, its significance as a pathogen (avian or otherwise) is unknown. Genetic characterization of DURV revealed that it contained a genomic organization that was not only similar to members of the *Vesiculovirus* genus [e.g., *Vesicular stomatitis Indiana virus* (VSINV), a pathogen of cattle and horses], but that the overprinted gene of DURV (i.e., C gene) exhibited amino acid identity to the fusion/attachment protein (i.e., G protein) of VSINV, a potentially extremely intriguing finding from an evolutionary standpoint. Although DURV superficially appears to be only a biological curiosity along the lines of HJV or FMV, much can potentially be gained from proactively characterizing novel viruses, even if they do not appear be of significant public or veterinary importance, as outlined below. The concept of virus evolution rests on the fact that selective forces are continually acting on viruses, and due to the inherent genetic pliability of many viruses, they can adapt quickly to such selection pressures which can readily transform a biological curiosity into public health threat.

As aforementioned, a second alternate definition of an 'emerging' virus is a virus that, although previously known to science, was never regarded as a significant pathogen (i.e., the incidence or severity of disease was either mild or was contained through preventative measures), but has since become a more serious veterinary or public health threat. As in the case of the emergence of novel viruses like HeV or NiV, research into understanding the potential factors responsible for the sudden increase in the incidence and/or severity of disease associated with a known pathogen generally escalates concomitantly with the increase in reported cases. Oftentimes, as the virus has not

previously posed a serious health threat, a thorough understanding of the genetics of the virus and the ecological conditions under which it circulates may be lacking.

A quintessential example of this scenario is the emergence of the alphavirus, *Chikungunya virus* (CHIKV), in Africa and Asia during the last decade. CHIKV was first isolated from a human patient exhibiting a mild febrile disease in Tanzania in 1953 (Ross, 1956). From a historical perspective, CHIKV infection generally presented as a mild to moderate polyarthralgia syndrome, and although it could be characterized by a persistent painful arthritis, infection in most patients was generally self-limiting and rarely, if ever, fatal (Brighton *et al*., 1983) However, from 1999-2007, large-scale outbreaks of CHIKV, involving cumulatively millions of people, were reported in the Democratic Republic of the Congo (Pastorino *et al*., 2004), Indonesia (Laras *et al*., 2005), the Indian subcontinent (Dash *et al*., 2007), and many island nations in the Indian Ocean (Sourisseau *et al*., 2007). Not only was the shear magnitude of the outbreaks unprecedented, but the clinical course of the disease presented with much more severe sequelae, including fatal encephalitis and maternoneonatal transmission leading to neonatal encephalopathy (Rampal *et al*., 2007; Renault *et al*., 2007; Sourisseau *et al*., 2007). In the 2006 outbreak on Reunion Island, located off the east coast of Madagascar, it was estimated that 34% (265,000) of the islands' population had clinical disease associated with CHIKV infection, resulting in 237 deaths (Charrel *et al*., 2007; Renault *et al*., 2007; Sourisseau *et al*., 2007). Thus, a virus previously regarded as an anecdotal human pathogen had become a very serious public health threat. Consequently, as CHIKV was previously of low concern from a public health standpoint, there was a limited knowledge base regarding CHIKV, leading to a flurry of research into many

aspects of the molecular biology, vector transmission, and pathogenesis of the virus (Ozden *et al*., 2007; Tsetsarkin *et al*., 2007; Wang *et al*. 2008; Ziegler *et al*., 2008; Kam *et al*., 2009).

One proposed molecular basis for the emergence of CHIKV, recognized in virus isolates from the 2005-2006 Indian Ocean outbreaks, was a single amino acid change in the E1 glycoprotein (Tsetsarkin *et al*., 2007). This single mutation (A226V) was demonstrated to increase CHIKV infectivity, dissemination, and transmissibility in *Aedes albopictus*, a mosquito species not normally associated with CHIKV transmission (Reiter *et al*., 2006), but incriminated as the vector in the Indian Ocean outbreaks. The take home message from the emergence of CHIKV is that under the right circumstances, a virus may evolve from a relatively low priority health concern to a serious pathogen even through very subtle genetic changes, suggesting that basic and applied research into viruses that are not considered as immediate public or veterinary health threats has a utility, not only in understanding general concepts in virus biology, but also in the advent that it may mutate, reassort, or recombine into a virus with altered transmissibility, host range, or virulence.

Another subject of this research is the characterization of a novel reassortant of *epizootic hemorrhagic disease virus* (EHDV; genus *Orbivirus*, family *Reoviridae*), a virus which infects ruminants (e.g., cattle, sheep, goats, deer) (Mertens *et al*., 2005). In the United States, EHDV is not considered a major concern from a veterinary health standpoint, whereas the closely-related *bluetongue virus* (BTV), owing primarily to its pathogenic potential in sheep, in association with trade restrictions of cattle and germplasm from BTV endemic areas, has been the focus of in-depth research for many

years. During the 1960's-1980's, intensive research elucidated many aspects of BTV epidemiology, vector-host relationships, and pathogenesis (Borden *et al*., 1971; Jones and Foster, 1974; Barber, 1979; Metcalf *et al*., 1981; Tabachnick, 1996). Subsequently, details from this research became widely accepted tenets, and thus, ongoing research into these areas has consequently subsided. However, as previously noted, a primary attribute of RNA virus biology is the ability to readily adapt to changing conditions.

In Europe, prior to 2006, BTV had not been documented in any nation north of the Alps; since then, four different serotypes (BTV-1, -6, -8, and -11) have been collectively introduced into Belgium, France, Germany, Denmark, the Netherlands, Switzerland, the Czech Republic, and the United Kingdom (Hoffman *et al*., 2008). The direct and indirect costs of the introduction of BTV into northern Europe due to trade restrictions, culling and/or decreased productivity of animals, vaccination campaigns, and loss of jobs are difficult to assess. However, it has been estimated that in northern Europe in 2007, bluetongue cost over £95 million in just direct losses alone (http://www.iah.bbsrc.ac.uk/press_release/2008/2008_14.htm). Prior to the emergence of BTV in the Netherlands in 2006, it was presumed that, based on the lack of appropriate vectors and the latitudinal extremes (i.e., BTV was believed to be confined to 40°N to 35°S), BTV would never pose a problem in the northern Europe. The economic consequences of the introduction of a pathogenic exotic serotype into the United States, such as the BTV-8 strain that is currently circulating in Europe (Conraths *et al*., 2009), is unknown; however, with the recent identification of additional exotic BTV serotypes previously never detected before in the United States (e.g., BTV-1, BTV-12; see Chapter 4, *Case histories* section), the question might not be what if, but when.

Although EHDV is not historically regarded as a pathogen of cattle, the incidence of morbidity and mortality associated with EHDV infection in recent years, similar to that observed with BTV, has been unprecedented. Historically, other than circumstantial isolated cases, no large-scale EHDV outbreaks (i.e., in domestic ruminants) had ever been documented. Since 2003, morbidity and mortality in association with EHDV outbreaks in cattle has been reported from Israel, Algeria, Morocco, and Reunion Island (Bréard *et al*., 2004; Yadin *et al*., 2008; Temizel *et al*., 2009), suggesting that the geographical distribution and/or the recognition of clinical disease due to EHDV infection is increasing. Other than the epidemic in Israel, the serotype responsible for the global outbreaks is EHDV-6, the same serotype that is described (for the first time in the United States) in this research. The reason(s) perpetuating the geographical expansion of orbiviruses are currently unknown, but may include climate change affecting the distribution of *Culicoides* vectors, increased movements of animals through international trade, or genetic changes in the virus allowing them to broaden their vector/host range and/or alter their pathogenicity, or potentially a combination of these (or undescribed) factors. The characterization of an reassortant EHDV containing the outer capsid proteins of an exotic serotype (EHDV-6) and the remaining structural and nonstructural proteins of an endemic serotype (EHDV-2) provides a unique example of genetic shift in orbiviruses and demonstrates the plasticity of RNA viruses, not only in terms of genetic plasticity, but the also in the ability to enter into, and become established in, novel environmental niches previously never occupied. Currently, the ramifications of the emergence of this novel reassortant in the United States remain to be seen.

Corollary to the previous examples of emerging viruses is the related term of a ‘reemerging’ virus. A ‘reemerging’ virus can be characterized as a virus that, historically, (unlike ‘emerging’ viruses) *was* considered to be a serious pathogen in a given area, but whose incidence had declined significantly, only to once again pose a serious health threat. The final example of viral evolution in this research is the detection and characterization of novel genetic and antigenic variants of *canine parvovirus* (CPV) (genus *Parvovirus*, family *Parvoviridae*) isolated from raccoons. CPV emerged as a recognized pathogen in 1978, apparently as a host range variant of *feline panleukopenia virus* (FPV) (Parrish *et al*., 1988). This species jump of a feline virus into dogs was facilitated by only a few amino acid changes in the capsid protein (VP2), the protein responsible for binding to the transferrin receptor of the host, thereby mediating cellular entry of the virus (Parker *et al*., 2001). Within two years of its emergence, CPV spread globally, initiating a pandemic of hemorrhagic gastroenteritis in dogs (Parrish *et al*., 1988). Through intensive vaccination programs and public awareness, canine parvoviral enteritis has become a manageable disease.

Within recent years, antigenic and genetic variants of CPV have emerged and spread in the dog population [e.g., CPV-2c in Europe and North America (Decaro *et al*., 2009)]. Consequential to the recognition of these new variants has been the question of the level of protection of current vaccines. So far, evidence to suggest that variants, such as CPV-2c, have undergone such substantial antigenic change to suggest vaccine failure is limited to a few isolated cases (Decaro *et al*., 2008). The novel CPV isolates obtained from raccoons that are discussed in this research have undergone more mutations than previously recognized variants, which is likely indicative of the recent adaptation of the

virus to a novel host. Additionally, the CPV variants appear to still be rapidly evolving, which has not only resulted in substantial genetic change, but may also significantly alter the antigenicity of the viruses. Whether these novel variants still have the ability to infect dogs *in vivo* remains to be seen. If the host range of these variants is wide (e.g., raccoons, cats, dogs), and if they are antigenically divergent to the point of vaccine failure, the consequences of this 'reemergence' of CPV could be far-reaching.

This leads into the overall concept of this research: defining the mechanisms of virus evolution and the factors which facilitate the emergence of novel viral pathogens. There are three specific aims of this research. The first specific aim, 1) involves the isolation and characterization of novel viruses or variants which emerged via a unique genetic process: a) recombination (the alphaviruses, HJV and FMV), b) reassortment [the orbivirus, EHDV-6 (Indiana)], c) overprinting (the rhabdovirus, DURV), and d) mutation (the parvovirus, CPV). Depending upon the virus, 'characterization' may involve genetic, phylogenetic, antigenic and/or pathological analysis. The second specific aim, 2) was to compare and contrast the novel recombinants, reassortants, or variants with the parental (or progenitor) viruses from a genetic, evolutionary, and/or ecological standpoint. For the examples of recombination and reassortment, this involved genetic exchange from two parental viruses: recombination between EEEV and a SIN-like virus to form the ancestor of HJV, FMV, and WEEV, and reassortment between EHDV-6 and EHDV-2 to produce EHDV-6 (Indiana). For the examples of overprinting and mutation, this involved a genetic change within a single virus: point mutations in the capsid protein gene conferring an expansion of host range in CPV, and the use of an overlapping reading frame to form a new gene in DURV. In the case of DURV, the comparison (i.e.,

between the novel virus and its parental viruses) was reciprocated, as DURV appears to be a novel, but likely ancestral, virus from which the currently described vesiculoviruses (e.g., VSINV) originated from. The third and final aim, 3) was to elucidate the putative determinants of why/how each emerging virus is viable in nature and/or the potential biological, epidemiological and evolutionary implications of the emergence of the new recombinants, reassortants, or variants.

References

Barber, T. L., 1979. Temporal appearance, geographic distribution, and species of origin of bluetongue virus serotypes in the United States. Am. J. Vet. Res. 40, 1654-1656.

Bonaparte, M. I., Dimitrov, A. S., Bossart, K. N., Crameri, G., Mungall, B. A., Bishop, K. A., Choudhry, V., Dimitrov, D. S., Wang, L. F., Eaton, B. T., Broder, C. C., 2005. Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. Proc. Natl. Acad. Sci. U. S. A. 102, 10652-10657.

Borden, E. C., Shope, R. E., Murphy, F. A., 1971. Physicochemical and morphological relationships of some arthropod-borne viruses to bluetongue virus - a new taxonomic group. Physicochemical and serological studies. J. Gen. Virol. 13, 261-271.

Bossart, K. N., Wang, L. F., Flora, M. N., Chua, K. B., Lam, S. K., Eaton, B. T., Broder, C. C., 2002. Membrane fusion tropism and heterotypic functional activities of the Nipah virus and Hendra virus envelope glycoproteins. J. Virol. 22, 11186-11198.

Bowden, T. A., Crispin, M., Harvey, D. J., Aricescu, A. R., Grimes, J. M., Jones, E. Y., Stuart, D. I., 2008. Crystal structure and carbohydrate analysis of Nipah virus attachment glycoprotein: a template for antiviral and vaccine design. J. Virol. 82, 11628-11636.

Bréard, E., Sailleau, C., Hamblin, C., Graham, S. D., Gourreau, J. M., Zientara, S., 2004. Outbreak of epizootic haemorrhagic disease on the island of Réunion. Vet. Rec. 155, 422-423.

Brighton, S. W., Prozesky, O. W., de la Harpe, A. L., 1983. Chikungunya virus infection. A retrospective study of 107 cases. S. Afr. Med. J. 63, 313-315.

Chan, Y. P., Chua, K. B., Koh, C. L., Lim, M. E, Lam, S. K. 2001. Complete nucleotide sequences of Nipah virus isolates from Malaysia. J. Gen. Virol. 82, 2151-2155.

Chant, K., Chan, R., Smith, M., Dwyer, D. E., Kirkland, P., and the NSW Expert Group, 1998. Probable human infection with a newly described virus in the family Paramyxoviridae. Emerg. Infect. Dis. 4, 273-275.

Charrel, R. N., de Lamballerie, X., Raoult, D., 2007. Chikungunya outbreaks - the globalization of vector-borne diseases. N. Engl. J. Med. 356, 769-771.

Chua, K. B., Wang, L. F., Lam, S. K., Crameri, G., Yu, M., Wise, T., Boyle, D., Hyatt, A. D., Eaton, B.T., 2001. Tioman virus, a novel paramyxovirus isolated from fruit bats in Malaysia. Virology 283, 215-229.

Chua, K. B., Chua, B. H., Wang, C.W., 2002a. Anthropogenic deforestation, El Niño and the emergence of Nipah virus in Malaysia. Malays. J. Pathol. 24, 15-21.

Chua, K. B., Koh, C.L., Hooi, P.S., Wee, K.F., Khong, J.H., Chua, B.H., Chan, Y.P., Lim, M. E., Lam, S.K., 2002b. Isolation of Nipah virus from Malaysian Island flying-foxes. Microbes Infect. 4, 145-151.

Conraths, F. J., Gethmann, J. M., Staubach, C., Mettenleiter, T. C., Beer, M., Hoffmann, B., 2009. Epidemiology of bluetongue virus serotype 8, Germany. Emerg. Infect. Dis. 15, 433-435.

Dash, P. K., Parida, M. M., Santhosh, S. R., Verma, S. K., Tripathi, N. K., Ambuj, S., Saxena, P., Gupta, N., Chaudhary, M., Babu, J. P., Lakshmi, V., Mamidi, N., Subhalaxmi, M. V., Lakshmana Rao, P. V., Sekhar, K., 2007. East Central South African genotype as the causative agent in reemergence of Chikungunya outbreak in India. Vector Borne Zoonotic Dis. 7, 519-527.

Decaro, N., Desario, C., Elia, G., Martella, V., Mari, V., Lavazza, A., Nardi, M., Buonavoglia, C., 2008. Evidence for immunisation failure in vaccinated adult dogs infected with canine parvovirus type 2c. New Microbiol. 31, 125-130.

Decaro, N., Desario, C., Parisi, A., Martella, V., Lorusso, A., Miccolupo, A., Mari, V., Colaianni, M. L., Cavalli, A., Di Trani, L., Buonavoglia, C., 2009. Genetic analysis of canine parvovirus type 2c. Virology 385, 5-10.

Epstein, J. H., Field, H. E., Luby, S., Pulliam, J. R., Daszak, P., 2006. Nipah virus: impact, origins, and causes of emergence. Curr. Infect. Dis. Rep. 8, 59-65.

Ficken, M.D., Wages, D.P., Guy, J.S., Quinn, J.A., Emory, W.H., 1993. High mortality of domestic turkeys associated with Highlands J virus and eastern equine encephalitis virus infections. Avian Dis. 37, 585-590.

Field, H., Young, P., Yob, J. M., Mills, J., Hall, L., Mackenzie, J., 2001. The natural history of Hendra and Nipah viruses. Microbes Infect. 3, 307-314.

Fraser, G. C., Hooper, P. T., Lunt, R. A., Gould, A. R., Gleeson, L. J., Hyatt, A. D., Russell, G. M., Kattenbelt, J. A., 1996. Encephalitis caused by a Lyssavirus in fruit bats in Australia. Emerg. Infect. Dis. 2, 327-331.

Garmashova, N., Gorchakov, R., Volkova, E., Paessler, S., Frolova, E., Frolov, I., 2007. The Old World and New World alphaviruses use different virus-specific proteins for induction of transcriptional shutoff. J. Virol. 81, 2472-2484.

Gould, A. R., Hyatt, A. D., Lunt, R., Kattenbelt, J. A., Hengstberger, S., Blacksell, S. D., 1998. Characterisation of a novel lyssavirus isolated from a pteropid bat in Australia. Virus Res. 54, 165-187.

Guillaume, V., Wong, K. T., Looi, R. Y., Georges-Courbot, M. C., Barrot, L., Buckland, R., Wild, T. F., Horvat, B., 2009. Acute Hendra virus infection: Analysis of the pathogenesis and passive antibody protection in the hamster model. Virology 387, 459-465.

Guy, J. S., Ficken, M. D., Barnes, H. J., Wages, D. P., Smith, L. G., 1993. Experimental infection of young turkeys with eastern equine encephalitis virus and highlands J virus. Avian Dis. 37, 389-395.

Guy, J. S., Barnes, H. J., Smith, L. G., 1994. Experimental infection of young broiler chickens with eastern equine encephalitis virus and Highlands J virus. Avian Dis. 38, 572-582.

Halpin, K., Young, P. L., Field, H. E., Mackenzie, J. S. 2000. Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus. J. Gen. Virol. 81, 1927-1932.

Hoffman, B., Saberath, M., Thalheim, S., Bunzenthal, C., Strebelow, G., and Beer, M., 2008. Bluetongue virus serotype 8 reemergence in Germany, 2007 and 2008. Emerg. Inf. Dis. 14, 1421-1423.

Hooper, P., Zaki, S., Daniels, P., Middleton, D., 2001. Comparative pathology of the diseases caused by Hendra and Nipah viruses. Microbes Infect. 3, 315-322.

Hyatt, A. D., Zaki, S. R., Goldsmith, C. S., Wise, T.G., Hengstberger, S. G., 2001. Ultrastructure of Hendra virus and Nipah virus within cultured cells and host animals. Microbes Infect. 3, 297-306.

Jones, R. H., Foster, N. M., 1974. Oral infection of Culicoides variipennis with bluetongue virus: development of susceptible and resistant lines from a colony population. J. Med. Entomol. 11, 316-323.

Kam, Y. W., Ong, E. K., Rénia, L., Tong, J. C., Ng, L. F., 2009. Immunobiology of Chikungunya and implications for disease intervention. Microbes Infect. 11, 1186-1196.

Laras, K., Sukri, N. C., Larasati, R. P., Bangs, M. J., Kosim, R., Djauzi, S., Wandra, T., Master, J., Kosasih, H., Hartati, S., Beckett, C., Sedyaningsih, E. R., Beecham, H. J. 3rd, Corwin, A. L., 2005. Tracking the re-emergence of epidemic chikungunya virus in Indonesia. Trans. R. Soc. Trop. Med. Hyg. 99, 128-141.

McCall, B. J., Epstein, J. H., Neill, A. S., Heel, K., Field, H., Barrett, J., Smith, G. A., Selvey, L. A., Rodwell, B., Lunt, R., 2000. Potential exposure to Australian bat lyssavirus, Queensland, 1996-1999. Emerg. Infect. Dis. 6, 259-264.

Mertens, P. P. C., Maan, S., Samuel, A., Attoui, H. 2005. Genus Orbivirus, Family Reoviridae. p. 466-483 *In* Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., Ball, L. A. (eds.), Virus Taxonomy: Eighth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, California.

Metcalf, H. E., Pearson, J. E., Klingsporn, A. L., 1981. Bluetongue in cattle: a serologic survey of slaughter cattle in the United States. Am. J. Vet. Res. 42, 1057-1061.

Middleton, D. J., Morrissy, C. J., van der Heide, B. M., Russell, G. M., Braun, M. A., Westbury, H. A., Halpin, K., Daniels, P. W., 2007. Experimental Nipah virus infection in pteropid bats (Pteropus poliocephalus). J. Comp. Pathol. 136, 266-272.

Murray, K., Selleck, P., Hooper, P., Hyatt, A., Gould, A., Gleeson, L., Westbury, H., Hiley, L., Selvey, L., Rodwell, B., Ketterer, P., 1995. A morbillivirus that caused fatal disease in horses and humans. Science 268, 94-97.

Ozden, S., Huerre, M., Riviere, J. P., Coffey, L. L., Afonso, P. V., Mouly, V., de Monredon, J., Roger, J. C., El Amrani, M., Yvin, J. L., Jaffar, M. C., Frenkiel, M. P., Sourisseau, M., Schwartz, O., Butler-Browne, G., Desprès, P., Gessain, A., Ceccaldi, P. E., 2007. Human muscle satellite cells as targets of Chikungunya virus infection. PLoS One 2, e527.

Paredes, A., Alwell-Warda, K., Weaver, S. C., Chiu, W., Watowich, S. J., 2001. Venezuelan equine encephalomyelitis virus structure and its divergence from old world alphaviruses. J. Virol. 75, 9532-9537.

Parker, J. S., Murphy, W. J., Wang, D., O'Brien, S. J., Parrish, C. R., 2001. Canine and feline parvoviruses can use human or feline transferrin receptors to bind, enter, and infect cells. J. Virol. 75, 3896-3902.

Parrish, C. R., Have, P., Foreyt, W. J., Evermann, J. F., Senda, M., Carmichael, L. E., 1988. The global spread and replacement of canine parvovirus strains. J. Gen. Virol. 69, 1111-1116.

Pastorino, B., Muyembe-Tamfum, J. J., Bessaud, M., Tock, F., Tolou, H., Durand, J. P, Peyrefitte, C. N., 2004. Epidemic resurgence of Chikungunya virus in democratic Republic of the Congo: identification of a new central African strain. J. Med. Virol. 74, 277-282.

Philbey, A. W., Kirkland, P. D., Ross, A. D., Davis, R. J., Gleeson, A. B., Love, R. J., Daniels, P. W., Gould, A. R., Hyatt, A. D., 1998. An apparently new virus (family Paramyxoviridae) infectious for pigs, humans, and fruit bats. Emerg. Infect. Dis. 4, 269-271.

Powers, A. M., Brault, A. C., Shirako, Y., Strauss, E. G., Kang, W., Strauss, J. H., Weaver, S. C., 2001. Evolutionary relationships and systematics of the alphaviruses. J. Virol. 75, 10118-10131.

Rampal, S. M., Sharda, M., Meena, H., 2007. Neurological complications in Chikungunya fever. J. Assoc. Physicians India 55, 765-769.

Reiter, P., Fontenille, D., Paupy, C., 2006. Aedes albopictus as an epidemic vector of chikungunya virus: another emerging problem? Lancet Infect. Dis. 6, 463-464.

Renault, P., Solet, J. L., Sissoko, D., Balleydier, E., Larrieu, S., Filleul, L., Lassalle, C., Thiria, J., Rachou, E., de Valk, H., Ilef, D., Ledrans, M., Quatresous, I., Quenel, P., Pierre, V., 2007. A major epidemic of chikungunya virus infection on Reunion Island, France, 2005-2006. Am. J. Trop. Med. Hyg. 77, 727-731.

Ross, R. W., 1956. The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic. J. Hyg. (Lond.) 54, 177-191.

Ryman, K. D., Klimstram, W.B., 2008. Host responses to alphavirus infection. Immunol. Rev. 225, 27-45.

Sourisseau, M., Schilte, C., Casartelli, N., Trouillet, C., Guivel-Benhassine, F., Rudnicka, D., Sol-Foulon, N., Le Roux, K., Prevost, M. C., Fsihi, H., Frenkiel, M. P., Blanchet, F., Afonso, P. V., Ceccaldi, P. E., Ozden, S., Gessain, A., Schuffenecker, I., Verhasselt, B., Zamborlini, A., Saïb, A., Rey, F. A., Arenzana-Seisdedos, F., Desprès, P., Michault, A., Albert, M. L., Schwartz, O., 2007. Characterization of reemerging chikungunya virus. PLoS Pathog. 3, e89.

Tabachnick, W. J., 1996. Culicoides variipennis and bluetongue-virus epidemiology in the United States. Annu. Rev. Entomol. 41, 23-43.

Temizel, E. M., Yesilbag, K., Batten, C., Senturk, S., Maan, N. S., Mertens, P. P. C., Batmaz, H., 2009. Epizootic hemorrhagic disease in cattle, Western Turkey. Emerg. Infect. Dis. 15, 317-319.

Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E., Higgs, S., 2007. A single mutation in chikungunya virus affects vector specificity and epidemic potential. PLoS Pathog. 3, e201.

Wang, E., Volkova, E., Adams, A. P., Forrester, N., Xiao, S. Y., Frolov, I., Weaver, S. C., 2008. Chimeric alphavirus vaccine candidates for chikungunya. Vaccine 26, 5030-5039.

Wang, L. F., Yu, M., Hansson, E., Pritchard, L. I., Shiell, B., Michalski, W. P., Eaton, B. T., 2000. The exceptionally large genome of Hendra virus: support for creation of a new genus within the family Paramyxoviridae. J. Virol. 74, 9972-9979.

Williamson, M. M., Hooper, P. T., Selleck, P. W., Gleeson, L. J., Daniels, P. W., Westbury, H. A., Murray, P. K., 1998. Transmission studies of Hendra virus (equine morbillivirus) in fruit bats, horses and cats. Aust. Vet. J. 76, 813-818.

Williamson, M. M., Hooper, P. T., Selleck, P. W., Westbury, H. A., Slocombe, R. F., 2001. A guinea-pig model of Hendra virus encephalitis. J. Comp. Pathol. 124, 273-279.

Wong, K. T., Shieh, W. J., Zaki, S. R., Tan, C. T., 2002. Nipah virus infection, an emerging paramyxoviral zoonosis. Springer Semin. Immunopathol. 24, 215-228.

Wong, K. T., Grosjean, I., Brisson, C., Blanquier, B., Fevre-Montange, M., Bernard, A., Loth, P., Georges-Courbot, M. C., Chevallier, M., Akaoka, H., Marianneau, P., Lam, S. K., Wild, T. F., Deubel, V., 2003. A golden hamster model for human acute Nipah virus infection. Am. J. Pathol. 163, 2127-2137.

Xu, K., Rajashankar, K. R., Chan, Y. P., Himanen, J. P., Broder, C. C., Nikolov, D. B., 2008. Host cell recognition by the henipaviruses: crystal structures of the Nipah G attachment glycoprotein and its complex with ephrin-B3. Proc. Natl. Acad. Sci. U. S. A. 105, 9953-9958.

Yadin, H., Brenner, J., Bumbrov, V., Oved, Z., Stram, Y., Klement, E., Perl, S., Anthony, S., Maan, S., Batten, C., Mertens, P. P. C., 2008. Epizootic haemorrhagic disease virus type 7 infection in cattle in Israel. Vet. Rec. 162, 53-56.

Yaiw, K. C., Crameri, G., Wang, L., Chong, H. T., Chua, K. B., Tan, C. T., Goh, K. J., Shamala, D., Wong, K. T., 2007. Serological evidence of possible human infection with Tioman virus, a newly described paramyxovirus of bat origin. J. Infect. Dis. 196, 884-886.

Yob, J. M., Field, H., Rashdi, A. M., Morrissy, C., van der Heide, B., Rota, P., bin Adzhar, A., White, J., Daniels, P., Jamaluddin, A., Ksiazek, T., 2001. Nipah virus infection in bats (order Chiroptera) in peninsular Malaysia. Emerg. Infect. Dis. 7, 439-441.

Ziegler, S. A., Lu, L., da Rosa, A. P., Xiao, S. Y., Tesh, R. B., 2008. An animal model for studying the pathogenesis of chikungunya virus infection. Am. J. Trop. Med. Hyg. 79, 133-139.

CHAPTER 2

LITERATURE REVIEW

Virus evolution

In a broad sense, ‘virus evolution’ can be defined as a change in the genetic composition of a virus population over time (Domingo, 2007) and, thus, the emergence and evolution of novel viruses from pre-existing viruses is explicitly dependent upon genetic variation. Genetic variation in viral populations can occur by three primary mechanisms: mutation, recombination, and reassortment (DeFilippis and Villarreal, 2001; Domingo, 2007). Once genetic variation is introduced into individuals in a population by one or more of these mechanisms, natural selection and genetic drift will shape their evolution through non-random and random processes, respectively. Thus, although a multitude of genetic variants will arise by mutation, recombination, or reassortment, competition with each other under the selective pressures imposed by their environment (e.g., host immune response), in conjunction with chance events (e.g., blood meal from a mosquito), will ultimately lead to the survival and perpetuation of only a small fraction of these variants (Domingo, 2007). Although mutation, recombination, and reassortment are generally recognized as the primary mechanisms responsible for the present-day genetic diversity of viruses, an often overlooked and potentially underappreciated genetic mechanism of virus evolution is the generation of novel genes from overlapping reading

frames, a process known as ‘overprinting’ (Keese and Gibbs, 1992). In this research, the genetic, antigenic, phylogenetic, and/or pathological characterization of five viruses are described, each of which has emerged by the one of these four mechanisms (recombination, reassortment, mutation, or overprinting; see below for details).

Taxonomic and phylogenetic groupings of viruses are determined, in part, by the nature of their genomes. A classification scheme based on viral genomic structure, and how that structure influences mRNA synthesis, was first instituted by Baltimore (1971). In a generalized sense, RNA and DNA viral genomes can classified as either being double-stranded (ds) or single-stranded (ss), in addition to being either segmented or non-segmented (monopartite). Although genome polarity [i.e., negative-sense (-), positive-sense (+)] may be applied to the ss DNA viruses, it is most commonly used in reference to delineating between ss RNA viruses whose genomes are in message-sense (+), and are thus directly infectious, and those genomes that are complementary to message (-) and therefore must first transcribe their mRNAs from the (-) genome (i.e., by the virion-associated RNA polymerase).

The viruses in this research represent a broad range of genomic configurations, including both RNA and DNA genomes, ds and ss genomes, and (-) and (+) RNA. In some instances, the type of nucleic acid and its structure and/or polarity may preferentially dictate which genetic mechanism may be most influential in shaping the evolution of the virus. For the research presented here, each mechanism and the name, taxonomic status, and genomic organization of the viruses studied, along with the host(s) from which each virus was obtained, were as follows:

1) Recombination: a) *Highlands J virus* (HJV), genus *Alphavirus*, family *Togaviridae*, single-stranded (+) RNA. Hosts: barred owl (*Strix varia*), red-tailed hawk (*Buteo jamaicensis*), blue jay (*Cyanocitta cristata*). Location: Georgia, Florida.

b) *Fort Morgan virus* (FMV), genus *Alphavirus*, family *Togaviridae*, single-stranded (+) RNA. Host: swallow bug (*Oeciacus vicarius*). Location: Colorado.

2) Reassortment: *Epizootic hemorrhagic disease virus* (EHDV) serotype 6, strain Indiana [EHDV-6 (Indiana)], genus *Orbivirus*, family *Reoviridae*, double-stranded segmented RNA. Host: white-tailed deer (*Odocoileus virginianus*). Location: Indiana, Illinois, Missouri, Kansas, Texas, Michigan.

3) Overprinting: *Durham virus* (DURV), genus *Vesiculovirus*, family *Rhabdoviridae*, single-stranded (-) RNA. Host: American coot (*Fulica americana*). Location: North Carolina.

4) Mutation: *Canine parvovirus* (CPV), genus *Parvovirus*, family *Parvoviridae*, single-stranded, linear DNA. Host: raccoon (*Procyon lotor*). Location: Georgia, Virginia, Tennessee, Kentucky, Florida, Wisconsin.

As mutation, in its simplest terms, involves a change in the base composition of a single (ribo)nucleotide, it is the primary genetic mechanism that is most commonly observed in all viruses. However, although mutation is a ubiquitous mechanism of evolution, mutation rates for RNA viruses far exceed those observed in DNA viruses, primarily as a result of the functional disparities between viral RNA polymerases and viral DNA or cellular polymerases (discussed below) (Steinhauer *et al*., 1992). As such, mutation is an intrinsic component of the evolutionary biology of RNA viruses, but is a

relatively minor driver in DNA virus evolution. Recombination, whether it is through the breakage and rejoining of DNA or by RNA template switching, like mutation, is common to all types of viruses; however, unlike mutation, there is a proclivity for recombination as a major source of evolution in both RNA [primarily ss (+)] and DNA viruses (Block *et al*., 1985; Worobey and Holmes, 1999; Esposito *et al*., 2006). Reassortment, as a prerequisite, is limited to viruses which have a segmented genome. Since genome segmentation is relatively uncommon among ss (+) RNA, ss DNA, and ds DNA viruses, reassortment as a major evolutionary mechanism is confined primarily to segmented viruses which have either a ds RNA or ss (-) RNA genome (e.g., see Webster *et al*., 1992; Mertens, 1999). Superficially, retroviruses may also undergo 'reassortment', as their (+) ssRNA genomes are diploid, and thus co-infection can result in heterozygous progeny containing a single copy of the genome from each parent. However, by definition, the retroviral genome, although diploid, is considered monopartite (non-segmented) (Murphy *et al*., 1999). Although each of the four mechanisms may operate individually to shape the evolutionary pathway of a virus, they do not necessarily have to function exclusively from one another. For example, as retroviruses are diploid, template switching and the lack of proofreading by reverse transcriptase during replication, along with the potential for strand transfer during co-infection, can theoretically promote simultaneous recombination, mutation, and heterodimer formation ('reassortment'), respectively.

Recombination

Recombination involves the exchange of nucleotide sequences between two different RNA or DNA molecules, such that the progenitor chimeric molecule contains

sequences derived from both parents. There are two proposed fundamental mechanisms by which recombination may occur in viruses: 1) a breakage-and-rejoining mechanism and 2) a copy-choice or template-switching mechanism (Lai, 1992; Nagy and Simon, 1997; Bujarski, 2008). The breakage-and-rejoining mechanism is the predominant form of recombination observed in DNA viruses, whereby two ds DNA fragments are paired together and religated by a number of enzyme-driven reactions (Ball, 1987; Umene, 1999; Bujarski, 2008); in theory, this mechanism has also been applied to both ds and ss RNA molecules (Bujarski and Dzianott, 1991). However, whether RNA-RNA recombination via a breakage-and-rejoining mechanism occurs naturally is controversial.

Breakage-and-rejoining mechanisms that could potentially operate for viral RNA molecules include cellular enzyme-catalyzed models such as RNase T1-dependent ligation, as observed for viroid transcripts (Tsagris *et al*., 1991), and site-specific cleavage and religation of exogenous RNAs, as seen with group II introns (ribozymes) (Morl and Schmelzer, 1990). Previously, recombination has been observed between replicating viral RNA and transfected RNA segments of the coronavirus, *mouse hepatitis virus* (MHV) (Liao and Lai, 1992), as well as recombination between two non-replicative RNA fragments of the alphavirus, *Sindbis virus* (SINV) (Raju *et al*., 1995). However, in both of these cases, recombination was dependent upon the presence of the cognate RNA-dependent RNA polymerase, suggesting that recombination of non-replicative RNA fragments was due to template switching (see below) by the polymerase. However, non-replicative recombination of RNA in the absense of any enzyme activity has been suggested to account for nonhomologous recombination in *Qβ bacteriophage* (Chetverin *et al*., 1997) and poliovirus (Gmyl *et al*., 1999) through a transesterification process

mediated by secondary structures in the two RNA fragments. This type of non-replicative RNA-RNA recombination would potentially suggest that viral RNAs are intrinsically recombinogenic (Chetverin *et al*., 1997). Although there is limited evidence to suggest that non-replicative RNA-RNA ligation occurs readily in natural systems (Jarvis and Kirkegaard, 1992; Nagy and Simon, 1997), if it does, it could have profound effects (e.g., ligation of discordant viral RNA genomes or viral and host RNA) on the evolution of RNA viruses (Gmyl *et al*., 2003).

The second proposed mechanism by which RNA recombination occurs is by a copy-choice or template-switching method, which is widely accepted as the primary mechanism by which recombination occurs in RNA viruses (Lai, 1992; Kim and Kao, 2001). The template-switching mechanism involves the RNA-dependent RNA polymerase pausing at some point during RNA synthesis on the template (donor) molecule, dissociating or 'jumping' from the donor and then binding to a new (acceptor) template and proceeding with RNA synthesis. It is believed that transcription by RNA-dependent RNA polymerases is inherently discontinuous and most pausing is initiated by regions of strong secondary structure (e.g., stem loops) or reiterative sequences (e.g., AU-rich regions) (Nagy and Simon, 1997). It is this inherent characteristic of pausing that likely facilitates the release of the polymerase, along with the nascent RNA strand, from the template. The detection of incomplete transcripts in infected cells and the sequential molar abundance of transcripts relative to the promoter site (transcriptional attenuation) support the notion of pausing and subsequent detachment of the RNA polymerase (Iverson and Rose, 1981). Although speculative, this may also suggest that the processivity of a specific viral RNA polymerase, and possibly the degree of secondary

structure in the specific viral RNA, could potentially dictate how prone a given virus is to recombination.

RNA recombination, mediated by template-switching, can be classified into three types: 1) homologous (or legitimate), 2) aberrant homologous, or 3) nonhomologous (or illegitimate) (Lai, 1992; Worobey and Holmes, 1999), although alternate versions (e.g., similarity-based recombination types) have also been proposed (Nagy and Simon, 1997). Homologous recombination involves an exchange of genetic elements of two nearly identical or closely related RNA molecules with extensive sequence homology. Crossovers generally occur at sites that are very closely or perfectly matched between the two RNA molecules, such that the recombinant RNA molecule retains a very similar or the exact sequence and structural organization of the parental RNA molecules. The alphaviruses presented in this research, HJV and FMV, arose by homologous recombination between EEEV and a SIN-like virus, such that the genomic organization and open reading frames (ORF) of the parental viruses were not compromised. Aberrant homologous recombination also involves similar or closely related RNA molecules, but crossovers occur at unrelated rather than homologous sites, although these sites are usually in close proximity to one another in the general organization of the two parental viruses (Lai, 1992). A virus that has arisen by aberrant homologous recombination may contain a deletion or insertion relative to the parental genomes. Nonhomologous recombination occurs between two RNA molecules that do not show any sequence homology. Although the basis for the selection of the crossover site between nonhomologous sequences is unclear, it has been proposed that similarity in the secondary structures of the two RNA molecules may play a role (Lai, 1992).

Nonhomologous recombination may result in the acquisition of completely unrelated (e.g., nonviral) genes into the viral genome, as witnessed by the insertion of cellular ubiquitin sequences in the genome of *bovine viral diarrhea virus* (BVDV) (Becher *et al*., 1999).

Although recombination can occur in all types of RNA viruses, it has been documented most often in those that have ss (+) genomes, such as picornaviruses (Simmonds, 2006), coronaviruses (Jackwood *et al*., 2010), flaviviruses (Twiddy and Holmes, 2003), caliciviruses (Forrester *et al*., 2008), and astroviruses (Pantin-Jackwood *et al*., 2006). In comparison, recombination has been reported in very few ss (-) RNA viruses (Sibold *et al*., 1999; Chare *et al*., 2003; McCarthy and Goodman, 2010). The reasons for this discrepancy is not completely clear, but it has been proposed that since the genome and antigenomes of ss (-) RNA viruses are bound by the viral nucleoprotein throughout the replication cycle, this prevents recombination (template switching) from occurring at a high frequency (Chare *et al*., 1993). However, paradoxically, defective-interfering particles (which arise by template-switching recombination) are a common occurrence during (-) RNA virus infections *in vitro* (Lazzarini *et al*., 1981), which suggests that recombination in (-) RNA viruses can occur readily. Additionally, it has been suggested that as many (-) RNA viral infections are characterized as being acute with short recovery periods, this limits the potential for superinfection and subsequent recombination due to the low probability that hosts are multiply infected at the same time (Chare *et al*., 1993); however, these properties (acute infections, short recovery periods) are also characteristic of many (+) RNA viruses, in which recombination appears to be relatively common (Twiddy and Holmes, 2003). Similar to that of the ss (-) RNA

viruses, recombination (in the strict sense) has not been historically regarded as an important mechanism of genetic variation in ds RNA viruses. However, the recent recognition of recombination in reoviruses (Phan *et al*., 2007) and birnaviruses (Hon *et al*., 2008) suggests that recombination plays a role in the evolution of all RNA viruses, albeit if at differential levels. Currently, the reason(s) for the apparent discrepancies in recombination rates among different RNA viruses [i.e., ss (+), ss (-), ds] remains speculative and requires further research.

The first alphavirus that was demonstrated to be derived from recombination was *Western equine encephalitis virus* (WEEV) (Hahn *et al*., 1988). Genetic analysis revealed that WEEV was descended from an ancestral recombinant derived from two other alphavirus species, *eastern equine encephalitis virus* (EEEV) and a *Sindbis* (SIN)-like virus. Not only was a putative cross-over point (i.e., the point at which the polymerase switched templates) identified within the coding region of an envelope protein gene (E3), but it was suggested that another cross-over occurred within the 3' UTR of the ancestral recombinant (Hahn *et al*., 1988). Thus, this recombination event is a unique study in virus evolution for a number of reasons: 1) it apparently required not one, but two cross-over events, 2) it occurred between two different alphaviruses (EEEV and a SIN-like virus), not variants or genotypes of the same species, and 3) it resulted in the evolution and subsequent establishment of three novel viruses (i.e., the ancestral recombinant radiated to produce WEEV, HJV, and FMV).

The recognition that alphaviruses could undergo reassortment in nature and produce viable recombinant viruses with novel genetic, antigenic, and pathological properties prompted a number of *in vitro* studies examining the genetic determinants of

generating a successful recombinant (i.e., between two different alphavirus species) (Lopez *et al*., 1994; Kuhn *et al*., 1996; Yao *et al*., 1996; Yao *et al*., 1998; Kim *et al*., 2000). These recombination studies were primarily aimed at analyzing the replicative abilities of different chimeric progeny derived from constructs developed between SINV and *Ross River virus* (RRV), two alphaviruses sharing 64 and 48% amino acid identity in their nonstructural and structural proteins, respectively (Kuhn *et al*., 1996). Generation of two reciprocal chimeras in which the nonstructural and structural proteins were switched (i.e., SINV-ns/RRV-s and RRV-ns/SINV-s) demonstrated that both recombinants were capable of replicating in mice and in various cell lines (Kuhn *et al*., 1996). However, for both chimeras, the *in vitro* and *in vivo* growth patterns along with the amount of RNA synthesized was attenuated relative to the parental viruses. This result suggested that although the structural and nonstructural proteins from two different viruses could interact to form a viable replicating virus, their interactions (e.g., protein-protein or protein-RNA) were suboptimal. In other words, although recombination could result in a functional virus, in all likelihood, these chimeras would not survive under natural conditions. This notion is reiterated by the recognition that although recombination in alphaviruses appears to occur readily under laboratory conditions as demonstrated by the formation of defective-interfering particles (Stollar, 1979), it is a rare phenomenon in nature, suggesting that most recombinants that do arise are quickly removed from the gene pool due to their reduced fitness.

Additional experiments aimed at determining the fate of chimeras in which more subtle changes were introduced were also instructive. A chimeric virus in which only the capsid protein of SINV was introduced into the backbone of RRV resulted in a virus that

grew to a titer of 10^{-5} of that of the parental RRV, and attempts at passing the recombinant in cell culture resulted in a loss of the virus (Lopez *et al*., 1994). The failure of the recombinant was demonstrated not be a defect in replication or packaging, but rather to be dependent upon the physical interaction of the SINV capsid with the RRV glycoproteins during the budding process. Analysis of the C-terminal cytoplasmic domain of the E2 (i.e., the portion of E2 that physically interacts with capsid) between RRV and SINV disclosed 11 amino acid differences; by changing 9 of these amino acids in the RRV E2 glycoprotein to the cognate SINV amino acids by site-directed mutagenesis, the severely attenuated chimeric virus grew 10^4-fold more efficiently. This result may suggest that, if a recombinant virus that is debilitated can mutate in order to allow its chimeric proteins to adapt to one another prior to being extinguished, it may be able to increase its fitness level significantly. The real-world emergence of WEEV may be similar in this scenario. Examination of the WEEV capsid (derived from EEEV) revealed that seven amino acids had changed from the EEEV sequence to that of the cognate SINV residues (Strauss, 1993). Additionally, analysis of the C-terminal domain of E2 in WEEV (derived from SINV) demonstrated that four amino acids had changed from the SINV residue to the EEEV sequence (Strauss, 1993). Thus, after recombination, selection and adaptation of the interacting amino acids of the capsid and E2 proteins in WEEV likely increased the stability of the virus during the budding process, which may have facilitated its survival in nature.

Another instructive aspect of WEEV is that besides the major internal cross-over occurring within E3, an additional cross-over is suspected to have occurred within the distal region of the 3' UTR (Hahn *et al*., 1988), although some authors suggest that this

non-coding cross-over did not occur (Netolitzky *et al*., 2000). An important aspect of replication in alphaviruses is that a number of *cis*-acting structural elements that reside within the 3' and 5' noncoding regions are well known to regulate transcription and replication. The 3' UTR in alphaviruses contains a number of conserved and repeat sequence elements, including a 19-nucleotide stretch directly upstream of the poly(A) tail that acts as the promoter for negative-strand synthesis (Hardy and Rice, 2005). Whether the non-coding template switch occurred at the same time as the E3 cross-over event is unknown; however, it has been postulated that in order for the virus to persist in nature, it may have been necessary to retain the terminal 3' UTR region (from EEEV) in order for the polymerase complex (also derived from EEEV) to replicate efficiently (Strauss and Strauss, 1997). In other words, it apparently took an extraordinary set of circumstances (i.e., two recombinations), potentially mediated by short-term selection (i.e., convergent mutations in E2 and the capsid) and possibly chance (initial survival prior to adaptation), in order for the ancestral alphavirus (from which WEEV, HJV and FMV evolved) to become established in nature.

The extent that recombination has shaped the evolution of the ss (+) RNA viruses, although apparently limited when examining viruses in the *Alphavirus* genus alone, may be far-reaching. Comparative analysis of the alphavirus genome with other ss (+) RNA viruses disclosed that the alphaviruses shared many common features and sequence homologies with a number of plant viruses (Haseloff *et al*., 1984; Ahlquist *et al*., 1985). The remarkable recognition of these shared features led to the formation of the alpha-like supergroup of viruses, which included, along with the alphaviruses, many ss (+) RNA plant viruses such as viruses in the genera *Tobamovirus*, *Tobravirus*, and *Hordeivirus*

(Strauss and Strauss, 1994). It was demonstrated that homologous domains and functions (i.e., nsP1 capping enzyme, nsP2 helicase, and nsP4 polymerase) were not only present in alphaviruses and ss (+) RNA plant viruses, but that these domains were also organized in similar locations in the genome (Rozanov *et al*., 1992; Strauss and Strauss, 1994). Furthermore, both groups were shown to have common replicative functions and sequence elements, in that they both transcribed a subgenomic RNA to express their structural proteins and contained an opal or amber termination codon, respectively, that permitted translational readthrough of the nonstructural polyprotein to produce the polymerase (Strauss *et al*., 1993; Strauss and Strauss, 1994). Although these similarities could possibly imply convergent evolution due to common functionalities, it is generally perceived that they are the result of a common origin. At the same time, many of the genes (i.e., glycoproteins) of alphaviruses are unique and do not have homologs in the plant viruses of the alpha-like supergroup (Strauss and Strauss, 1994). This may suggest that the ancestral alphavirus originated as a plant virus, only to acquire additional genes (e.g., glycoproteins) via recombination from an unknown source. Alternatively, the alphaviruses may have originated as insect viruses, which subsequently recombined with a plant virus at some point during their evolution; however, the chain of events leading to the evolution of the present-day alphaviruses is unknown. However, it appears likely that all ss (+) RNA viruses, whether they are plant, animal, or insect viruses, originated from a common ancestor, with their subsequent divergence mediated, in part, by recombination (Koonin and Dolja, 1993; Weaver, 1995). The genus *Alphavirus* currently contains 29 species, many of which are important human and veterinary pathogens. As a group, they have a worldwide distribution. Thus, although recombination among present-day

alphaviruses is limited to WEEV, HJV, and FMV, it appears that recombination likely played a role in the emergence and subsequent spread and establishment of the genus as a whole.

Reassortment

Reassortment is defined as the exchange of complete RNA or DNA segments between two (or more) viruses co-infecting the same cell. The progeny from this reassortment, termed the reassortant, is therefore composed of RNA or DNA segments derived from different parental viruses. Since reassortment involves the exchange of genetic material between parental viruses, although by horizontal gene transfer (i.e., without template switching), it is sometimes referred to as a form of recombination, or alternatively as 'pseudo-recombination' (Gibbs and Keese, 1995). However, for the purposes of this research, the terms 'recombination' and 'reassortment' relate to discrete mechanisms and are not interchangeable. Since reassortment involves the exchange of complete gene segments, only viruses which have segmented genomes are capable of reassortment. One demarcation of the concept of a viral species is the ability, or inability, to reassort with members of the same, or different, species, respectively; hence, unlike recombination, reassortment can not occur between closely related, but distinct, species (see Brown *et al*., 1988; Nuttal and Moss, 1989). In the case of segmented viruses that infect vertebrates, reassortment has been described extensively for viruses with both ss (-) RNA [e.g., arenaviruses (Charrel *et al*., 2003), bunyaviruses (Cheng *et al*., 1999), orthomyxoviruses (Webster *et al*., 1992)] and ds RNA [e.g., birnaviruses (Hon *et al*., 2006), reoviruses (Iturriza-Gómara *et al*., 2001)] genomes. On the other hand, genome

segmentation in vertebrate ss (+) RNA viruses is rare and is limited to a single family (i.e., *Nodaviridae*) (Olverira *et al*., 2009). Similarly, genome segmentation is only observed in three families of DNA viruses [*Geminiviridae* (Surendranath *et al*., 2005), *Nanoviridae* (Hu *et al*., 2006), and *Polydnaviridae* (Kroemer and Webb, 2004)], none of which infect vertebrates. Consequently, reassortment, in contrast to recombination, does not appear to be an important mechanism in the evolution of DNA viruses.

Although reassortment to produce phenotypic variants appears to occur readily *in vitro*, there are only a few examples of the isolation of a novel reassortant in nature that clearly demonstrates that the new segment configuration confers some beneficial characteristic for viral fitness not previously observed in other reassortants or non-reassortants (e.g., increased vector competency, expanded host range, antigenic diversity). This lack of recognition among field isolates may be due, in part, to 1) the inherent difficulty in detecting most reassortants (e.g., segments are closely related and thus reassortants are not easily identified), 2) the fact that reassortment occurs only between viruses of the same species (i.e., similar viruses, for the most part, have similar genotypes/phenotypes), or that 3) most new reassortment configurations are neutral or deleterious (i.e., phenotypes remain unchanged or RNA and/or protein interactions are incompatible and, thus, the virus does not survive in nature).

The genetic analysis of novel reassortants derived under natural conditions and the subsequent epidemiological consequences of such reassortment events has been studied most extensively with (avian) *influenza A virus* (AIV) (genus *Influenzavirus A*, family *Orthomyxoviridae*). The concept of antigenic shift (i.e., reassortment involving an exchange of a segment(s) encoding a surface antigen) is most often associated with

AIV, whereby a new subtype is generated through the exchange of the segments containing the hemagglutinin (HA) and/or neuraminidase (NA) genes (Webster *et al*., 1992). Whether the 1918 H1N1 pandemic strain of AIV, which killed an estimated 20-50 million people worldwide, was a reassortant (Gibbs and Gibbs, 2006), or was an non-reassortant avian virus that adapted to humans (Taubenberger *et al*., 2005; Tumpey *et al*., 2005), is a matter of debate. Nevertheless, it is unequivocally accepted that the two subsequent pandemic strains of 1957 and 1968 were reassortants derived from parental avian and human viruses (Kawaoka *et al*., 1989; Belshe, 2005). Although reassortment may not be an absolute prerequisite for the development of a pandemic strain, as suggested by Taubenberger *et al*. (2005) for the 1918 H1N1 virus, in addition to the 1947 pandemic-like outbreak of H1N1 resulting from antigenic drift (Reid and Taubenberger, 2003), reassortment can allow a virus to acquire large beneficial fitness advantages instantaneously, mediating the rescue of a virus with deleterious mutations or, in the case of the pandemic strains of 1957 and 1968, the emergence of a new, possibly highly fit, antigenic variant into an immunologically naïve population. As it has been estimated that a present-day pandemic on a severity scale of the 1918 H1N1 virus could result in 62 million deaths worldwide (Murray *et al*., 2006), and that the next pandemic of AIV has been suggested to be not a question of if, but when (Webby and Webster, 2003), understanding the dynamics of reassortment and how reassortment can effect fitness and transmission of RNA viruses is of obvious concern.

An interesting aspect of the 1957 and 1968 pandemic AIV strains is the recognition that, in both cases, RNA segments encoding novel avian HA and PB1 genes were present in the new reassortant (Kawaoka *et al*., 1989). Was this just coincidental or

was their a functional reason? Besides chance, possible reasons that the HA and PB1 gene segments reassorted together in both pandemic strains was that 1) the two genes interact with each other in some way, such that if one segment was transferred without the other, it would lead to a decrease in fitness, resulting in less transmission and the ultimate removal of the virus from the population, or 2) both novel segments retained necessary, but independent, features (e.g., immune evasion due to a novel HA and increased replication and/or pathogenesis due to a novel PB1) that together (in association with the novel NA obtained in 1957), resulted in a pathogenic virus that spread efficiently in a human population without protective immunity, and thus initiated a pandemic (Kawaoka *et al*., 1989; Conenello *et al*., 2007). As direct functional interactions between HA (a surface antigen) and PB1 (a polymerase component) are not required (i.e., at least not at the protein-protein level), it is possible that the reason for the association of HA and PB1 in pandemic strains may have been the synergistic phenotype conferred by their dual reassortment. Other reassortants that may have emerged at the same time, but without these novel tandem segments, simply may not have had pandemic potential (e.g., antigenic novelty coupled to increased replication). In the case of the reassortant EHDV presented in this research, two RNA segments, which encode the two outer capsid proteins (which would be structurally analogous to HA and NA of AIV), were derived from the exotic parental virus. Was this by chance or was there a functional reason for the incorporation of both of these segments? As discussed later, it appears that both segments may have reassorted together due to mandatory physical interactions between the two proteins in order to form a structurally viable virus.

As aforementioned, the reassortant described in this study is EHDV, which is a *Culicoides* midge-borne member of the genus *Orbivirus*, family *Reoviridae*. Reoviruses are characterized as having a ds segmented RNA genome, which depending upon the genus, can be composed of 10-12 segments (e.g., orbiviruses have 10 segments) (Mertens *et al*., 2005). If two different serotypes of EHDV (each containing 10 segments of dsRNA) were to infect a single cell, the number of potential different configurations that could emerge from such a co-infection would be 2^{10}, or 1,024. In theory, if this were a completely random process, 1,022 of the progeny would be reassortants and the remaining two would be the parental viruses (Nibert *et al*., 1996). However, this estimation would only be valid if, besides being totally random, the concentration of each segment for each virus was present at identical molar quantities during packaging, which, in all likelihood does not occur (Urquidi and Bishop, 1992). Nevertheless, an important observation from this theoretical estimation is that, although the potential to create an enormous diversity of different reassortment configurations is possible, some sort of mechanism(s) must be in place that inhibits the unrestricted reassortment of co-infecting genomes. Extensive experimental analysis of many reoviruses suggests that this is the case, as although reassortment in reoviruses has been demonstrated to occur at a high frequency (e.g., 42% in *Culicoides* midges coinfected with BTV-10 and BTV-17) (Samal *et al*., 1987), it does not approach levels that would be suggested if it were a completely non-random and non-restrictive process. If mechanisms exist to restrict random reassortment of segments between two viruses co-infecting the same cell, what are they and how do they operate? Although much is still to be learned regarding the process of reassortment in segmented RNA viruses, the main barrier to unrestricted reassortment is

the inability of RNA segments (and/or their products) from one parent to structurally and functional interact with segments from a different parent, such that the viability of the reassortant is not compromised. This barrier can exist at a 1) RNA-RNA level (e.g., two RNAs must recognize one another in order to be packaged), 2) RNA-protein level (e.g., a polymerase must recognize a RNA in order to bind to it and initiate transcription/replication), or 3) protein/protein level (e.g., two structural proteins must functionally interact to form a stable virion).

In order to understand how RNA-RNA or RNA-protein interactions may potentially affect reassortment, it is instructive to look at how genomes are normally assorted in segmented viruses. The assortment process by which a virus normally packages its RNA involves two major obstacles: 1) recognizing viral RNA segments from all the other RNAs in the cell, and 2) specifically packaging each of the appropriate viral segments into the virion. With segmented ss (-) RNA viruses like AIV, since the RNA is heavily encapsidated by the nucleoprotein, it has been historically proposed that, although the assortment of genomic viral RNA is specific, the assortment of each specific genome segment into a virus particle is random (Jolik and Roner, 1995). Random assortment, by convention, implies that virions will contain a range in both the number and configuration of segments; for AIV, the estimated low infectious to total particle ratio (0.01-0.1 or ~1-10% of progeny virions are infectious), along with observation of virions containing more than eight segments, has been used as evidence that random assortment is the mechanism by which AIV segments are packaged (Lamb and Choppin, 1983; Jolik and Roner, 1995). At the other extreme, a highly-ordered RNA segment-specific assortment has been demonstration in some ds RNA viruses. In the case of bacteriophage

φ6, three RNA segments (S, M, L), have been shown to be sequentially packaged (S→M→L) due to the temporal exposure, and then masking, of segment-specific binding sites on the preassembled procapsid (Mindich, 1999). As each segment must be packaged in sequential order, bacteriophage φ6 has an estimated infectious to total particle ratio close to 1 (Mindich, 1999). Does packaging efficiency actually differ that drastically (i.e., theoretically from 1-100%) among different segmented RNA viruses? If so, how does this affect the frequency of reassortment? Are segmented viruses with strict packaging requirements, like bacteriophage φ6, less likely to undergo reassortment? Are the 'loose' packaging requirements of AIV responsible for the establishment of pandemic viruses?

Whether each segment of the AIV genome is incorporated at random (segment-nonspecific) or that each of the eight segments are incorporated specifically has been a matter of much debate (Bancroft and Parslow, 2002; Liang *et al*., 2005; Muramoto *et al*., 2006; Gog *et al*., 2007; Hutchinson *et al*., 2010). If each of the segments of AIV and other (-) ss RNA viruses are, in fact, randomly incorporated as has been suggested (Bancroft and Parslow, 2002), this might imply that, although it would appear to an extremely inefficient process, the potential for generating novel reassortants may be greater than those viruses that require a stricter segment-specific assortment of genomes. However, recent advances in determining the signals required for genome packaging in AIV have been made, as the generic packaging signal (i.e., viral-specific, segment non-specific) for AIV has been demonstrated to be located within the terminal promoter regions at the 3' and 5' ends of each segment (Tchatalbachev *et al*., 2001). The co-localization of important structural and sequence elements (i.e., promoter and packaging

signals) in ss (-) RNA viruses, as opposed to the localization of such signals to discrete genomic regions as in (+) RNA viruses [e.g., alphaviruses (Frolova *et al*., 1997)], may be a reflection of the limited access to the genomic RNA due to the bound nucleocapsid. Moreover, it has recently been suggested that both the viral-specific and segment-specific packaging signals for AIV are located adjacent to one another, as the segment-specific packaging signals, although apparently discontinuous, appear to stretch from the 5' and 3' UTRs into the coding region of each segment (Hutchinson *et al*., 2010). Exactly how each of the eight segments then recognize each other and form a complex that is incorporated into a budding virion remains unclear.

Although speculative, one potential consequence of an overlap of important replicative and packaging signals (such as in AIV) is that these sequences may be more conserved among different strains, thereby facilitating reassortment more so than viruses which have clearly distinct promoter and packaging signals, which may inhibit reassortment between genetically variable serotypes or strains. Determining how different viruses recognize and package their RNA may disclose a differential proclivity to reassort, not only among different segmented RNA viruses [e.g., ss (-) RNA versus ds RNA], but also the degree to which packaging signals may dictate how efficiently, for instance, different human and avian strains of AIV may reassort with one another.

Although reovirus genomes are composed of ds RNA, the structure of the nucleic acid that is packaged into the immature core particle is ss RNA, whereby in association with the bound RNA polymerase within the core particle, complementary (-) strands are synthesized to produce the ds molecules (Patton and Spencer, 2000). Since packaging of ss RNA precedes ds RNA synthesis, it can be inferred that the packaging signals for

reoviruses are on the mRNAs rather than the ds RNA molecules. Similar to AIV, defining the exact signals and mechanisms of segment packaging in reoviruses has been difficult. However, contrary to AIV, owing to the recognition that the genome is composed of equimolar quantities of each segment and that the infectious to total particle ratio is high (>0.5) (Jolik and Roner, 1995), the incorporation of each individual segment into the reovirion has never been questioned to be a random process. What makes reovirus packaging apparently more efficient than that postulated with AIV is unknown, but it may be related to the fact that, even though reovirus ss RNAs are initially bound by many nonstructural and structural proteins (see below), they are not fully encapsidated as ribonucleoproteins as are ss (-) RNAs and therefore may have more 'naked' sequence available for more complex RNA-RNA recognition signals. Indeed, it has been postulated that protein binding to reoviral mRNAs may induce steric changes that expose signal sequences in each mRNA needed to recognize each another (Patton and Spencer, 2000).

After cell entry and destabilization of the outer capsid, reovirus mRNA transcripts that are released from the core particle into the cytoplasm which, if not translated, are bound by viral nonstructural RNA-binding proteins, and then later, by the RNA polymerase and capping enzyme to form assortment complexes (Taraporewala *et al*., 2004). Association of the RNA-binding proteins with the viral RNA is not sequence-specific, and although the polymerase recognizes conserved sequences in association with the 5' cap on each segment, none of the viral proteins which bind to the mRNA selectively bind to a specific RNA segment (Patton and Spencer, 2000). Thus, during a co-infection of two serotypes, binding of viral proteins to a heterologous mRNA should

not be theoretically inhibited by sequence variation, as long as the polymerase recognizes the conserved terminal elements common of capped viral mRNAs (Tao *et al*., 2002; Lu *et al*., 2008), which are well conserved among different serotypes (Mertens *et al*., 2005).

Although the exact process by which reovirus segments are co-packaged has yet to be fully elucidated (Roner and Steele, 2007a; Lu *et al*., 2008), if RNA-RNA interactions dictate the assembly of the 10-segmental complex in orbiviruses, could RNA segment-specific packaging sites inhibit the ability of two reoviruses to reassort, if they were genetically divergent (as observed among different serotypes)? For mammalian orthoreoviruses, the molecular basis of segment selection during reassortment has been reported to be dictated by sequences contained in the 5' termini of each of the ss RNA segments, such that these sequences may help maintain ('identify') different serotypes (Roner and Steele, 2007b). In other words, as alluded to earlier with AIV, the ability of different strains and serotypes of a given reovirus, to even begin the reassortment process, many depend upon, in part, how closely related their segment-specific packaging signals are.

As the EHDV reassortant reported in this research contained two RNA segments derived from one parent, and the remaining eight from another, it must be assumed that the packaging signals between the two serotypes were not prohibitive to reassortment (i.e., there was RNA-RNA compatibility). That the two exotic RNA segments identified in the reassortant encoded the outer capsid proteins, a question that arised was whether there were certain structural or functional requirements between these two segments beyond the RNA-RNA level, such that co-reassortment was dependent upon protein compatibilities between the segments. Previously, the non-random segregation of

specific inter-parental pairings of segments has been demonstrated in co-infection studies using the serotype 1 Lang and serotype 3 Dearing strains of *mammalian reovirus* (MRV) (Nibert *et al*., 1996). Electropherotype analysis of 83 reassortants disclosed that certain gene segments, which encoded proteins that were known to functionally interact, were significantly underrepresented in terms of inter-parental pairings. However, the observation that other underrepresented heterologous pairings that did not involve known interacting proteins also suggested that additional non-protein interactions (i.e., RNA-protein or RNA-RNA), as discussed earlier, may also dictate whether certain segments can reassort.

In a generalized sense, these observations collectively imply that the ability of two serotypes or strains to reassort likely decreases as genetic diversity between the two viruses increases. The point at which genetic disparity functionally inhibits RNA-RNA, RNA-protein, and protein-protein interactions is impossible to gauge for any given virus and ultimately depends upon the specific nucleotide or amino acid interactions in question. However, in order to compensate for genetic differences between two different parental viruses, it has been demonstrated that (similar to the adaptation of the E2 and capsid interactions in recombinant alphaviruses), reassortants have been commonly shown to contain one or more mutations in their ‘interacting’ segments relative to the original parent (Joklik and Roner, 1995; Roner *et al*., 1995; Nibert *et al*., 1996). These findings suggest that in order to compensate for incompatibility between two segments, whether it is at the RNA or protein level, mutations may occur that allow the reassortant to gain fitness increases (i.e., allow for proper interactions and, hence, increase function) that are needed in order to survive. Thus, initial incompatibility issues derived from

reassortment, similar to recombination, may be overcome by subsequent adaptive mutations.

Previous analysis of BTV reassortants, similar to that described with MRV, demonstrated that certain inter-parental pairings of gene segments, mainly segment 2 (VP2; major outer capsid protein) and segment 6 (VP5; minor outer capsid protein) were underrepresented relative to homologous pairings (Mertens, 1999). The reason for this apparent under representation was suggested to rely on the fact that both VP2 and VP5 functionally interact to collectively form the outer capsid of the virus. As such, replacing either VP2 or VP5 with a heterologous substitute from another serotype may result in a virus that is structurally unstable and/or functionally impaired. When heterologous VP2-VP5 pairings were observed, they were oftentimes associated with reduced rates of replication, suggesting that they would likely not persist in nature (Mertens, 1999). The reassortment of both of the outer proteins (VP2 and VP5) of EHDV-6 into an EHDV-2 background detailed in this research likely relies on this principle of protein-protein compatibility.

The pandemic strains of AIV are reminders of how a novel reassortant can affect transmission, and subsequently the pathogenicity of a virus. In other viruses that undergo reassortment, whether the transfer of one or two gene segments from one phenotypic variant (e.g., virulent) to that of a different phenotype (e.g., avirulent) results in a concomitant transfer of the phenotype is largely unknown and likely depends upon the gene segment(s) in question. Reassortment has been described for many different reoviruses, including those that are important human and veterinary pathogens such as *rotavirus A* (SiRV-A), which causes 870,000 human deaths annually in the developing

world (Gouvea and Brantly, 1995) and *bluetongue virus* (BTV), which has killed more than 2,000,000 animals since it introduction into Europe in 1998 (Batten *et al*., 2008). For both SiRV-A and BTV, a major step in attempting to prevent the spread of these reoviruses has been the advent of massive vaccination programs. One unintended consequence of vaccination has been the recognition of reassortment between field and vaccine strains of both SiRV-A and BTV (Gouvea and Brantly, 1995; Batten *et al*., 2008), leading to the potential to create reassortants with novel phenotypic (e.g., more virulent) characteristics that are also well-adapted to replication in the host. Similar to AIV, understanding how reassortment can alter the phenotypic characteristics of different reoviruses is of concern.

In terms of reassortment and its effects on pathogenicity in orbiviruses, the most comprehensive *in vitro* studies have been performed with serotypes of *Great Island virus* (GIV), a serogroup of tick-borne viruses that are associated with seabirds found in polar regions (e.g., the prototype GIV is maintained in an *Ixodes* tick-Atlantic puffin cycle) (Moss *et al*., 1988; Nuttal *et al*., 1989; Nuttal *et al*., 1990; Nuttal *et al*., 1992). In a study by Nuttal *et al*. (1989), *in vitro* reassortment experiments were performed using three serotypes of GIV: *Wexford virus* (WEXV), *Nugget virus* (NUGV), and the prototype GIV. Survival times of mice infected with the parental viruses disclosed that while NUGV and GIV were neurovirulent, WEXV had an attenuated phenotype in terms of pathogenicity. Thus, it was surmised that viruses derived from either NUGV or GIV reassorting with WEXV would provide ideal candidates for deducing the genetic determinants of neurovirulence. Different reassortants, in which segments of NUGV or GIV were intermixed with WEXV, disclosed that the pathogenic phenotype could only

be transferred when the reassortant contained segment 4 from NUGV or GIV. In the case of the NUGV-WEXV reassortants, although 90% of the parental segments of WEXV remained unchanged, reassortment of segment 4 alone from NUGV conferred the virulent phenotype of the minority parent. Although at the time these experiments were performed, the protein encoded of segment 4 in the GIV serotypes was unknown, cryoelectron reconstruction of the three dimensional structure of *Broadhaven virus* (BRDV), another serotype of GIV, demonstrated that segment 4 encoded the major outer capsid protein VP4 (analogous to VP2 in EHDV), and that VP4 was likely responsible for cell receptor recognition (Schoehn *et al*., 1997). Thus, the neurovirulent phenotype of GIV or NUGV could apparently be transferred to WEXV solely through the reassortment of the RNA segment encoding the major outer capsid protein.

Alteration in the pathogenicity of novel reassortants (i.e., relative to the majority parental) as demonstrated in orbiviruses, has also been observed in other vector-borne RNA viruses, such as bunyaviruses. Viruses in the family *Bunyaviridae* are composed of three ss (-) RNA segments [denoted as large (L), medium (M) and small (S)] and include a number of animal and human pathogens found in the United States, including *La Crosse virus* (LACV) and *Jamestown Canyon virus* (JCV), two strains of *California encephalitis virus* (CEV). Although JCV has been implicated in a single human case of encephalitis (Grimstad *et al*., 1982), LACV is more of an immediate public health concern, as it is recognized as the leading cause of pediatric arboviral encephalitis in the United States (Gerhardt *et al*., 2001). Reassortment studies between LACV and JCV by co-infecting *Ae. albopictus*, an Asian mosquito species that was introduced into the United States in 1985 (Sprenger and Wuithiranyagool, 1986) and also the vector species

responsible for the CHIKV outbreaks mentioned in the *Introduction* section, demonstrated that reassortment between the two viruses occurred readily, as all six potential configurations were isolated (Cheng *et al*., 1999). All reassortants were demonstrated to be infectious to *Ae. albopictus* and could be transmitted to, and were neurovirulent in, suckling mice. However, in comparison to the parental JCV or other reassortment configurations of JCV-LACV, reassortants containing the M segment (encoding the glycoproteins responsible for host cell receptor binding) of LACV, but the remaining L and S segments from JCV, were shown to more neuroinvasive than other reassortants and could infect gerbils and *Ochlerotatus* (*Oc*.) *triseriatus* (the normal vector of LACV), both characteristics of LACV, but not JCV. As such, acquirement of the glycoprotein of LACV into a JCV background not only allowed the reassortant to gain a more virulent phenotype that was characteristic of LACV, but also to obtain a new host range (i.e., gerbils and *Oc. triseriatus*), of which the parental JCV was incapable of.

The fact that the reassortment of the gene segment encoding a protein involved in antigenicity and/or host cell binding in orthomyxoviruses (HA), orbiviruses (VP4), and bunyaviruses (G) has been shown to lead to an increase in transmission, pathogenicity, and/or host range is of direct interest to the EHDV reassortant in this research, as the two segments obtained from exotic parental EHDV-6 encode the surface proteins (VP2 and VP5) responsible for antigenicity and mammalian receptor binding. What the long-term epidemiological impacts of this introduction will be, if any, remains to be seen.

Overprinting

A unifying characteristic of all RNA virus genomes is their compact size. Most RNA viruses have genomes in the range of ~5 to 15kB, with the *beluga whale coronavirus* SW1 genome, the largest known viral RNA genome, reaching ~31.7kB (Mihindukulasuriya *et al*., 2008). In comparison, the largest known viral DNA genome belongs to that of *Acanthamoeba polyphaga mimivirus*, which is ~1.2 million base pairs in length (Claverie *et al*., 2006), and is thus ~37 times the size of *beluga whale coronavirus* SW1. The compact size of RNA genomes is believed to be a consequence of the error-prone replication of the RNA-dependent RNA polymerase, an enzyme unique to RNA viruses. In theory, as RNA genome size increases, so does the number of ribonucleotides misincorporated into the growing strand. As a consequence, the high mutation rate places an upper limit on the size of the viral RNA genome due to the process of 'error catastrophe', which theorizes that there is a threshold level of mutations that a RNA genome can incorporate before it becomes non-viable (Eigen, 1971; Holmes, 2003b). Accordingly, experimental evidence has demonstrated that many RNA viruses replicate just below the error threshold, and that mutagenically inducing an increase in the mutation rate, even if only by one or two-fold, can result in virus extinction (Crotty *et al*., 2001).

In theory, one consequence of the compact size of the RNA genome imposed by the error threshold may be that many viral genes encode proteins that are multifunctional. For example, the nsP2 protein of the alphaviruses (discussed in *Chapter 3*, *part II*) has NTPase, RNA helicase, RNA triphosphatase, and protease activity, as well as numerous non-replicative functions. In other words, a single gene encodes a single protein that

performs multiple biological functions. However, there is a point at which a single open reading frame (ORF) of a gene is maximized in terms of the biological functions it can encode. Due to the compact size of viral RNA genomes, it would therefore appear that just a few genes must be responsible for encoding all the functions needed for the virus to replicate and subvert cellular responses in order to survive, and that acquiring novel genetic elements that may expand the biological and functional repertoire of the virus must be gained either through processes such as recombination or reassortment. However, one way RNA viruses can maximize the genetic potential of a small genome is to encode two proteins (which are functionally divergent) from the same gene, but in alternate reading frames by a process known as 'overprinting'. The first documentation of the use of overprinting in a virus was demonstrated by Barrell *et al*. (1976) in the bacteriophage φX174, whereby two proteins of different function (protein D, a scaffolding protein and protein E, a lysis protein) were shown to be translated from the same exact gene. Since then, additional overlapping genes have been described in bacteriophage φX174 and other members of the family *Microviridae* (Tessman *et al*., 1980; Gillam *et al*., 1985; Pavesi, 2006).

Computational and experimental analysis has divulged that the use of overprinting as a mechanism of genetic variation in viruses is widely evident, such that overprinted genes have been found in ~50% of all viral genera (Gibbs and Keese, 1995) and in more than 120 RNA viruses in which complete genomic sequences are available (Holmes, 2003). However, often times, identifying overlapping genes may not be inherently obvious, as PB1-F2, the 11th AIV gene product that is expressed from an overlapping reading frame of PB1, was only identified in 2001 despite the fact that the genetics of

AIV had been studied extensively for decades (Chen *et al*., 2001). As PB1-F2 was characterized as a virulence factor that contributed to a high pathogenicity phenotype, it was theorized that its incorporation into a newly reassorted virus may allow it to replicate more efficiently in a new host and, thus, facilitate its transmission (Conenello *et al*., 2007). In lieu of this theory and the demonstrated ability to induce apoptosis in immune cells, it was additionally suggested that PB1-F2 may have been a factor in precipitating the emergence and high lethality of pandemic AIV strains (Conenello *et al*., 2007). This reiterates the notion that the genetic mechanisms discussed in this research may not be mutually exclusive and they may, in fact, operate in tandem and synergistically in the emergence of novel pathogens.

For overlapping genes, as two proteins are encoded in the same exact nucleotide sequence, but in alternate reading frames, mutations in one frame, although not affecting the coding sequence of one protein, may be detrimental to the second. For example, a third base codon transition, although resulting in a silent synonymous amino acid substitution in one reading frame (e.g., *CTT*AT→*CTC*AT = L→L), would encode a nonsynonymous amino acid substitution in either of the two alternate reading frames (e.g., C*TTA*T→C*TCA*T = L→S or CT*TAT*→CT*CAT* = Y→H). Therefore, there must be constraints imposed on overlapping genes (i.e., that are not in place for monocistronic genes) that maintain the integrity of both gene sequences simultaneously. For ORFs that encode a single protein, synonymous nucleotide substitutions theoretically accumulate at a uniform rate as they are independent of protein function and consequently are considered neutral (Fujii *et al*., 2001), although they could affect RNA secondary structure. However, in overlapping genes, if both of the genes in alternate reading frames

are to be conserved, this would theoretically suggest that a much lower rate of synonymous nucleotide substitutions should be observed relative to the remainder of the genome (Pavesi, 2000).

As nonsynonymous nucleotide changes alter the amino acid residue of the protein, their substitution rate should be confined by the functionality of the protein, such that amino acid residues integral for protein activity and stability are conserved through selection (Fujii *et al*., 2001). As such, the ratio of nonsynonymous to synonymous nucleotide substitutions (d_n/d_s) can be used to measure of the degree of evolutionary constraint and selection being place on a gene, with d_n/d_s values >1 being indicative of positive selection and d_n/d_s values <1 implying negative, or purifying, selection (Zhao *et al*., 2007). In a monocistronic gene in which protein function needs to be highly conserved (e.g., a polymerase), it would be expected to have a low d_n and a uniform (i.e., higher) d_s and thus likely have a d_n/d_s <1, suggesting that purifying selection is acting to conserve the protein. In an overlapping gene, whether only one or both proteins are being conserved can potentially be indicated by comparative d_n/d_s values between the distinct gene regions of overlap (ancestral and derived).

In each case where overprinting is observed, a fundamental problem is to elucidate which gene is ancestral and which gene is novel. One way of determining ancestral from derived genes is to analyze codon usage in the overlapping versus non-overlapping genes in a particular genome, such that the overlapping gene which utilizes the same or most similar codon usage as the rest of the (monocistronic) genome is believed to be ancestral (Pavesi, 2006). An alternate method for defining ancestral and derived genes has been to determine the extent to which the overlapping gene in question

is observed among species belonging to the same family (Rancurel *et al*., 2009). The gene that is present in most, if not all, related viruses (e.g., common to all genera within a family) is generally considered to be the ancestral gene, while the gene which is more restricted phylogenetically (e.g., found only in members of a single genus) is believed to be the derived gene. For example, the overprinted gene of the rhabdovirus described in this research [*Durham virus* (DURV)] encodes two putative proteins, the phosphoprotein (P) and the C protein. The P protein is found in all viruses within the family *Rhabdoviridae*, while the C protein has only been reported from a single genus of rhabdoviruses, the *Vesiculovirus* genus. Therefore, it can be inferred that, with DURV, the P gene is ancestral while the C gene is derived.

The primary function of the rhabdovirus P protein is to associate with the L protein (polymerase) and the ribonucleoprotein [nucleoprotein (N) bound to the RNA] to initiate transcription; indeed, without the P protein, transcription does not occur (Emerson and Yu, 1975). Hence, although the P protein has no enzymatic function of its own, it can be referred to as an 'essential' protein, as it is required in order for the virus to survive (i.e., polymerase activity). On the other hand, the function of the rhabdovirus C protein is unknown, but it has been speculated to enhance transcriptional activity or to be involved in host pathogenicity or insect transmission (Kretzschmar *et al*., 1996; Peluso *et al*., 1996). A common theme of most novel overprinted proteins (i.e., in which the function has been determined) is that they appear to have an accessory role in the viral life cycle; that is, they are neither structural nor enzymatic, but rather provide some secondary function that enhances the survival and/or transmission of the virus [e.g., inhibit apoptosis (Hong *et al*., 2002) or IFN signaling (Shaffer *et al*., 2003); suppress

RNA interference (Scholthof, 2006); promote cell-to-cell movement (Krishnamurthy *et al*., 2003)]. Since derived proteins are believed to be important, but nevertheless dispensable, for replication and/or survival of the virus, as has been previously shown for the C protein of the rhabdovirus, *Vesicular stomatitis Indiana virus* (VSINV) (Kretzschmar *et al*., 1996), they are normally characterized as 'non-essential' proteins. Although the fortuitous *de novo* synthesis of a protein from a second overlapping reading frame in a pre-existing gene essentially implies it that it must have an accessory role, it is possible that these proteins may evolve over time to become an 'essential' component of the virus life cycle (Gibbs and Keese, 1995). Although paramyxovirus C proteins, similar to the C proteins of rhabdoviruses, are also categorized as non-essential accessory proteins, mutants of the paramyxovirus *Sendai virus* (SeV) in which the C protein was silenced were shown to be severely attenuated in mice, suggesting that, rather than being non-essential, the C protein was indispensable for *in vivo* replication and pathogenesis (Kurotani *et al*., 1998). Interestingly, evolutionary analysis by comparative d_n/d_s of the overlapping P and C genes in both rhabdoviruses and paramyxoviruses indicated that, in the rhabdoviruses, negative selection was acting to conserve the P region, but not the C gene sequence, while in paramyxoviruses, the P region was evolving rapidly relative to the C gene (Jordan *et al*., 2000; Fujii *et al*., 2001). Thus, both molecular and experimental analysis suggest that the C proteins encoded within the P gene sequences in paramyxoviruses and rhabdoviruses may be evolving under different constraints, potentially becoming essential (non-dispensable) and non-essential (dispensable) proteins, respectively (Kretzschmar *et al*., 1996; Kurotani *et al*., 1998; Escoffier *et al*., 1999; Jordan *et al*., 2000).

As having two genes encoded in the same sequence places constraints on both genes, such that neither may become optimally adapted, one method of alleviating such constraints is through the duplication of the overlapping sequences (Keese and Gibbs, 1992). Gene duplication of overlapping genes in eukaryotes is a common occurrence and is believed to alleviate such constraints, whereas the genomes of RNA viruses are more likely to maintain overlapping genes due to size restrictions (Keese and Gibbs, 1992). However, it is of interest that the family *Rhabdoviridae*, as a whole, appears to be unique in that many recently characterized viruses, in which their full genomes have been sequenced, has revealed a genomic organization that departs from the standard prototypical rhabdovirus genome in that they include a number of additional novel ORFs (e.g., see Kurath *et al*., 1985; Springfeld *et al*., 2005; Gubala *et al*., 2008). Three things are important about this recognition: 1) contrary to the notion that the size of RNA genomes are constrained due the their mutation rates and the error threshold, rhabdovirus genomes appear to be quite amendable to increases in size, as the incorporation of these additional ORFs attests to; 2) in some rhabdoviruses, the novel ORF appears to be a result of gene duplication (e.g., ephemeroviral G_{NS} proteins are believed to be gene duplications of the upstream G protein) (Walker *et al*., 1992; Wang and Walker, 1993), and 3) the derived overlapping gene described in this study shares homology and identity to a different gene of a related rhabdovirus (i.e., DURV C shows identity to VSINV G). Thus, one hypothetical scenario by which a virus may alleviate the functional constraint of overprinting is by autologous gene duplication or possibly by intertypic recombination. Whether many viral genes (e.g., those that to do not have any apparent common ancestry) arose by gene duplication or intermolecular recombination of a derived overprinted gene

is unknown, but such a scenario could theoretically play a role in the molecular diversity of viruses observed today.

It is also of interest that not all viruses may have the same capabilities to encode additional proteins from a single gene sequence. As aforementioned, the alphavirus nsP2 protein is a prototypical example of a multifunctional viral protein, exemplifying the concept of maximizing the biological function out of a short sequence of RNA. The alphavirus nonstructural proteins, including nsP2, are initially translated as a single polyprotein, whereby each individual protein is then produced by post-translational cleavage. For rhabdoviruses, each protein is translated from an individual transcript of each gene, such that read-through of the initial start codon in the P transcript to the next encountered methionine codon is all that is needed in order to produce the overlapping protein. For alphaviruses, translation of all nonstructural proteins is dependent upon initiation from a single start codon in nsP1 and protein formation is entirely dependent upon site-specific post-translational cleavage. Thus, the multi-functionality of viral proteins recognized among many different RNA viruses may be due, in some cases, to the fact that the mechanism by which they are produced may be prohibitive to expressing proteins from a secondary ORF.

Mutation

A mutation is an inheritable change in the sequence of a nucleotide(s) that can occur through a number of different mechanisms such as intercalation by chemical mutagens, enzyme-mediated deamination, or polymerization of an incorrect nucleotide (Domingo, 2007). Mutations that involve a single nucleotide change (point mutations)

can be grouped into either transitions, where similar structured nitrogen-containing bases are substituted (e.g., one type of purine for another), or transversions, where different bases are switched (e.g., purine for a pyrimidine). Mutational transitions and transversions can be further grouped in silent, missense, or nonsense mutations depending upon if the mutation results in a synonymous or nonsynonymous amino acid change, or a stop codon, respectively. Besides point mutations, single or multiple base insertions or deletions (collectively referred to as indels) are commonly encountered mutations which, if not in triplicate form, can alter the reading frame of the gene and thus may have profound effects in they occur in a coding region.

In the context of RNA virus evolution, mutational analysis is most often associated with measuring point mutation rates or frequencies that are inherently dependent upon by the low-level fidelity of the RNA-dependent polymerases used in copying the genome. Owing to the lack of a 3' to 5' exonuclease activity, RNA-dependent RNA polymerases and RNA-dependent DNA polymerases (i.e., reverse transcriptases) cannot remove misincorporated ribonucleotides once they are inserted into a growing strand (Steinhauer *et al*., 1992). In contrast, DNA-dependent RNA or DNA polymerases (i.e., enzymes found in DNA viruses, prokaryotes, and eukaryotes) have a 3' to 5' proofreading-repair activity and thus can remove such misincorporations (Abbotts and Loeb, 1985; Abbotts *et al*., 1987). An evolutionary consequence of this difference in proofreading function is that RNA viruses have an extremely high mutation rate which can exceed that of DNA viruses by several orders of magnitude. As RNA virus mutation rates vary from 10^{-3} to 10^{-5} per nucleotide per round of replication, an RNA virus with a genome of 10kB is likely to incorporate one or more mutations every time its genome is

copied (Domingo and Holland, 1997). Coupled to the high mutation rate, many RNA viruses replicate to extremely high levels in their hosts, resulting in very large population sizes relative to DNA viruses (Holland *et al*., 1982).

Consequential to this error-prone, high-level replication is that any given RNA virus exists as a heterogeneous and dynamic population of mutant genomes, commonly referred to as a 'quasispecies' (Eigen and Schuster, 1977; Domingo, 2007). A key concept of the quasispecies nature of RNA viruses is that the existence of such a large distribution of mutants (e.g., 10^{12} virions may be produced per host/day) confers great adaptability to selection pressures that may be encountered (Domingo, 1998). For an RNA virus of 10kB, the total sequence space available (i.e., the theoretical representation of all the possible variations of its sequence) is $4^{10,000}$, an astronomical number (Domingo, 2007). Although only a infinitesimal fraction of the available sequence space of a given virus will ever be 'transversed' due to functional constraints, the biological and epidemiological implications of the mutational adaptability conferred by the RNA quasispecies to disease prevention and control are numerous: e.g., 1) reversion of attenuated viruses to virulence, which may hamper vaccine development; 2) selection of antibody or inhibitor escape mutants, which may result in disease progression, even in the presence of virus-specific antibody or antiviral agents; or 3) selection of mutants with increased replicative capacity, altered cell tropism, or expanded host range which can lead to the emergence of 'novel' viruses (Duarte *et al*., 1994; Domingo, 2003). It is this last implication, mutations leading to a host range expansion, which is the basis for this research. However, contrary to conventional wisdom, the virus described in this study, which evolves at a rate that is characteristic of some RNA viruses (1×10^{-4}

substitutions/site/yr) (Shackelton *et al*., 2005), is in fact *canine parvovirus* (CPV), a ss linear DNA virus.

The reason(s) for the high rate of evolution of CPV is unclear. For DNA viruses which utilize their own DNA polymerases, such as bacteriophage T4, mutation rates can be as low as 2×10^{-8} per base pair (Drake, 1991). If bacteriophage T4 were to utilize its host cell (*E. coli*) DNA polymerase, mutation rates could decrease even further, to 10^{-10} misincorporations per nucleotide, or approximately one mutation for every 10 billion nucleotides incorporated (Domingo, 2007). The CPV genome is small, being ~5.3kb, and encodes only four proteins, two nonstructural (NS1 and NS2) and two structural (VP1 and VP2); none of these proteins have polymerase activity (Berns and Parrish, 2007). As such, CPV is inherently dependent upon cellular DNA polymerases in order for its replication. However, as stated above, mutation rates in CPV far exceed those that would be expected with the fidelity of cellular DNA polymerases. One possible explanation is the ss nature of the CPV genome and it mechanism for replication. Parvoviruses replicate by a unidirectional, strand-displacement, rolling-hairpin mechanism, which in *murine minute virus* (MVM), is presumed to be mediated by DNA polymerase δ and its accessory proteins (Cotmore and Tattersall, 2006). A similar type of replication strategy (rolling-circle replication) is also observed in circular ss DNA viruses that infect animals (i.e., *Circoviridae*), as well as those that infect plants (i.e., *Geminiviridae*) and bacteria (i.e., *Microviridae*) (Lefeuvre *et al*., 2009). In addition to CPV, high mutation rates have been documented in other ss DNA viruses such as circovirus SENV (7.32×10^{-4} nt/site/yr) (Umemura *et al*., 2002) and in the plant geminivirus *Tomato yellow leaf curl virus* (2.88×10^{-4} nt/site/yr) (Duffy and Holmes,

2007), suggesting that high mutation rates are the norm for ss DNA viruses (Duffy *et al*., 2008), irrespective of whether the polymerase being used is of animal or plant origin. Thus, the nature of the replication strategy, rather than the host cell polymerase, may account for the high mutation rate observed in CPV and other ss DNA viruses (Lopez-Bueno *et al*., 2006). Owing in part to the rolling circle mechanism of replication, recombination appears to be a common mechanism of genetic variation among ss DNA viruses (Lefeuvre *et al*., 2009), suggesting that recombination in CPV could partially contribute to an apparent high rate of mutation. However, recombination has been demonstrated not to play a role in the observed mutation rates of CPV (Shackelton *et al*., 2005).

Additionally, nucleotide substitution rates for a single virus species have been shown to fluctuate depending on the degree of replication that is occurring. For *human T-cell lymphotropic virus* types 1 and 2 (HTLV-1/2), mutation rates were demonstrated to be 1000-fold lower during times of restricted replication (i.e., latent, endemic transmission) than during times of rapid replication (i.e., epidemic transmission) in the same host species (Van Dooren *et al*., 2001; Holmes, 2003a; Duffy *et al*., 2008). By analogy, the intrinsically high rate of mutation in CPV and other ss DNA viruses could also be a consequence of their high replication rate (Duffy *et al*., 2008). Experimental infections of dogs with CPV has demonstrated that infections are acute and may be associated with high viral loads (Azetaka *et al*., 1981; Truyen and Parrish, 1992), suggesting CPV may normally replicate to high levels *in vivo* without inducing a persistent carrier state, qualities more consistent with that of an RNA virus than many DNA viruses.

Another possible mechanism for the rapid evolution of CPV is that estimates are derived at points during host switching, when it would be expected to observe a high rate of evolution. However, the estimated basal mutation rate of CPV is also much higher relative to what would be expected in most DNA viruses (Shackelton *et al*., 2005). Additionally, analysis of the surface-exposed VP2 protein and the nonstructural NS1 protein showed relatively equal rates of mutation, suggesting that although positive selection for neutralization-escape mutants or host range variants in VP2 may drive the high mutation rate to some extent, it does not explain the analogous rate in NS1 (Shackelton *et al*., 2005). Therefore, it was suggested that for CPV, positive selection coupled to an intrinsically high rate of mutation, in the absense of detectable recombination, is likely responsible for its rapid evolution (Shackelton *et al*., 2005). Although the underlying factor(s) governing the high mutation rate are yet to be resolved, the biological consequences of many such mutations are well documented, as the amino acid changes that alter host range (i.e., transferrin receptor binding) and antigenicity in CPV have been extensively mapped (Hueffer *et al*., 2003; Palermo *et al*., 2003; Palmero *et al*., 2006; Govindasamy *et al*., 2003). Since a vast amount of information relating to the genetics of the carnivore parvoviruses and their associated variants is available, CPV provides a well-established and unique model system for examining and potentially determining mutations that are responsible for species jumps. Interestingly, for the parvoviruses, large-scale epidemiological shifts (i.e., the acquirement or loss of host range) can be obtained through very minor genetic changes that, in some instances, may involve a single mutation.

The mutation of a single nucleotide, although seemingly insignificant, can potentially be lethal to a virus, or at the other end of the spectrum, can have far-reaching beneficial biological effects, which can unfortunately be coupled with extremely adverse epidemiological consequences in certain cases. For the overprinted PB1-F2 gene of AIV, a single mutation (N66S) observed in the 1997 Hong Kong H5N1 viruses, as well as in the reconstituted 1918 H1N1 pandemic strain, was not only demonstrated to increase pathogenicity *in vivo*, but it was additionally suggested that this mutation may have been a factor relating to the high lethality observed during the 1918-1919 pandemic (Conenello *et al*., 2007). Although most mutations of functional importance are normally associated with nonsynonymous amino acid changes, point mutations in non-coding regions can also have dramatic effects on the biological properties of a virus. Single nucleotide changes in the 5' UTR of the oral polio (live, attenuated) vaccine, that were shown to be selected upon passage in the human gut, were associated with a reversion to the wild-type neurovirulent phenotype and induced vaccine-associated paralytic poliomyelitis in a number of cases (Evans *et al*., 1985; Guillot *et al*., 1994; Svitkin *et al*., 1990). The effects of single nucleotide substitutions on fitness have also been analyzed on a broad scale. In a study by Sanjuán *et al*. (2004), 91 different single mutations were introduced into a clone of VSINV by site-directed mutagenesis. Overall, the majority of mutations were deleterious (i.e., they reduced fitness), with up to 40% of random mutations being lethal, suggesting that most random mutations are removed by negative selection. However, as large populations sizes are oftentimes coupled to the high mutation rate, a beneficial mutation that is buried among a multitude of neutral and deleterious mutations, may be selected under the right circumstances.

In the case of the emergence of CHIKV on Reunion Island in 2006 (as mentioned in the *Introduction* section), a single mutation conferring an A226V substitution in the E1 glycoprotein allowed the virus to infect a new vector species not previously incriminated in CHIKV transmission. Subsequently, the newly acquired vector species could efficiently transmit the virus to a population without protective immunity, resulting in an epidemiological crisis. Similarly, for CPV, mutations in the protein responsible for cellular receptor binding (VP2) have also been demonstrated to have far reaching epidemiological consequences. The progenitor virus of CPV was presumably *feline panleukopenia virus* (FPV) or a variant of FPV that may have 'jumped' into another carnivore species (e.g., fox) prior to its transmission and subsequent adaptation to dogs (Truyen, 2006). Although this new variant (known as CPV-2) could infect and replicate in dogs, its presumable parent, FPV, could not. The host range discrepancy between FPV and CPV-2 was later deduced by demonstrating that by altering two amino acids in the VP2 protein of FPV (K93N and D323N) allowed the virus to bind to and infect canine cells (Hueffer *et al*., 2003). These two residues were not only demonstrated to be responsible for CPV-2 binding to the canine transferrin receptor, but were also were responsible for antigenic (i.e., monoclonal antibody binding) and pH-dependent hemagglutination differences observed between FPV and CPV (Parrish, 1999), therefore affording biological differentiation of the two viruses prior to the advent of genetic testing. An additional three other amino acid substitutions (A300G, D305Y, and M87Y), which were also demonstrated to be in contact with the transferrin receptor and thus involved in controlling canine host range, became fixed in the population as new CPV variants (CPV-2a/b) became widespread, replacing the original CPV (CPV-2) (Truyen *et*

al., 1994). An additional consequence of these mutations was the ability of CPV-2a/b to infect cats, a property lost in CPV-2. Although CPV was first recognized as a pathogen in 1978, it has been suggested that it may have first emerged up to 10 years before it was first described, possibly circulating in the dog population while acquiring beneficial mutations that increased its infectivity, and hence transmissibility, allowing it to spread globally (Shackelton *et al*., 2005). As such, the cases represented in this research may present a similar potential scenario, whereby CPV has recently jumped into a novel host and appears to evolving rapidly; whether adaptation to its novel host results in the evolution of a new species-specific (raccoon) variant, or a novel multi-species (e.g., raccoon, cat, and dog) pathogen that is antigenically and genetically distinct, remains to be determined.

References

Abbotts, J., Loeb, L. A., 1985. DNA polymerase alpha and models for proofreading. Nucleic Acids Res. 13, 261-274.

Abbotts, J., Nishiyama, Y., Yoshida, S., Loeb, L. A., 1987. On the fidelity of DNA replication: herpes DNA polymerase and its associated exonuclease. Nucleic Acids Res. 15, 1185-1198.

Ahlquist, P., Strauss, E. G., Rice, C. M., Strauss, J. H., Haseloff, J., Zimmern, D., 1985. Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. J. Virol. 53, 536-542.

Azetaka, M., Hirasawa, T., Konishi, S., Ogata, M., 1981. Studies on canine parvovirus isolation, experimental infection and serologic survey. Jpn. J. Vet. Sci. 43, 243-255.

Ball, L. A., 1987. High-frequency homologous recombination in vaccinia virus DNA. J. Virol. 61, 1788-1795.

Baltimore, D., 1971. Expression of animal virus genomes. Bacteriol. Rev. 35, 235-241.

Bancroft, C. T., Parslow, T. G., 2002. Evidence for segment-nonspecific packaging of the influenza A virus genome. J. Virol. 76, 7133-7139.

Barrell, B. G., Air, G. M., Hutchison III, C. A., 1976. Overlapping genes in bacteriophage φX174. Nature 264, 34-41.

Batten, C. A., Maan, S., Shaw, A. E., Maan, N. S., Mertens, P. P. C., 2008. A European field strain of bluetongue virus derived from two parental vaccine strains by genome segment reassortment. Virus Res. 137, 56-63.

Becher, P., Orlich, M., König, M., Thiel, H. J., 1999. Nonhomologous RNA recombination in bovine viral diarrhea virus: molecular characterization of a variety of subgenomic RNAs isolated during an outbreak of fatal mucosal disease. J. Virol. 73, 5646-5653.

Belshe, R. B., 2005. The origins of pandemic influenza - lessons from the 1918 virus. N. Engl. J. Med. 353, 2209-2211.

Berns, K., Parrish, C. R., 2007. Parvoviridae, p. 2437-2477 *In* Knipe, D. M., Howley, P. M., Griffin, D. E., Martin, M. A., Lamb, R. A., Roizman, B., and Straus, S. E. (ed.), Fields Virology, Vol. 2, Fifth ed. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.

Block, W., Upton, C., McFadden, G., 1985. Tumorigenic poxviruses: genomic organization of malignant rabbit virus, a recombinant between Shope fibroma virus and myxoma virus. Virology 140, 113-124.

Brown, S. E., Gonzalez, H. A., Bodkin, D. K., Tesh, R. B., Knudson, D. L., 1988. Intra- and inter-serogroup genetic relatedness of orbiviruses. II. Blot hybridization and reassortment in vitro of epizootic haemorrhagic disease serogroup, bluetongue type 10 and Pata viruses. J Gen Virol. 69, 135-147.

Bujarski, J. J., Dzianott, A. M., 1991. Generation and analysis of nonhomologous RNA-RNA recombinants in brome mosaic virus: sequence complementarities at crossover sites J. Virol. 65, 4153-4159.

Bujarski, J. J., 2008. Recombination, p. 199-207 *In* Mahy, B. W. H. J., van Regenmortel, M. H. V. (ed.)., Desk Encyclopedia of General Virology. Academic Press, San Diego, California.

Chare, E. R., Gould, E. A., Holmes, E. C., 2003. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. J. Gen. Virol. 84, 2691-2703.

Charrel, R. N., Lemasson, J. J., Garbutt, M., Khelifa, R., De Micco, P., Feldmann, H., de Lamballerie, X., 2003. New insights into the evolutionary relationships between arenaviruses provided by comparative analysis of small and large segment sequences. Virology 317, 191-196.

Chen, W., Calvo, P. A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., Basta, S., O'Neill, R., Schickli, J., Palese, P., Henklein, P., Bennink, J. R., Yewdell, J. W., 2001. A novel influenza A virus mitochondrial protein that induces cell death. Nat. Med. 7, 1306-1312.

Cheng, L. L., Rodas, J. D., Schultz, K. T., Christensen, B. M., Yuill, T. M., Israel, B. A., 1999. Potential for evolution of California serogroup bunyaviruses by genome reassortment in Aedes albopictus. Am. J. Trop. Med. Hyg. 60, 430-438.

Chetverin, A. B., Chetverina, H.V., Demidenko, A. A., Ugarov, V. I., 1997. Nonhomologous RNA recombination in a cell-free system: evidence for a transesterification mechanism guided by secondary structure. Cell 88, 503-513.

Claverie, J. M., Ogata, H., Audic, S., Abergel, C., Suhre, K., Fournier, P. E., 2006. Mimivirus and the emerging concept of "giant" virus. Virus Res. 117, 133-144.

Conenello, G. M., Zamarin, D., Perrone, L. A., Tumpey, T., Palese, P., 2007. A single mutation in the PB1-F2 of H5N1 (HK/97) and 1918 influenza A viruses contributes to increased virulence. PLoS Pathog. 3, 1414-1421.

Cotmore, S. F., Tattersall, P., 2006. A rolling-hairpin strategy: basic mechanisms of DNA replication in the parvoviruses, p. 171-188 *In* Kerr, J. R., Cotmore, S. F., Bloom, M. E., Linden, R. M., Parrish, C. R. (ed.), Parvoviruses. Oxford University Press, New York, New York.

Crotty, S., Cameron, C. E., Andino, R., 2001. RNA virus error catastrophe: direct molecular test by using ribavirin. Proc. Natl. Acad. Sci. U. S. A. 98, 6895-6900.

DeFilippis, V. R., Villarreal, L. P., 2001. Virus evolution, p. 353-370 *In* Knipe, D. M., Howley, P. M., Griffin, D. E., Martin, M. A., Lamb, R. A., Roizman, B., Straus, S. E. (ed.), Fields Virology, Vol. 1, Fourth ed. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.

Domingo, E., 1998. Quasispecies and the implications for virus persistence and escape. Clin. Diagn. Virol. 10, 97-101.

Domingo, E., 2003. Quasispecies and the development of new antiviral strategies. Prog. Drug Res. 60, 133-58.

Domingo, E., 2007. Virus evolution, p. 389-421 *In* Knipe, D. M., Howley, P. M., Griffin, D. E., Martin, M. A., Lamb, R. A., Roizman, B., Straus, S. E. (ed.), Fields Virology, Vol. 1, Fifth ed. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.

Domingo, E., Holland. J. J., 1997. RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. 51, 151-178.

Drake, J. W., 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. U. S. A. 88, 7160-7164.

Duarte, E. A., Novella, I. S., Weaver, S. C., Domingo, E., Wain-Hobson, S., Clarke, D. K., Moya, A., Elena, S. F., de la Torre, J. C., Holland, J. J., 1994. RNA virus quasispecies: significance for viral disease and epidemiology. Infect. Agents Dis. 3, 201-124.

Duffy, S., Holmes, E. C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. J. Virol. 82, 957-965.

Duffy, S., Shackelton, L. A., Holmes, E. C., 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. 9, 267-276.

Eigen, M., 1971. Molecular self-organization and the early stages of evolution. Q. Rev. Biophys. 4, 149-212.

Eigen, M., Schuster, P., 1977. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. Naturwissenschaften 64, 541-565.

Emerson, S. U., Yu, Y., 1975. Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus. J. Virol. 15, 1348-1356.

Escoffier, C., Manié, S., Vincent, S., Muller, C. P., Billeter, M., Gerlier, D., 1999. Nonstructural C protein is required for efficient measles virus replication in human peripheral blood cells. J. Virol. 73, 1695-1698.

Esposito, J. J., Sammons, S. A., Frace, A. M., Osborne, J. D., Olsen-Rasmussen, M., Zhang, M., Govil, D., Damon, I. K., Kline, R., Laker, M., Li, Y., Smith, G. L., Meyer, H., Leduc, J. W., Wohlhueter, R. M., 2006. Genome sequence diversity and clues to the evolution of variola (smallpox) virus. Science 313, 807-812.

Evans, D. M., Dunn, G., Minor, P. D., Schild, G. C., Cann, A. J., Stanway, G., Almond, J. W., Currey, K., Maizel, J. V. Jr., 1985. Increased neurovirulence associated with a single nucleotide change in a noncoding region of the Sabin type 3 poliovaccine genome. Nature 314, 548-550.

Forrester, N. L., Moss, S. R., Turner, S. L., Schirrmeier, H., Gould, E. A., 2008. Recombination in rabbit haemorrhagic disease virus: possible impact on evolution and epidemiology. Virology 376, 390-396.

Frolova, E., Frolov, I., Schlesinger, S., 1997. Packaging signals in alphaviruses. J. Virol. 71, 248-258.

Fujii, Y., Kiyotani, K., Yoshida, T., Sakaguchi, T., 2001. Conserved and non-conserved regions in the Sendai virus genome: evolution of a gene possessing overlapping reading frames. Virus Genes 22, 47-52.

Gerhardt, R. R., Gottfried, K. L., Apperson, C. S., Davis, B. S., Erwin, P. C., Smith, A. B., Panella, N. A., Powell, E. E., Nasci, R. S., 2001. First isolation of La Crosse virus from naturally infected Aedes albopictus. Emerg. Infect. Dis. 5, 807-11.

Gibbs, A. J., Keese, P. K., 1995. In search of the origins of viral genes, p. 76-90 *In* Gibbs, A. J., Calisher, C. H., Garcia-Arenal, F. (ed.), Molecular basis of virus evolution. Cambridge University Press, Cambridge, United Kingdom.

Gibbs, M. J., Gibbs, A. J., 2006. Was the 1918 pandemic caused by a bird flu? Nature 440, E8.28.

Gillam, S., Atkinson, T., Markham, A., Smith, M., 1985. Gene K of bacteriophage φ X174 codes for a protein which affects the burst size of phage production. J. Virol. 53, 708-709.

Gmyl, A. P., Belousov, E. V., Maslova, S. V., Khitrina, E. V., Chetverin, A. B., Agol, V. I., 1999. Nonreplicative RNA recombination in poliovirus. J. Virol. 73, 8958-8965.

Gmyl, A. P., Korshenko, S. A. , Belousov, E. V., Khitrina, E. V., Agol, V. I., 2003. Nonreplicative homologous RNA recombination: promiscuous joining of RNA pieces? RNA 9, 1221-1231.

Gog, J. R., Afonso-Edos, S., Dalton, R. M., Leclercq, I., Tiley, L., Elton, D., von Kirchbach, J. C., Naffakh, N., Escriou, N., Digard, P., 2007. Codon conservation in the influenza A virus genome defines RNA packaging signals. Nucleic Acids Res. 35, 1897-1907.

Gouvea, V., Brantly, M., 1995. Is rotavirus a population of reassortants? Trends Microbiol. 3, 159-162.

Govindasamy, L., Hueffer, K., Parrish, C. R., Agbandje-McKenna, M., 2003. Structures of host range-controlling regions of the capsids of canine and feline parvoviruses and mutants. J. Virol. 77, 12211-12221.

Grimstad, P. R., Shabino, C. L., Calisher, C. H., Waldman, R. J., 1982. A case of encephalitis in a human associated with a serologic rise to Jamestown Canyon virus. Am. J. Trop. Med. Hyg. 31, 1238-1244.

Gubala, A. J., Proll, D. F., Barnard, R. T., Cowled, C. J., Crameri, S. G., Hyatt, A. D., Boyle, D. B., 2008. Genomic characterisation of Wongabel virus reveals novel genes within the Rhabdoviridae. Virology 376, 13-23.

Guillot, S., Otelea, D., Delpeyroux, F., Crainic, R., 1994. Point mutations involved in the attenuation/neurovirulence alternation in type 1 and 2 oral polio vaccine strains detected by site-specific polymerase chain reaction. Vaccine 12, 503-507.

Hahn, C. S., Lustig, S., Strauss, E. G., Strauss, J. H., 1988. Western equine encephalitis virus is a recombinant virus. Proc. Natl. Acad. Sci. U. S. A. 85, 5997-6001.

Hardy, R. W., Rice, C. M., 2005. Requirements at the 3' end of the Sindbis virus genome for efficient synthesis of minus-strand RNA. J. Virol. 79, 4630-4639.

Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R., Kaesberg, P., 1984. Striking similarities in amino acid sequence among nonstructural proteins encoded by RNA viruses that have dissimilar genomic organization. Proc. Natl. Acad. Sci. U. S. A. 81, 4358-4362.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., VandePol, S., 1982. Rapid evolution of RNA genomes. Science 215, 1577-1585.

Holmes, E. C., 2003a. Molecular clocks and the puzzle of RNA virus origins. J. Virol. 77, 3893-3897.

Holmes, E. C., 2003b. Error thresholds and the constraints to RNA virus evolution. Trends Microbiol. 11, 543-546.

Hon, C. C., Lam, T. Y., Drummond, A., Rambaut, A., Lee, Y. F., Yip, C. W., Zeng, F., Lam, P. Y., Ng, P. T., Leung, F. C., 2006. Phylogenetic analysis reveals a correlation between the expansion of very virulent infectious bursal disease virus and reassortment of its genome segment B. J. Virol. 80, 8503-8509.

Hon, C. C, Lam, T. T., Yip, C. W., Wong, R.T., Shi, M., Jiang, J., Zeng, F., Leung, F. C., 2008. Phylogenetic evidence for homologous recombination within the family Birnaviridae. J. Gen. Virol. 89, 3156-3164.

Hong, J. R., Gong, H. Y., Wu, J. L., 2002. IPNV VP5, a novel anti-apoptosis gene of the Bcl-2 family, regulates Mcl-1 and viral protein expression. Virology 295, 217-229.

Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J., Yeh, H. H., 2007. Reassortment and concerted evolution in banana bunchy top virus genomes. J. Virol. 81, 1746-1761.

Hueffer, K., Parker, J. S., Weichert, W. S., Geisel, R. E., Sgro, J. Y., Parrish, C. R., 2003. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. J. Virol. 77, 1718-1726.

Hutchinson, E. C., von Kirchbach, J. C., Gog, J. R., Digard, P., 2010. Genome packaging in influenza A virus. J. Gen. Virol. 91, 313-328.

Iturriza-Gómara, M., Isherwood, B., Desselberger, U., Gray, J., 2001. Reassortment in vivo: driving force for diversity of human rotavirus strains isolated in the United Kingdom between 1995 and 1999. J Virol. 75, 3696-3705.

Iverson, L. E., Rose, J. K., 1981. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. Cell 23, 477-484.

Jackwood, M. W., Boynton, T. O., Hilt, D. A., McKinley, E. T., Kissinger, J. C., Paterson, A. H., Robertson, J., Lemke, C., McCall, A. W., Williams, S. M., Jackwood, J. W., Byrd, L. A., 2010. Emergence of a group 3 coronavirus through recombination. Virology 398, 98-108.

Jarvis, T. C., Kirkegaard, K., 1992. Poliovirus RNA recombination: mechanistic studies in the absence of selection. EMBO J. 11, 3135-3145.

Joklik, W. K., Roner, M. R., 1995. What reassorts when reovirus genome segments reassort? J. Biol. Chem. 270, 4181-4184.

Jordan, I. K., Sutter IV, B. A., McClure, M. A., 2000. Molecular evolution of the Paramyxoviridae and Rhabdoviridae multiple-protein-encoding P gene. Mol. Evol. 17, 75-86.

Kawaoka, Y., Krauss, S., Webster, R. G., 1989. Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. J. Virol. 63, 4603-4608.

Keese, P. K., Gibbs, A. J., 1992. Origins of genes: "big bang" or continuous creation? Proc. Natl. Acad. Sci. U. S. A. 89, 9489-9493.

Kim, K. H., Strauss, E. G., Strauss, J. H., 2000. Adaptive mutations in Sindbis virus E2 and Ross River virus E1 that allow efficient budding of chimeric viruses. J. Virol. 74, 2663-2670.

Kim, M. J., Kao, C., 2001. Factors regulating template switch in vitro by viral RNA-dependent RNA polymerases: implications for RNA-RNA recombination. Proc. Natl. Acad. Sci. U. S. A. 98, 4972-4977.

Koonin, E. V., Dolja, V. V., 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem. Mol. Biol. 28, 375-430.

Kretzschmar, E., Peluso, R., Schnell, M. J., Whitt, M. A., Rose, J. K., 1996. Normal replication of vesicular stomatitis virus without C proteins. Virology 216, 309-316.

Krishnamurthy, K., Heppler, M., Mitra, R., Blancaflor, E., Payton, M., Nelson, R. S., Verchot-Lubicz, J., 2003. The Potato virus X TGBp3 protein associates with the ER network for virus cell-to-cell movement. Virology 309, 135-151.

Kroemer, J. A., Webb, B. A., 2004. Polydnavirus genes and genomes: emerging gene families and new insights into polydnavirus replication Annu. Rev. Entomol. 49, 431-456.

Kuhn, R. J., Griffin, D. E., Owen, K. E., Niesters, H. G., Strauss, J. H., 1996. Chimeric Sindbis-Ross River viruses to study interactions between alphavirus nonstructural and structural regions. J. Virol. 70, 7900-7909.

Kurath, G., Leong, J. C., 1985. Characterization of infectious hematopoietic necrosis virus mRNA species reveals a nonvirion rhabdovirus protein. J. Virol. 53, 462-468.

Kurotani, A., Kiyotani, K., Kato, A., Shioda, T., Sakai, Y., Mizumoto, K., Yoshida, T., Nagai, Y., 1998. Sendai virus C proteins are categorically nonessential gene products but silencing their expression severely impairs viral replication and pathogenesis. Genes Cells 3, 111-124.

Lai, M. M., 1992. RNA recombination in animal and plant viruses. Microbiol. Rev. 56, 61-79.

Lamb, R. A., Choppin, P. W., 1983. The gene structure and replication of influenza virus. Annu. Rev. Biochem. 52, 467-506.

Lazzarini, R. A., Keene, J. D., Schubert, M., 1981. The origins of defective interfering particles of the negative-strand RNA viruses. Cell 26, 145-154.

Lefeuvre, P., Lett, J. M., Varsani, A., Martin, D. P., 2009. Widely conserved recombination patterns among single-stranded DNA viruses. J Virol. 83, 2697-2707.

Liang, Y., Hong, Y., Parslow, T. G., 2005. Cis-acting packaging signals in the influenza virus PB1, PB2, and PA genomic RNA segments. J. Virol. 79, 10348-10355.

Liao, C. L., Lai, M. M., 1992. RNA recombination in a coronavirus: recombination between viral genomic RNA and transfected RNA fragments. J. Virol. 66, 6117-6124.

Lopez, S., Yao, J. S., Kuhn, R. J., Strauss, E. G., Strauss, J. H., 1994. Nucleocapsid-glycoprotein interactions required for assembly of alphaviruses. J. Virol. 68, 1316-1323.

López-Bueno, A., Villarreal, L. P., Almendral, J. M., 2006. Parvovirus variation for disease: a difference with RNA viruses? Curr. Top. Microbiol. Immunol. 299, 349-370.

Lu, X., McDonald, S. M., Tortorici, M. A., Tao, Y. J., Vasquez-Del Carpio, R., Nibert, M. L., Patton, J. T., Harrison, S. C., 2008. Mechanism for coordinated RNA packaging and genome replication by rotavirus polymerase VP1. Structure 16, 1678-1688.

McCarthy, A. J., Goodman, S. J., 2010. Reassessing conflicting evolutionary histories of the Paramyxoviridae and the origins of respiroviruses with Bayesian multigene phylogenies. Infect. Genet. Evol. 10, 97-107.

Mertens, P. P. C., 1999. Orbiviruses and Coltiviruses (*Reoviridae*): General Features, p. 1043-1061 *In* Granoff, A., Webster, R. G. (ed.), Encyclopedia of Virology, Second ed. Elsevier Academic Press, San Diego, California.

Mertens, P. P. C., Maan, S., Samuel, A., Attoui, H., 2005. Genus Orbivirus, Family Reoviridae. p. 466-483 *In* Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., Ball, L. A. (eds.), Virus Taxonomy: Eighth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, California.

Mihindukulasuriya, K. A., Wu, G., St Leger, J., Nordhausen, R. W., Wang, D., 2008. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. J Virol. 82, 5084-5088.

Mindich, L., 1999. Precise packaging of the three genomic segments of the double-stranded RNA bacteriophage φ6. Microbiol. Mol. Biol. Rev. 63, 149-160.

Mörl, M., Schmelzer, C., 1990. Group II intron RNA-catalyzed recombination of RNA in vitro. Nucleic Acids Res. 18, 6545-6551.

Moss, S. R., Ayres, C. M., Nuttall, P. A., 1988. The Great Island subgroup of tick-borne orbiviruses represents a single gene pool. J. Gen. Virol. 69, 2721-2727.

Muramoto, Y., Takada, A., Fujii, K., Noda, T., Iwatsuki-Horimoto, K., Watanabe, S., Horimoto, T., Kida, H., Kawaoka, Y., 2006. Hierarchy among viral RNA (vRNA) segments in their role in vRNA incorporation into influenza A virions. J. Virol. 80, 2318-2325.

Murphy, F. A., Gibbs, E. P. J., Horzinek, M. C., Studdert, M. J. (ed.), 1999. Veterinary Virology. Academic Press, San Diego, California.

Murray, C. J., Lopez, A. D., Chin, B., Feehan, D., Hill, K. H., 2006. Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918-20 pandemic: a quantitative analysis. Lancet 368, 2211-2218.

Nagy, P. D., Simon, A. E., 1997. New insights into the mechanisms of RNA recombination. Virology 235, 1-9.

Netolitzky, D. J., Schmaltz, F. L., Parker, M. D., Rayner, G. A., Fisher, G. R., Trent, D. W., Bader, D. E., Nagata, L. P., 2000. Complete genomic RNA sequence of western equine encephalitis virus and expression of the structural genes. J. Gen. Virol. 81, 151-159.

Nibert, M. L., Margraf, R. L., Coombs, K. M., 1996. Nonrandom segregation of parental alleles in reovirus reassortants. J. Virol. 70, 7295-7300.

Nuttall, P. A., Moss, S. R., 1989. Genetic reassortment indicates a new grouping for tick-borne orbiviruses. Virology 171, 156-161.

Nuttall, P. A., Moss, S. R., Jones, L. D., Carey, D., 1989. Identification of the major genetic determinant for neurovirulence of tick-borne orbiviruses. Virology 172, 428-434.

Nuttall, P. A., Moss, S. R., Carey, D., Jones, L. D., Jacobs, S. C., 1990. Genetic determinants modulating the pathogenic phenotype of tick-borne orbiviruses. Virology 174, 430-435.

Nuttall, P. A., Jacobs, S. C., Jones, L. D., Carey, D., Moss, S. R., 1992. Enhanced neurovirulence of tick-borne orbiviruses resulting from genetic modulation. Virology. 187, 407-412.

Olveira, J. G., Souto, S., Dopazo, C. P., Thiéry, R., Barja, J. L., Bandín, I., 2009. Comparative analysis of both genomic segments of betanodaviruses isolated from epizootic outbreaks in farmed fish species provides evidence for genetic reassortment. J. Gen. Virol. 90, 2940-2951.

Palermo, L. M., Hueffer, K., Parrish, C. R., 2003. Residues in the apical domain of the feline and canine transferrin receptors control host-specific binding and cell infection of canine and feline parvoviruses. J. Virol. 77, 8915-8923.

Palermo, L. M., Hafenstein, S. L., Parrish, C. R., 2006. Purified feline and canine transferrin receptors reveal complex interactions with the capsids of canine and feline parvoviruses that correspond to their host ranges. J Virol. 80, 8482-8492.

Parrish, C. R., 1999. Host range relationships and the evolution of canine parvovirus. Vet. Microbiol. 69, 29-40.

Pantin-Jackwood, M. J., Spackman, E., Woolcock, P. R., 2006. Phylogenetic analysis of Turkey astroviruses reveals evidence of recombination. Virus Genes 32, 187-192.

Patton, J. T., Spencer, E., 2000. Genome replication and packaging of segmented double-stranded RNA viruses. Virology 277, 217-25.

Pavesi, A., 2000. Detection of signature sequences in overlapping genes and prediction of a novel overlapping gene in hepatitis G virus. J. Mol. Evol. 50, 284-295.

Pavesi A., 2006. Origin and evolution of overlapping genes in the family Microviridae. J Gen. Virol. 87, 1013-1017.

Peluso, R. W., Richardson, J. C., Talon, J., Lock, M., 1996. Identification of a set of proteins (C' and C) encoded by the bicistronic P gene of the Indiana serotype of vesicular stomatitis virus and analysis of their effect on transcription by the viral RNA polymerase. Virology 218, 335-342.

Phan, T. G., Okitsu, S., Maneekarn, N., Ushijima, H., 2007. Evidence of intragenic recombination in G1 rotavirus VP7 genes. J. Virol. 81, 10188-10194.

Raju, R., Subramaniam, S. V., Hajjou, M., 1995. Genesis of Sindbis virus by in vivo recombination of nonreplicative RNA precursors. J. Virol. 69, 7391-7401.

Rancurel, C., Khosravi, M., Dunker, A. K., Romero, P. R., Karlin, D., 2009. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. J. Virol. 83, 10719-10736.

Reid, A. H., Taubenberger, J. K., 2003. The origin of the 1918 pandemic influenza virus: a continuing enigma. J. Gen.Virol. 84, 2285-2292.

Roner, M. R., Lin, P. N., Nepluev, I., Kong, L. J., Joklik, W. K., 1995. Identification of signals required for the insertion of heterologous genome segments into the reovirus genome. Proc. Natl. Acad. Sci. U. S. A. 92, 12362-12366.

Roner, M. R., Steele, B. G., 2007a. Localizing the reovirus packaging signals using an engineered m1 and s2 ssRNA. Virology 358, 89-97.

Roner, M. R., Steele, B. G., 2007b. Features of the mammalian orthoreovirus 3 Dearing l1 single-stranded RNA that direct packaging and serotype restriction. J. Gen. Virol. 88, 3401-3412.

Rozanov, M. N., Koonin, E. V., Gorbalenya, A. E., 1992. Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. J. Gen. Virol. 73, 2129-2134.

Samal, S. K., el-Hussein, A., Holbrook, F. R., Beaty, B. J., Ramig, R. F., 1987. Mixed infection of Culicoides variipennis with bluetongue virus serotypes 10 and 17: evidence for high frequency reassortment in the vector. J. Gen. Virol. 68, 2319-2329.

Sanjuán, R., Moya, A., Elena, S. F., 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. Acad. Sci. U. S. A. 101, 8396-8401.

Schoehn, G., Moss, S. R., Nuttall, P. A., Hewat, E. A., 1997. Structure of Broadhaven virus by cryoelectron microscopy: correlation of structural and antigenic properties of Broadhaven virus and bluetongue virus outer capsid proteins. Virology 235, 191-200.

Scholthof, H. B., 2006. The tombusvirus-encoded P19: from irrelevance to elegance. Nat. Rev. Microbiol. 4, 405-411.

Shackelton, L. A., Parrish, C. R., Truyen, U., Holmes, E. C., 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc. Natl. Acad. Sci. U. S. A. 102, 379-384.

Shaffer, J. A., Bellini, W. J., Rota, P. A., 2003. The C protein of measles virus inhibits the type I interferon response. Virology 315, 389-397.

Sibold, C., Meisel, H., Krüger, D. H., Labuda, M., Lysy, J., Kozuch, O., Pejcoch, M., Vaheri, A., Plyusnin, A., 1999. Recombination in Tula hantavirus evolution: analysis of genetic lineages from Slovakia. J. Virol. 73, 667-675.

Simmonds, P., 2006. Recombination and selection in the evolution of picornaviruses and other mammalian positive-stranded RNA viruses. J. Virol. 80, 11124-11140.

Sprenger, D., Wuithiranyagool, T., 1986. The discovery and distribution of Aedes albopictus in Harris County, Texas. J. Am. Mosq. Control Assoc. 2, 217-219.

Springfeld, C., Darai, G., Cattaneo, R., 2005. Characterization of the tupaia rhabdovirus genome reveals a long open reading frame overlapping with P and a novel gene encoding a small hydrophobic protein. J. Virol. 79, 6781-6790.

Steinhauer, D. A., Domingo, E., Holland, J. J., 1992. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. Gene 122, 281-288.

Stollar, V., 1979. Defective interfering particles of togaviruses. Curr. Top. Microbiol. Immunol. 86, 35-66.

Strauss, E. G., Rice, C. M., Strauss, J. H., 1983. Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. Proc. Natl. Acad. Sci. U. S. A. 80, 5271-5275.

Strauss, J. H., 1983. Recombination in the evolution of RNA viruses, p. 241-251 *In* Morse, S. S. (ed.), Emerging Viruses. Oxford University Press, New York, New York.

Strauss, J.H., Strauss, E.G., 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. 58, 491-562.

Strauss, J. H., Strauss, E. G., 1997. Recombination in alphaviruses. Semin. Virol. 8, 85-94.

Surendranath, B., Usharani, K. S., Nagma, A., Victoria, A. K., Malathi, V. G., 2005. Absence of interaction of genomic components and complementation between Mungbean yellow mosaic India virus isolates in cowpea. Arch. Virol. 150, 1833-1844.

Svitkin, Y. V., Cammack, N., Minor, P. D., Almond, J. W., 1990. Translation deficiency of the Sabin type 3 poliovirus genome: association with an attenuating mutation C472U. Virology 175, 103-109.

Tao, Y., Farsetta, D. L., Nibert, M. L., Harrison, S. C., 2002. RNA synthesis in a cage - structural studies of reovirus polymerase lambda3. Cell 111, 733-745.

Taraporewala, Z. F., Patton, J. T., 2004. Nonstructural proteins involved in genome packaging and replication of rotaviruses and other members of the Reoviridae. Virus Res. 101, 57-66.

Taubenberger, J. K., Reid, A. H., Lourens, R. M., Wang, R., Jin, G., Fanning, T. G., 2005. Characterization of the 1918 influenza virus polymerase genes. Nature 437, 889-893.

Tchatalbachev, S., Flick, R., Hobom, G., 2001. The packaging signal of influenza viral RNA molecules. RNA 7, 979-989.

Tessman, E. S., Tessman, I., Pollock, T. J., 1980. Gene K of bacteriophage φX174 codes for a nonessential protein. J. Virol. 33, 557-560.

Truyen, U., Parrish, C. R., 1992. Canine and feline host ranges of canine parvovirus and feline panleukopenia virus: distinct host cell tropisms of each virus in vitro and in vivo. J. Virol. 66, 5399-5408.

Truyen, U., 2006. Evolution of canine parvovirus - a need for new vaccines? Vet. Microbiol. 117, 9-13.

Tsagris, M., Tabler, M., Sänger, H. L., 1991. Ribonuclease T1 generates circular RNA molecules from viroid-specific RNA transcripts by cleavage and intramolecular ligation. Nucleic Acids Res. 19, 1605-1612.

Tumpey, T. M., Basler, C. F., Aguilar, P. V., Zeng, H., Solórzano, A., Swayne, D. E., Cox, N. J., Katz, J. M., Taubenberger, J. K., Palese, P., García-Sastre, A., 2005. Characterization of the reconstructed 1918 Spanish influenza pandemic virus. Science. 310, 77-80.

Twiddy, S. S., Holmes, E. C., 2003. The extent of homologous recombination in members of the genus Flavivirus. J. Gen. Virol. 84, 429-440.

Umemura, T., Tanaka, Y., Kiyosawa, K., Alter, H. J., Shih, J. W., 2002. Observation of positive selection within hypervariable regions of a newly identified DNA virus (SEN virus). FEBS Lett. 510, 171-174.

Umene, K., 1999. Mechanism and application of genetic recombination in herpesviruses. Rev. Med. Virol. 9, 171-182.

Urquidi, V., Bishop, D. H., 1992. Non-random reassortment between the tripartite RNA genomes of La Crosse and snowshoe hare viruses. J. Gen. Virol. 73, 2255-2265.

Van Dooren, S., Salemi, M., Vandamme, A. M., 2001. Dating the origin of the African human T-cell lymphotropic virus type-I (HTLV-I) subtypes. Mol. Biol. Evol. 18, 661-671.

Walker, P. J., Byrne, K. A., Riding, G. A., Cowley, J. A., Wang, Y., McWilliam, S., 1992. The genome of bovine ephemeral fever rhabdovirus contains two related glycoprotein genes. Virology 191, 49-61.

Wang, Y., Walker, P. J., 1993. Adelaide River rhabdovirus expresses consecutive glycoprotein genes as polycistronic mRNA: new evidence of gene duplication as an evolutionary process. Virology 195, 719-731.

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

Webby, R. J., Webster, R. G., 2003. Are we ready for pandemic influenza? Science 302, 1519-1522.

Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., Kawaoka, Y., 1992. Evolution and ecology of influenza A viruses. Microbiol. Rev. 56, 152-179.

Worobey, M., Holmes, E. C., 1999. Evolutionary aspects of recombination in RNA viruses. J. Gen. Virol. 80, 2535-2543.

Yao, J. S., Strauss, E. G., Strauss, J. H., 1996. Interactions between PE2, E1, and 6K required for assembly of alphaviruses studied with chimeric viruses. J. Virol. 70, 7910-7920.

Yao, J., Strauss, E. G., Strauss, J. H., 1998. Molecular genetic study of the interaction of Sindbis virus E2 with Ross River virus E1 for virus budding. J. Virol. 72, 1418-1423.

Zhao, X., McGirr, K. M., Buehring, G. C., 2007. Potential evolutionary influences on overlapping reading frames in the bovine leukemia virus pXBL region. Genomics 89, 502-511.

CHAPTER 3

RECOMBINATION, PART I:

A COMPARATIVE GENOMIC ANALYSIS OF HIGHLANDS J VIRUS WITH WESTERN AND EASTERN EQUINE ENCEPHALITIS VIRUSES[1]

[1]Allison, A. B., and D. E. Stallknecht. 2009. *Virus Research* 145:334-40.

Reprinted here in unabridged format with permission of publisher.

Abstract

Highlands J virus (HJV) is a member of the genus *Alphavirus*, family *Togaviridae*. HJV is the sole representative of the western equine encephalitis (WEE) serocomplex found in the eastern United States, and circulates in nature in an apparently identical transmission cycle as *eastern equine encephalitis virus* (EEEV). North American representatives of the WEE serocomplex [HJV, WEE virus (WEEV), and *Fort Morgan virus* (FMV)] are believed to be derived from a recombination event involving EEEV and a *Sindbis* (SIN)-like virus, such that the nonstructural polyprotein, the capsid, and the terminal end of the 3' UTR were derived from EEEV, while the major surface glycoproteins (E1 and E2) encoded in the subgenomic RNA were derived from the SIN-like virus. In this report, the complete nucleotide sequence of HJV is described, along with a comparative analysis of the HJV genome to WEEV and EEEV. Analysis of the 3' UTR of the parental viruses disclosed that both EEEV and SINV share similar sequence motifs, that in conjunction with the double cross-over, may have facilitated the survival of the ancestral recombinant leading to the diversification of the recombinant alphaviruses in the New World.

Introduction

The genus *Alphavirus* within the family *Togaviridae* is comprised of 29 species encompassing eight antigenic complexes based on serological cross-reactivity and genetic similarity (Weaver *et al*., 2005). Four of these complexes (eastern, western, and Venezuelan equine encephalitis; Semliki Forest) contain members found in the New World. Although some New World representatives of the genus (i.e., *Mayaro virus*) can cause a severe, but generally non-fatal, polyarthritis or rash/arthralgia syndrome typical of Old World alphaviruses, North and South American complexes are characterized by human and veterinary pathogens which may induce fatal encephalitis (Powers *et al*., 2006).

The western equine encephalitis (WEE) complex is composed of 6 species: *WEE virus* (WEEV), *Highlands J virus* (HJV), *Fort Morgan virus* (FMV) [including Buggy Creek virus, a variant of FMV], *Aura virus* (AURAV), *Sindbis virus* (SINV), and *Whataroa virus* (WHAV) (Weaver *et al*., 1997). Of these, only HJV, WEEV, and FMV are present in the North America, with HJV being the only WEE complex member found in the eastern United States. The only other alphavirus present in the eastern United States is *eastern equine encephalitis virus* (EEEV), whose geographical distribution is nearly identical to that of HJV. Although HJV was initially described as an 'eastern' variant of WEEV (Hayes and Wallis, 1977), serological testing and oligonucleotide fingerprint analysis of the E2 glycoprotein of WEEV and HJV isolates (Trent and Grant, 1980), in addition to cross neutralization assays among the WEE complex viruses (Calisher *et al*., 1988), determined that HJV is a close, but distinct, genetic and antigenic relative of WEEV.

Both EEEV and HJV circulate in the United States under apparently identical transmission cycles, sharing the same enzootic mosquito vector [*Culiseta* (*Cs.*) *melanura*], vertebrate amplifying hosts (passerine and wading birds), and geographical range (coastal areas from Maine south to Texas, in addition to a number of inland foci) (McLean *et al*., 1985; Scott and Weaver, 1989; Cilnis *et al*., 1996). The earlier seasonal appearance and/or higher prevalence of HJV in local mosquito or bird populations, relative to that of EEEV, has been suggested to be, at times, a useful predictor of oncoming EEEV activity (Williams *et al*., 1971; Dalrymple, *et al*., 1972; Edman *et al*., 1993). However, as virus activity likely varies both regionally and temporally, HJV may represent more of a casual indicative, rather than predictive, sentinel for potential EEEV activity (Andreadis *et al*., 1998). WEEV is more widespread geographically than either HJV or EEEV, stretching from western Canada south to Argentina (Waters, 1976; Mitchell *et al*., 1987). Like HJV and EEEV, passerine birds serve as the primary amplifying hosts for WEEV in North America. However, WEEV is predominately vectored (enzootically and epizootically) in North America by *Culex* (*Cx.*) *tarsalis* (Reisen and Monath, 1988), which likely facilitates its broader geographical distribution, although the absense of WEEV in the eastern United States may be due in part to the rarity of *Cx. tarsalis* east of the Mississippi (Darsie and Ward, 2005). Although both EEEV and WEEV are additionally found in the tropics of Central and/or South America (Reisen and Monath, 1988; Brault *et al*., 1999), HJV has only been described from North America.

EEEV, WEEV, and HJV differ substantially in their virulence in mammalian hosts. EEEV is a severe human and equine neuropathogen with apparent case fatality

rates of ~30% and 90%, respectively (Przelomski *et al*., 1988; Deresiewicz *et al*., 1997; Fenner *et al*., 1993). WEEV is intermediate in its virulence, with apparent case fatality rates of 5% in humans and 20-30% in horses (Hayes, 1981; Englund *et al*., 1986). Although rare serologically confirmed cases of equine encephalitis due to HJV have been documented (Jennings *et al*., 1967; Hoff *et al*., 1978), in addition to the isolation of HJV from the brain of an encephalitic Florida horse in 1964 (Jennings *et al*., 1966; Karabatsos *et al*., 1988), HJV is generally not considered to be an equine pathogen. Likewise, HJV is not regarded as a public health threat, as no cases of human encephalitis due to HJV infection have been reported. However, HJV, like EEEV (and to a lesser extent WEEV), has been documented as an avian pathogen (Ficken *et al*., 1993; Guy *et al*., 1993; Guy *et al*., 1994; Wages *et al*., 1993; Eleazer and Hill, 1994; Randolph *et al*., 1994).

Full-length genomic sequences of HJV are currently not available, although the structural polyprotein and portions of nsP1, nsP4, and the 5' UTR have been previously sequenced (i.e., 5.17kB or 44% of the genome) (Ou *et al*., 1982a; Ou *et al*., 1983; Cilnis *et al*., 1996; Pfeffer *et al*., 1997; Weaver *et al*., 1997; Pfeffer *et al*., 1998). As the complete sequences of each of the four nonstructural proteins of HJV, encompassing approximately 6.36kB, were previously unknown, the genomic sequence of a HJV isolate (585-01) obtained from the brain of a red-tailed hawk (*Buteo jamaicensis*) during arbovirus surveillance in Georgia (NCBI accession FJ827631), in addition to the prototype strain HJV B-230 (NCBI accession GQ227789), was determined.

Materials and methods

Virus isolation and identification

On 2 August 2001, a red-tailed hawk (*Buteo jamaicensis*) was submitted from Kings Bay, Camden County, Georgia, to the Southeastern Cooperative Wildlife Disease Study for arbovirus testing (dead bird submission 585-01). For virus isolation, samples of brain and heart (~0.5cm^3) was mechanically homogenized in 650µl of virus isolation media [1X minimum essential media, 2.2g/l $NaHCO_3$, 10% fetal bovine serum, 400 units/ml penicillin, 400µg/ml streptomycin, 1µg/ml amphotericin B]. Homogenized tissues were centrifuged (6700 x *g* for 10 min) to pellet debris and an aliquot (100µl) of clarified supernatant was used to inoculate 2-day-old Vero cells in a 12-well plate format. Plates were incubated at 37 °C in a humidified atmosphere containing 5% CO_2. For wells exhibiting cytopathology, cell culture supernatant was harvested and RNA was extracted using a QIAamp® Viral RNA Mini kit (Qiagen) according to the manufacturer's instructions.

HJV-specific primers based on the E1 glycoprotein, HJV-E1-10115/10644 (5'-CTGCGATAGCGAAAACAC-3' and 5'-CACAGTCAACCGCCCTTA-3'), were used for initial identification. A single-tube RT-PCR (50µl) was set up using 10mM Tris-HCl (pH 9.0), 50mM KCl, 0.1% Triton® X-100, 1.5mM $MgCl_2$, 250µM deoxynucleotide triphosphates (dNTPs), 0.5µM of each primer, 2U of AMV reverse transcriptase (Promega), 1U of *Taq* DNA polymerase (Promega), and 3µl of extracted RNA. Cycling parameters were reverse transcription at 42 °C for 20 min, followed by 40 cycles of denaturation at 94 °C for 1 min, annealing at 50 °C for 1 min, and extension at 72 °C for

1 min. Reaction products were visualized by electrophoresis in a 2% agarose TAE gel stained with ethidium bromide (EtBr; 0.1µg/ml) and subsequently purified using a QIAquick® Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. Sequencing of RT-PCR products was performed at the Integrated Biotechnology Laboratories at the University of Georgia using a 3100 Genetic Analyzer (Applied Biosystems, Inc.). HJV was isolated from both the brain and heart of 585-01.

Genomic sequencing

In order to amplify the portion of the HJV genome encoding the nonstructural polyprotein, degenerate primers were designed based on a CLUSTAL alignment of genomic sequences of EEEV (AY705240) and WEEV (NC003908). In most cases, HJV-specific primers were developed from RT-PCR products amplified using the degenerate primers in order to increase specificity and to fill in any gaps in sequence. To obtain the full-length nucleotide sequence of the structural genes, primers were developed based on the sequence of HJV prototype strain B-230 (AF339476). A list of all primers used in sequencing of the HJV genome may be obtained from the authors. The terminal 5' 21 nucleotides (initially amplified using the primer HJV-5'UTR-1 5'-ATAGGGYRTGGTATAGAGTGA-3') were verified by 5' RACE using a commercial kit (First Choice® RLM-RACE Kit, Ambion, Austin, TX). Briefly, an antisense primer, HJV-nsP1-557 (5'-ATAGTTCTTACGCCTTTTAG-3'), was used in conjunction with a sense primer complementary to a 5'-end ligated adaptor to amplify the 5' end. Similarly, the terminal 3' 30 nucleotides (initially amplified using the primer HJV-3'UTR-11526, 5'- GAAATTTTAAAAACAAAATAAAAGAAAAAT-3') were verified by 3' RACE.

Briefly, a sense primer, HJV-E1-10965 (5'-GAGTGTAAGCCCCCATCAGA-3'), was used in conjunction with an Oligo-d(T)25 primer to amplify the complete 3' end. RACE amplicons were electrophoresed on a 2% agarose EtBr-stained TAE gel, excised and purified from agarose, cloned using a PCR Cloning^{Plus}® Kit (Qiagen), and plasmid was subsequently purified using a QIAprep® Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. Sequencing of RT-PCR products or recombinant plasmid was performed as noted in the virus isolation section.

RNA and protein analysis

The complete genome of HJV was determined by construction of a contig of overlapping RT-PCR products using Sequencher version 4.1.4. (Gene Codes Corporation). Each nucleotide position was sequenced a minimum of two times. For comparative sequence analysis of HJV with other alphaviruses, nucleotide and deduced amino acid sequences were aligned using CLUSTALW (Chenna *et al*., 2003). All pairwise comparisons of HJV 585-01 with EEEV or WEEV were made against NCBI accession numbers EF151502 or AF214039/AF214040, respectively. Evolutionary relationships of HJV with representative alphaviruses were derived by construction of phylogenetic trees using the Molecular Evolutionary Genetics Analysis (MEGA) version 4 program (Tamura *et al*., 2007) and by similarity plots using the Simplot program version 3.5 (Lole *et al*., 1999). Phylogenetic trees were generated with a CLUSTALW alignment of the amino acid sequences of the nonstructural and structural polyproteins of representative alphaviruses using the Neighbor-joining method in the MEGA4 program. Bootstrap values were determined using 2000 replicates. The trees were calculated using

Poisson correction and evolutionary distances were represented as the number of amino acid substitutions per site. Similarity plots were generated with the complete nucleotide and nonstructural/structural polyprotein sequences of HJV and representative alphaviruses in the Simplot program using the 2-parameter (Kimura) distance model in a sliding window of 200 nucleotides, step size of 20. Predicted minimum energy secondary structures of the UTR sequences of selected alphaviruses were generated using the RNAfold program using default parameters (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) (Hofacker, 2003). Theoretical isoelectric points and molecular weights of the HJV nonstructural proteins were determined using the Compute pI/Mw tool on the ExPASy server (http://www.expasy.ch/tools/pi_tool.html) (Gasteiger *et al*., 2003). Phosphorylation sites of the nsP3 protein were predicted using the NetPhos 2.0 server (http://www.cbs.dtu.dk/services/NetPhos/) (Blom *et al*., 1999).

Results and discussion

Complete genome analysis

The HJV genome, like other alphaviruses, contains four genes encoded in a nonstructural polyprotein open reading frame (ORF) translated directly from the genomic RNA (nsP1, nsP2, nsP3, and nsP4) and five genes encoded on a subgenomic structural polyprotein ORF [capsid (C), E3, E2, 6K, and E1]. These nonstructural and structural gene sequences are separated and flanked by untranslated regions (UTRs). The alphavirus genome, schematically represented as 5'-UTR-nsP1-nsP2-nsP3-nsP4-UTR-C-E3-E2-6K-E1-UTR-3' (Fig. 3.1), contains a 5' cap0 structure [m7G(5')ppp(5')N]

(Vasiljeva *et al*., 2000) and a poly(A) tail (Raju *et al*., 1999), and ranges from approximately 11.4-11.9 kilobases (kB) in length (Strauss and Strauss, 1994). Excluding the cap nucleotide and poly(A) tail, the complete genome of HJV was 11,526 nucleotides long and coded for 3,686 amino acids. The nonstructural polyprotein gene sequence contained 7,350 nucleotides encoding 2,450 amino acids and represented 63.8% of the genome. The structural polyprotein gene sequence was approximately half that size, containing 3,708 nucleotides encoding 1,236 amino acids and occupied 32.2% of the genome. The three non-coding regions, the 5' UTR, the subgenomic junction UTR, and the 3' UTR, were 45, 44, and 379 nucleotides in length, respectively, and collectively occupied 4% of the genome. The nucleotide composition of the HJV genome was 28.4% A, 24.9% C, 24.1% G, and 22.5% U.

In comparison to EEEV and WEEV, the HJV genome was intermediate in length, being 18 nucleotides longer than WEEV (11,508) and 156 nucleotides shorter than EEEV (11,682), although HJV has the smallest coding capacity (3,686aa) of the three viruses. Over the complete genome, HJV shared a 75% nucleotide and an 87% amino acid identity to WEEV. Similarity plot analysis of the complete nucleotide sequence of EEEV, when compared to the HJV genome, demonstrated that, other than the C-terminal region of nsP3, similarity between the two viruses ends abruptly in the E3 gene sequence, consistent with the location of the coding region recombination crossover event reported for WEEV (Hahn *et al*., 1988; Strauss and Strauss, 1997) (Fig. 3.2A). Additionally, similarity plot analysis of the HJV and WEEV genomes, when compared against EEEV, were remarkably similar to one another throughout their entire nucleotide sequence, confirming that both HJV and WEEV are derived from the same recombination event

(Fig. 3.2B). A comprehensive comparative analysis of the genomes of HJV, EEEV, and WEEV is presented in Table 3.1.

5’ UTR analysis

The 5’ UTRs of EEEV, WEEV, and HJV were similar in length, ranging from 45-48 nucleotides. Two conserved regions are found in the 5’ end of the alphavirus genome: a 51-nucleotide (nt) conserved sequence element (CSE) residing within nsP1, and the terminal 5’ UTR itself, the latter being more conserved in structure than in sequence (Strauss and Strauss, 1994). The complement of the 5’ UTR (the 3’ UTR of the minus strand) is believed to function as a promoter for plus-strand synthesis (Strauss and Strauss, 1994). Additionally, the 5’ UTR (in positive polarity) has been demonstrated to interact with the 3’ end of the genome for the initiation of minus-strand synthesis (Frolov *et al*., 2001). The 51-nt CSE, which forms a double stem loop structure within the coding region of nsP1, has been reported to be an enhancer for both minus and plus-strand synthesis, although it is not essential for replication (Frolov *et al*., 2001).

Secondary structure predictions of the 5’ UTR demonstrated a close structural similarity between the three viruses, with nucleotides 2-30 forming a stem loop structure (SL-1) in HJV and WEEV, while the predicted SL-1 was slightly more elongated in EEEV, occupying nucleotides 1-31 of the 5’ UTR (Fig. 3.3A-B). The 51-nt CSE of HJV occupies nucleotide positions 134-185 (Fig. 3.3B) or amino acids positions 31-47. Comparative nucleotide and predicted secondary structure analysis of the 51-nt CSE region of HJV with WEEV indicated an overall sequence identity of 94% (49/52), with SL-3 and SL-4 covering a total of 52 nucleotides. The three nucleotides differing

between HJV and WEEV were at HJV positions 134 and 159 (constituting the first base pairing in SL-3) and position 144 (in the loop of SL-3). Both of the predicted stem loops for HJV and WEEV in the 51-nt CSE are structurally identical, with SL-4 incorporating a bulge (Fig. 3.3A), as both viruses contain a mismatched double guanine base-pairing previously reported for HJV (Niesters and Strauss, 1990). The predicted double stem loop structure for the 51-nt CSE of EEEV shared a 88% (45/51) nucleotide identity to both HJV and WEEV and was energetically more favorable, as it lacked the additional bulge in the second stem loop (Fig. 3.3A).

Nonstructural gene and polyprotein analysis

In pairwise comparisons of the amino acid sequences of the HJV nonstructural proteins to EEEV and WEEV, nsP1 was the most conserved (86-93%), followed by nsP4 (88-90%), nsP2 (84-89%), and nsP3 (69-73%) [Table 3.1]. Overall, the HJV nonstructural polyprotein shared an 82% and an 87% amino acid identity to EEEV and WEEV, respectively. In the nonstructural polyprotein, there are three sites (nsP1/nsP2, nsP2/nsP3, and nsP3/nsP4) that are cleaved by the protease activity of nsP2 in order to produce the four individual nonstructural proteins. The cleavage motifs for both EEEV and WEEV at the nsP1/nsP2, nsP2/nsP3, and nsP3/nsP4 junctions are EAGA/GSVE, EAGR/APAY, and RYEAGA/YIFS, respectively (Strauss and Strauss, 1994). The P1 arginine and P3' alanine in the nsP2/nsP3 motif (EAGR/APAY) are unique to EEEV and WEEV (Strauss and Strauss, 1994). All three cleavage site sequence motifs in the nonstructural polyprotein of HJV were identical to both EEEV and WEEV. Based on the cleavage site motifs, the theoretical isoelectric point and molecular weight of each of

the HJV nonstructural proteins were: nsP1 (6.64; 59.75kDa), nsP2 (8.76; 88.79kDa), nsP3 (7.69; 56.56kDa), and nsP4 (6.92; 67.98kDa). In addition to the cleavage site sequences, other known conserved motifs in the nonstructural genes, such as the nsP2 NTP binding motifs (GVPGSGKS and DEAF) (Vasiljeva *et al*., 2000; Sawicki *et al*., 2006), the nsP2 proteinase motif (CWA) (Hardy and Strauss, 1989), and the nsP4 polymerase motif (GDD) (Tomar *et al*., 2006), were invariant between HJV, EEEV, and WEEV.

All of the nonstructural proteins were identical in length between HJV, EEEV, and WEEV, except that the nsP3 gene of HJV (1,545nt) is truncated in relation to both WEEV (1,596nt) and EEEV (1,677nt). nsP3 is comprised of two primary regions: a highly conserved N-terminal region (which can be further divided into "X" and alphavirus-like conserved domains) and a hypervariable C-terminal region, the latter varying in both sequence and length among the alphaviruses (LaStraza *et al*., 1994; Vihinen *et al*., 2001). nsP3 is a phosphoprotein whose functional role in replication is more obscure than other nonstructural proteins, although it has been proposed to function in both minus-strand and subgenomic RNA synthesis (Wang *et al*., 1994; LaStraza *et al*., 1994; De *et al*., 2003), in addition to serving as an anchoring protein for replication complexes on membranes (Peränen and Kääriäinen, 1991).

The N-terminal portion of nsP3 (amino acids 1-325) is conserved among HJV, EEEV, and WEEV, with HJV sharing identities of 86% to WEEV and 83% to EEEV. The C-terminal region of nsP3 in HJV (residues 326-515), however, shared a 52% and 44% identity, and was truncated by 17 or 44 amino acids, in relation to WEEV and EEEV, respectively. Interestingly, deletions in the hypervariable C-terminal region of

nsP3 of Semliki Forest virus (SFV) have been associated with reduced levels of RNA synthesis and virus production due to a reduction in the formation of replication complexes on intracellular membranes, ultimately resulting in decreased virulence in mice (Galbraith *et al*., 2006). In nsP3, HJV had collectively fewer, and an overall lower frequency of, serine, threonine, and tyrosine residues (94; 18.25%) than either WEEV (102; 19.17%) or EEEV (115; 20.57%). Previously, it has been demonstrated with SFV that, as the C-terminal region of nsP3 contains a greater concentration of phosphorylation sites than does the conserved N-terminus, deletions in the C-terminus appear to reduce the level of phosphorylation of nsP3 (Vihinen *et al*., 2001). As with nsP3 C-terminal deletion mutants, SFV mutants defective in nsP3 phosphorylation have also been shown to be greatly reduced in pathogenicity in mice (Vihinen and Saarinen, 2000). Whether the degree of truncation (and/or phosphorylation) of nsP3 has an effect on the substantial differences in the pathogenicity of EEEV, WEEV, and HJV in mammalian hosts is unknown. However, as it has been recently demonstrated that EEEV isolates from South American lineages had nsP3 truncations ranging from 60-72 nucleotides with respect to North American EEEV (Kondig *et al*., 2007), this may suggest that the discrepancies in virulence recognized between North and South American strains of EEEV could possibly be due, in part, to the length of the nsP3 gene.

Similar to EEEV and WEEV, HJV contains an in-frame opal termination codon (UGA) directly preceding the RYEAGA/YIFS nsP3/nsP4 cleavage site. Termination at, or readthrough of, the codon in HJV would result in nonstructural polyproteins of 1,835 (P123) or 2,449 (P1234) amino acids in length (Strauss *et al*., 1983). Like all other alphaviruses sequenced, a C nucleotide follows the opal termination codon in HJV

(Strauss and Strauss, 1994). The phylogenetic tree of the nonstructural polyprotein (nsP1-nsP2-nsP3-nsP4) of HJV with other representative alphaviruses showed two major clades: New World encephalitic viruses (HJV, WEEV, EEEV, VEEV) and Old World arthralgic viruses (SINV, SFV) (Fig. 3.4B, left). Groupings within the New World encephalitic viruses indicated that the nonstructural polyprotein of HJV, like WEEV, was mostly closely related to EEEV. Comparison of the branch lengths between HJV and WEEV in the phylogenetic tree derived from the nonstructural proteins (Fig. 3.4B, left), or from the nonstructural proteins and capsid (not shown), indicated that HJV is more similar to the hypothetical progenitor of the HJV-WEEV group than is WEEV.

Subgenomic UTR and promoter region analysis

In contrast to the similarly sized 5' UTR of the three viruses, the subgenomic UTR is significantly larger in EEEV (66nt) than either WEEV (47nt) or HJV (44nt). Besides the C-terminal region of nsP3, the 3' end of the subgenomic UTR apparently represents the only other region where genomic sequence, originally obtained from EEEV, is truncated in both HJV and WEEV. Previous functional studies of the junction UTR and the terminal 3' end of nsP4 have identified two conserved regions involved in promoter activity for the transcription of the subgenomic RNA from the negative-sense strand: a 24-nt stretch, which spans the boundary between the 3' end of nsP4 and the 5' end of the junction region in EEEV, WEEV, and HJV (denoted as -19/+5 with respect to the transcription start site of the subgenomic RNA), and a 6-nt stretch located upstream in nsP4 (-35/-30) (Fig. 3.5A) (Ou *et al*., 1982b; Levis *et al*., 1989; Wielgosz *et al*., 2001). Two additional sequence elements adjacent to these conserved regions at positions -40/-

20 and +6/+14 have been demonstrated to be required for increased rates of transcription (Wielgosz *et al*., 2001). As the nsP4 termination codon for the three viruses lies at positions -6 to -8, positions from -5 to the start of the capsid protein (+37 in HJV) are not under coding constraints and would likely not be conserved if not essential for transcription (Wielgosz *et al*., 2001). Positions -5 to +5, encompassing the junction UTR portion of the minimal promoter, is invariant between the three viruses barring a single transition at position -2 for HJV (Fig. 3.5A).

The portion of the subgenomic UTR that is truncated in HJV and WEEV in relation to EEEV lies beyond position +27 (the last identifiable position of identity between the three viruses) (Fig. 3.5A). The first 13 nucleotides after position +27 in EEEV are either A or U, with positions +27 to the capsid start codon being 80% (25/31) AU. As AU-rich insertions in the 3' UTR of alphaviruses have been postulated to act as motifs for cellular RNA-binding proteins involved in replication (George and Raju, 2000), as well as the recognition of UA or UU insertions directly in the 3' terminal 19-nt CSE (Pfeffer *et al*., 1998), this AU rich region in EEEV may represent an insertion acquired after the recombination event. Alternatively, the truncated region may represent a deletion in HJV and WEEV, suggesting that (if present during recombination) the terminal third of the junction UTR may have not been essential for the transcription (or translation) of the subgenomic RNA in the ancestral recombinant.

Secondary structures (in positive-sense) generated from the portion of the subgenomic UTR incorporated into the subgenomic RNA exhibited a stem loop structure from positions 2-27 for HJV and WEEV and positions 1-28 for EEEV, followed by a lack of any secondary structure in the remaining sequence (Fig. 3.5B). Previously, secondary

structure predictions for the first 170 nucleotides of the subgenomic RNA of EEEV also demonstrated a hairpin structure from positions 1-28, followed a lack of any secondary structure until the three nucleotides preceding the capsid start codon (Frolov and Schlesinger, 1996). Additionally, secondary structures downstream of the capsid initiation codon, rather than upstream, were shown to be important for efficient translation of the subgenomic RNA (Frolov and Schlesinger, 1996). Moreover, the extreme 5' end of the subgenomic RNA of SFV (i.e., equivalent to the 5' cap and nucleotide 1 in HJV and WEEV) has been demonstrated to be integral for efficient translation via recognition by cap binding proteins (Berben-Bloemheuvel *et al*., 1992). Therefore, the portion of the subgenomic UTR of EEEV that is truncated in HJV and WEEV, does not appear to contain sequences, or generate any predicted secondary structures, that have been shown to be essential in enhancing transcription or translation of the subgenomic RNA. As South American varieties of EEEV also contain an extended subgenomic UTR (e.g., GenBank accessions DQ241304 and DQ241303; 76-79nt) which exhibit a high nucleotide identity to North American EEEV only up to the vicinity of position +27, this suggests that the 3' terminal region of the EEEV subgenomic UTR, or its complement in the 5' terminal region of the subgenomic RNA, is not conserved and therefore may have been non-essential, and represents a deletion, in the ancestral recombinant.

Structural gene and polyprotein analysis

The only structural protein of HJV (and WEEV) derived from EEEV is the capsid, which has been recently postulated to have been a necessary component of the

recombination event in order for the ancestral chimeric virus to inhibit cellular transcription (Garmashova *et al*., 2007). The capsid protein is well conserved among the three viruses, as the HJV capsid shares an 88% and 90% amino acid identity to EEEV and WEEV, respectively. Hong *et al*. (2006) reported that the alphavirus capsid can be divided into three regions (I, II, and III) based on structure and function. Region I contains a high level of basic amino acids involved in charge neutralization and nonspecific binding of the viral RNA. Within this region lies a conserved stretch of amino acids called helix I that contains leucine heptad repeats and functions as an essential checkpoint for capsid dimer formation and core assembly. The helix I structure in HJV, ^{34}-RPPLAAQIEDLRRSIANLTLKQR-^{56} (with the leucine heptad repeats underlined), is invariant to EEEV. WEEV contained two substitutions, (A49V and L53→F) relative to HJV and EEEV.

Region II of the capsid has been shown to be involved in binding to the encapsidation signal on the RNA, as well as containing the ribosome-binding site (Wengler *et al*., 1992; Owen and Kuhn, 1996; Hong *et al*., 2006). Capsid protein residues 93-109, believed to be important in both RNA encapsidation and nucleocapsid assembly/disassembly (Owen and Kuhn, 1996), are invariant between EEEV, WEEV, and HJV (^{93}-KPKPGKRQRMCMKLESD-^{109}). Region III of the capsid contains the protease domain and also contains the residues that interact with the C-terminal domain of the E2 glycoprotein (Hong *et al*., 2006). It was the structural and functional compatibility of these capsid residues in the ancestral EEEV with their glycoprotein counterparts in the SIN-like virus which gave rise to HJV, WEEV, and FMV. Previous analysis of WEEV identified seven amino acids in the capsid and four amino acids in the

E2 glycoprotein that had reverted to the amino acid residue of the alternate parental virus, suggesting that WEEV had undergone selection after the recombination event in order to maximize adaptation of the chimeric capsid-glycoprotein interaction (Hahn *et al*., 1988; Strauss, 1993). Of these 11 residues, only Ser233 in the capsid of HJV was not the same as the SINV residue (Thr233). As expected, the structural genes and proteins (other than the capsid) of HJV showed a high degree of identity with WEEV and minimal identity to EEEV (Table 3.1).

Although mutation frequencies for most RNA viruses are similar, arboviruses, and North American alphaviruses in particular, evolve at slower rates than most other RNA viruses (Weaver *et al*., 1991; Weaver *et al*., 1992; Powers *et al*., 2001). Tentative reasons for this restricted genetic diversity, as it relates to HJV, include 1) obligate dual host cycling (i.e., multiple selective constraints exists in having to adapt to replicate efficiently in two phylogenetically divergent hosts - passerine birds and *Cs. melanura*), 2) efficient virus dispersal (i.e., geographical movement of HJV by birds constrains genetic drift by facilitating dissemination of genotypes between populations), and 3) the restricted transmission season (i.e., the lack of mosquito activity during the temperate winter months restricts the amount of time HJV has to evolve in comparison to viruses in subtropical or tropical regions) (Cilnis *et al*., 1996).

Previous sequence analysis of 19 strains of HJV isolated over a period of 43 years estimated an evolutionary rate of 0.9 - 1.6 x 10^{-4} substitutions per nucleotide per year (substitutions/nt/yr), based on a 1,200 nt region covering the 3' UTR and a portion of the E1 gene (Cilnis *et al*., 1996). Comparison of the genomes between HJV 585-01, isolated in 2001, and HJV B-230, isolated in 1960, revealed a 99.0% (11,413/11,526) nucleotide

identity. Based on the entire genomes, the estimated evolutionary rate between HJV 585-01 and HJV B-230, was 2.28 x 10^{-4} substitutions/nt/yr, slightly higher than, but comparable to, previous estimates for alphaviruses endemic to North America (Weaver *et al*., 1993; Weaver *et al*., 1994; Cilnis *et al*., 1996). Comparison of the amino acid sequences of the nonstructural and structural polyproteins between HJV 585-01 and HJV B-230 revealed a 99.7% (2,442/2,450) and a 99.8% (1,234/1,236) identity, respectively, with an estimated evolutionary rate of 5.29 x 10^{-5} substitutions/aa/yr. The low level of genetic divergence at the nucleotide and amino acid level reiterates previous findings regarding the sequence conservation of HJV isolates and the comparative slower rates of evolution for alphaviruses relative to that of other RNA viruses (Cilnis *et al*., 1996).

Phylogenetic analysis of the structural polyprotein of HJV with that of other representative alphaviruses showed two major clades: viruses within the WEE serogroup (HJV, WEEV, FMV, SINV, AURAV) and those belonging to different antigenic complexes (EEEV, VEEV) (Fig. 3.4B, right). The WEE serocomplex viruses can be further divided into two groupings: those that are recombinant (WEEV, HJV, FMV) and those that are not (SINV, AURAV). The recombinant group in the structural polyprotein tree is then subdivided into the two lineages, HJV-WEEV and FMV (including Buggy Creek virus), as previously demonstrated by Weaver *et al*. (1997). Similarity plot analysis of the entire nonstructural and structural protein sequences of HJV indicated that the structural genes (other than the capsid) are more closely related to SINV than EEEV (Fig. 3.4A), as reported for WEEV (Hahn *et al*., 1988). All of the structural genes of HJV were identical in length to WEEV and ranged from 81%-90% in amino acid identity (Table 3.1).

3’ UTR analysis

The 3’ UTR in HJV was 379 nucleotides in length, being 17 and 75 nucleotides longer than EEEV and WEEV, respectively. Previous studies on the 3’ UTRs of different alphaviruses have identified a number of common features: 1) a highly conserved 19-nt CSE at the terminal 3’ end, which acts as a core promoter for negative-strand synthesis, 2) an AU-rich region upstream of the 19-nt CSE, and 3) a number of repeat sequence elements (RSEs) distributed throughout the 3’ UTR, which are variable in length, sequence, and abundance among different alphaviruses (Ou *et al*., 1981; Kuhn *et al*., 1990; Strauss and Strauss, 1994; Pfeffer *et al*., 1998). The function of the alphavirus RSEs is unknown, but they may function to bind cellular proteins during replication and/or translation (Strauss and Strauss, 1994). The terminal 3’ end of WEEV and HJV is believed to be derived from EEEV rather than the SIN-like virus, implying that the ancestral recombinant arose via a double cross-over event (Hahn *et al*., 1988; Strauss and Strauss, 1997). The exact location of the 3’ UTR crossover site for WEEV varies according to studies (Hahn *et al*., 1988; Pfeffer *et al*., 1998) or is hypothesized not to have occurred (Netolitzky *et al*., 2000).

The 3’ UTR in HJV and WEEV shared a high degree of nucleotide identity (82%) and similar secondary structural profiles (Fig. 3.6A-B), although there have been apparent deletions or insertions in WEEV or HJV, respectively, as the 3’ UTR in HJV is 75 nucleotides longer. Comparison of the HJV 585 3’ UTR with that of the HJV B-230 revealed a 98% nucleotide identity (375/379); none of the four substitutions occurred in either of the two SINV-like 40-nt RSEs or in the 19-nt CSE (not shown). Similar to that previously reported by Pfeffer *et al*. (1998), HJV 585-01 contains an A at position -6 (3’

UTR position 373) in the 19-nt CSE and both HJV and WEEV contain two SINV-like 40-nt RSEs in their 3' UTRs (Fig. 3.6A-B).

Alignment of the 3' UTRs of HJV, WEEV, and EEEV with SINV revealed that all three viruses contain at least one SINV-like 40-nt repeat, which based on predicted secondary structure, forms a double-stem loop (DSL) (Fig. 3.6A-B). Both HJV and WEEV contain two DSLs (DSL-1 and DSL-2) similar to SINV, except that the length of sequence between DSL-1 and DSL-2 in SINV is elongated relative to HJV and WEEV. SINV contains a third 40-nt RSE (DSL-3) not found in either WEEV or HJV (Fig. 3.6A). Comparative alignment of HJV and WEEV with SINV indicates a high nucleotide identity up to positions 129 in WEEV or 130 in HJV (Fig. 3.6B), suggesting that the 3' UTR crossover event occurred beyond this position as reported by Pfeffer *et al*. (1998). Based on a comparison of the SINV or SINV-like 40-nt RSE sequence and their predicted secondary structures in SINV, WEEV, and HJV (DSL-2 only), the first four nucleotides (AAAA) of the RSE precedes the stem, and each of stem loops in the DSL are separated by three nucleotides (GAA) (Fig. 3.6A-B). Unlike WEEV, DSL-1 in HJV does not contain the lead-in four adenylate nucleotides (AAAA→AGCA) common to the SINV 40-nt RSE.

The predicted SINV-like RSE in EEEV (DSL-1) differs from HJV and WEEV in that it 1) is elongated in the number of nucleotides that are incorporated into the double stem loop (55 in EEEV, 38 in SINV, 38 in WEEV, 38 in HJV DSL-2 and 43 in HJV DSL-1), 2) does not contain the first 4 nucleotides of the SINV-like RSE (AAAA→CTAG), and 3) does not contain the GAA gap. The 3' UTR of EEEV contains five 16-nt RSEs (GYRGYGYAUAAKGCYG) which encompass nucleotides 27-40 of

the 40-nt SINV RSE (Figure 3.6B). In addition to the reported terminal end recombination cross-over in order to acquire the distal 3' UTR sequence of the EEEV genome (Hahn *et al*., 1988), it is of interest that EEEV and SINV contain RSEs of similar sequence throughout their 3' UTRs, as this similarity may have additionally facilitated the recognition and binding of the polymerase complex (i.e., the nonstructural proteins derived from EEEV) to the 3' UTR in order for the ancestral recombinant to replicate efficiently. Besides EEEV, no other non-SINV-related alphavirus appears to share the same degree of homology in RSEs to SINV in their 3' UTRs, suggesting that the five 16-nt RSEs of EEEV that share identity to the 40-nt SINV RSE may have further enhanced the ability of the ancestral recombinant to survive in nature. Based on the sequence alignment, EEEV RSE-2 and -4 are equivalent to the second stem loop in DSL-1 and -2, respectively, in SINV, HJV, and WEEV. However, only two of the five EEEV RSEs (RSE-2 and -5) formed stem loop structures when the entire 3' UTR was used for secondary structure prediction. Similarly, a stem loop was predicted from HJV nucleotides 361-376 (corresponding to the 5' 16 nucleotides of the 19nt CSE), but not for WEEV, even though both viruses have identical sequences, suggesting that secondary structure predictions based on large sequences has some limitations.

Table 3.1. Comparison of the genomes of HJV, WEEV, and EEEV.

Genomic region	HJV[a]		WEEV				EEEV				EEEV-WEEV
	nt (aa)[b]	nt position	nt (aa)	ND[c]	% ID[d]	nt position	nt (aa)	ND	% ID	nt position	% ID
5' UTR	45	1-45	48	+3	95	1-48	46	+1	77	1-46	54
nsP1	1,599 (533)	46-1,644	1,599 (533)	0	80 (93)	49-1,647	1,599 (533)	0	75 (86)	47-1,645	75 (85)
nsP2	2,382 (794)	1,645-4,026	2,382 (794)	0	75 (89)	1,648-4,029	2,382 (794)	0	72 (84)	1,646-4,027	73 (84)
nsP3	1,545 (515)	4,027-5,571	1,596 (532)	+51 (+17)	69 (73)	4,030-5,625	1,677 (559)	+132 (+44)	54 (69)	4,028-5,704	66 (65)
nsP4	1,824 (608)	5,572-7,395	1,824 (608)	0	74 (90)	5,626-7,449	1,824 (608)	0	73 (88)	5,705-7,528	74 (88)
Junction UTR	44[e]	7,396-7,439	47	+3	80	7,450-7,496	66	+22	54	7,529-7,594	53
Capsid	777 (259)	7,440-8,216	777 (259)	0	78 (90)	7,497-8,273	783 (261)	+6 (+2)	75 (88)	7,595-8,377	74 (88)
E3	180 (60)	8,217-8,396	180 (60)	0	75 (90)	8,274-8,453	189 (63)	+9 (+3)	51 (56)	8,378-8,566	57 (53)
E2	1,269 (423)	8,397-9,665	1,269 (423)	0	72 (85)	8,454-9,722	1,260 (420)	-9 (-3)	47 (45)	8,567-9,826	52 (45)
6K	165 (55)	9,666-9,830	165 (55)	0	78 (81)	9,723-9,887	171 (57)	+6 (+2)	47 (40)	9,827-9,997	55 (43)
E1	1,317 (439)	9,831-11,147	1,317 (439)	0	77 (89)	9,888-11,204	1,323 (441)	+6 (+2)	57 (51)	9,998-11,320	58 (50)
3' UTR	379	11,148-11,526	304	-75	82	11,205-11,508	362[f]	-17	55	11,321-11,682	61
total	11,526 (3,686)		11,508 (3,703)	-18 (+17)	75 (87)		11,682 (3,736)	+156 (+50)	66 (73)		68 (73)

[a]Comparisons were generated using the following NCBI accession numbers: HJV (FJ827631), WEEV (AF214039/AF214040), and EEEV (EF151502).
[b]The length of the gene or non-coding region in nucleotides (nt), with, where applicable, the number of deduced amino acids (aa) in parentheses.
[c]ND = numerical differential, expressed as the difference in the number of nucleotides versus the cognate genomic region of HJV, with, where applicable, the corresponding differential in the number of amino acids in parentheses.
[d]% ID = percent identity, expressed as the CLUSTALW pairwise alignment score, calculated as the number of identities between the best alignment of the two sequences divided by the number of residues compared, with gap positions excluded (Chenna et al., 1996).
[e]Nucleotide sequences encoding stop codons at the end of the nonstructural and structural polyproteins are included in the junction UTR and 3' UTR, respectively.
[f]The poly(A) tail of EEEV was removed for alignment purposes.

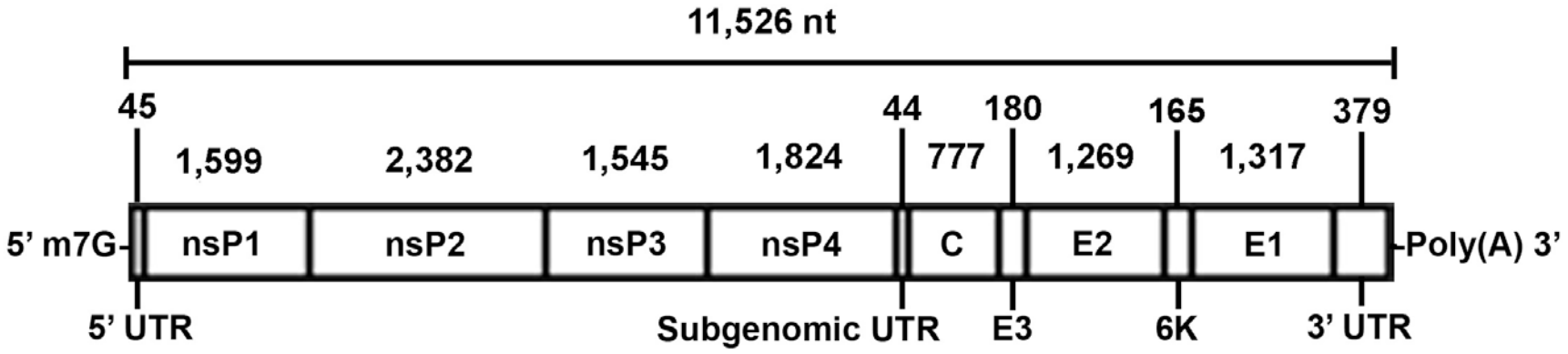


Figure 3.1. Schematic organization of the HJV genome. Excluding the 5' cap nucleotide and the 3' poly(A) tail, the HJV genome was 11,526 nucleotides (nt) in length. The nonstructural polyprotein gene sequence was 7,350 nt long, encoding 2,450 amino acids (aa), and represented 63.8% of the genome. The structural polyprotein gene sequence was approximately half that size, containing 3,708 nt that encoded 1,236 aa and occupied 32.2% of the genome. The three non-coding regions, the 5' UTR, the subgenomic junction UTR, and the 3' UTR, were 45, 44, and 379 nt in length, respectively, and collectively occupied 4% of the genome. The size of each gene and UTR, in nt, is indicated.

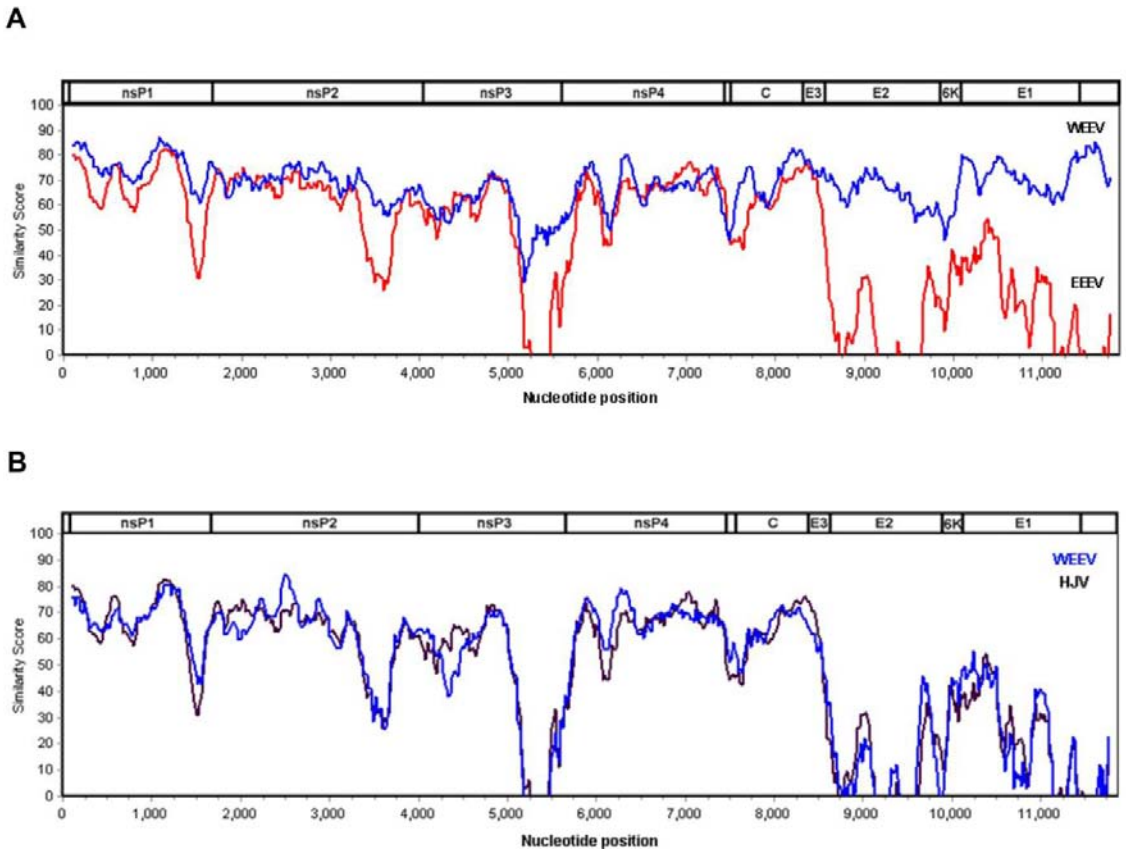


Figure 3.2. Genomic relatedness of HJV, WEEV, and EEEV. Genomic sequences used to generate similarity plots were HJV (FJ827631), EEEV (EF151502), and WEEV (AF214039/AF214040). (A) Lines for WEEV (blue) and EEEV (red) represent percent nucleotide similarity to HJV. A schematic representation of the HJV genome is shown to scale above the graph. (B) Lines for WEEV (blue) and HJV (black) represent percent nucleotide similarity to EEEV. A schematic representation of the EEEV genome is shown to scale above the graph.

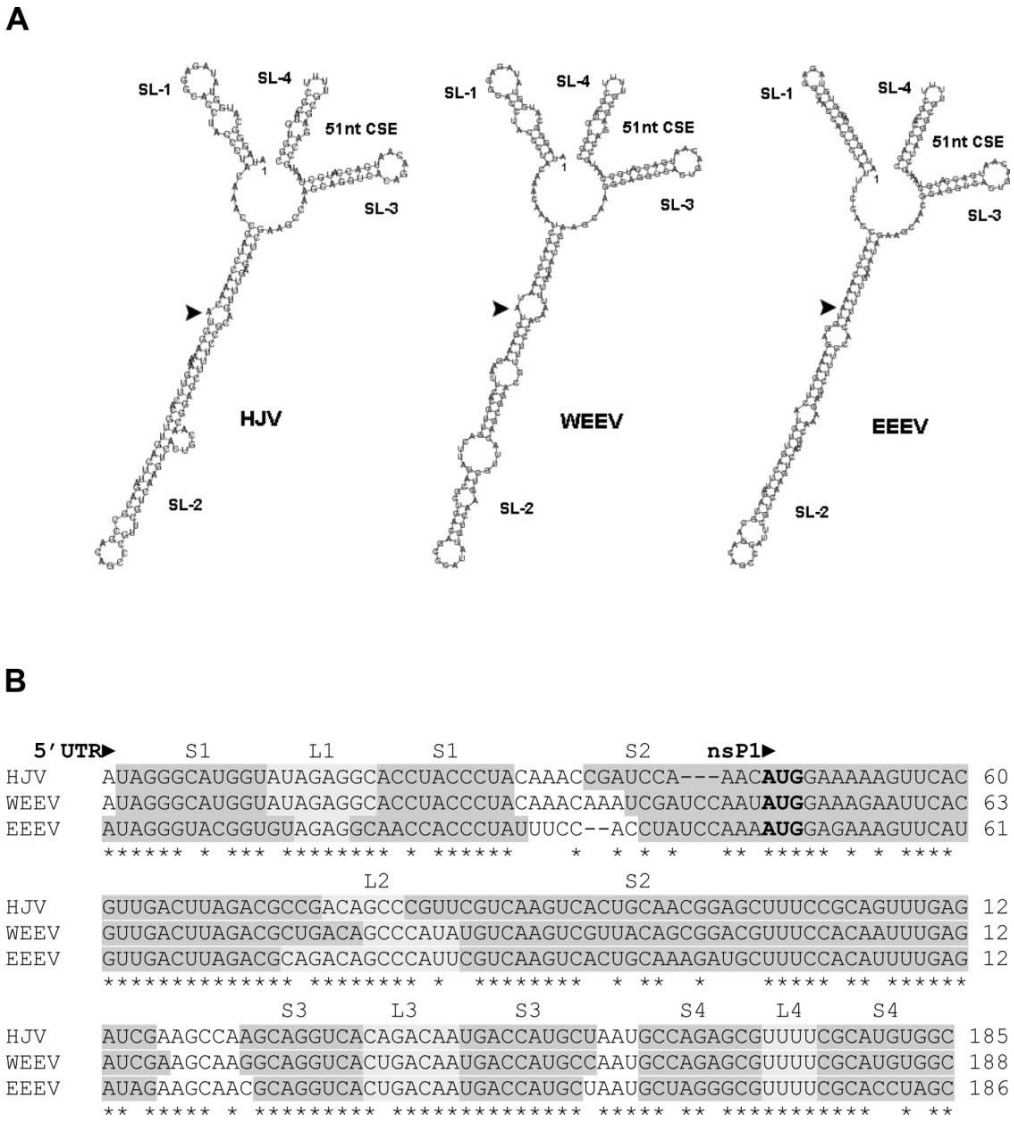
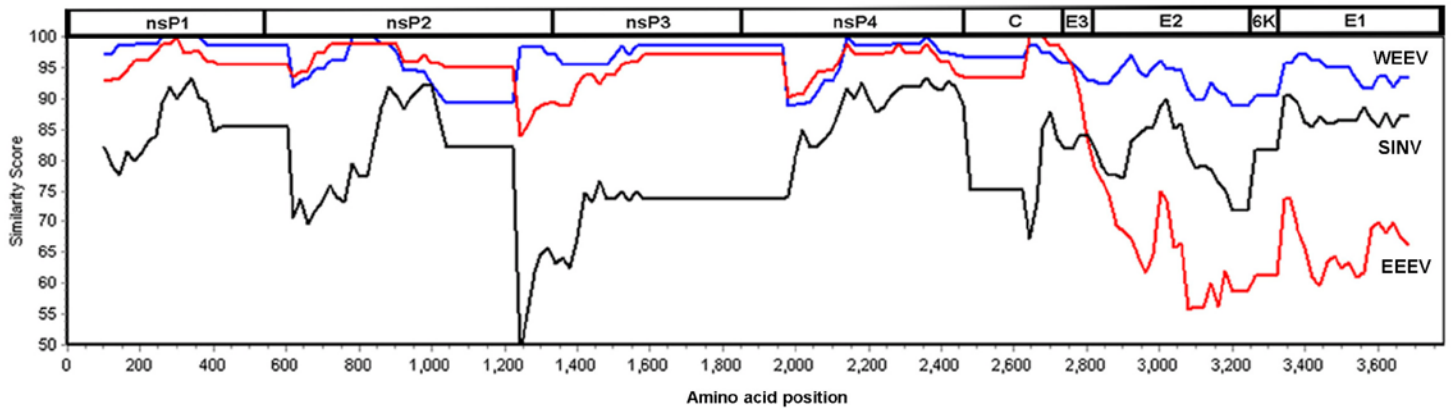


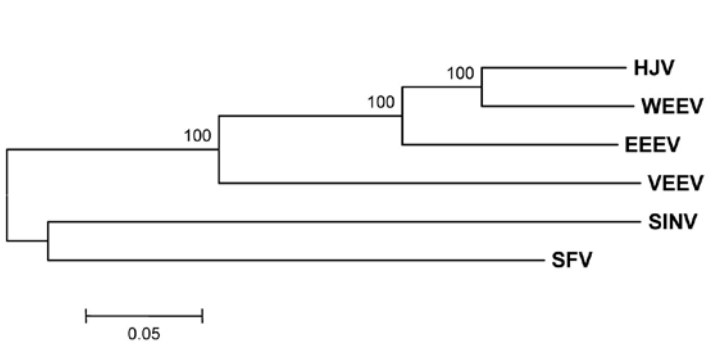
Figure 3.3. Comparison of the 5’ UTR and the region of the nsP1 gene encompassing the 51-nucleotide (nt) conserved sequence element (CSE) between HJV, WEEV, and EEEV in positive-sense orientation. (A) Predicted secondary structures for HJV, WEEV, and EEEV. Each of the four stem loops (SL) are noted, with SL-3 and SL-4 collectively forming the 51-nt CSE. The first nucleotide (1) in the 5’ UTR is identified. The nsP1 start codon is indicated by an arrowhead. (B) Pairwise alignment of the nucleotide sequences for HJV, WEEV, and EEEV. Dark-shaded nucleotides represent predicted stems (S); light-shaded nucleotides represent terminal stem loops (L); non-shaded nucleotides represent nucleotides not incorporated into the SL structure. Each SL is identified as an upward stem, the terminal loop(s), and a downward stem (e.g., S1-L1-S1). Asterisks denote identity. The positions of the 5’ UTR and nsP1 are indicated with arrowheads. The nsP1 start codon is in bold. Numbering to the right of the alignment indicates the respective genomic positions for each virus.

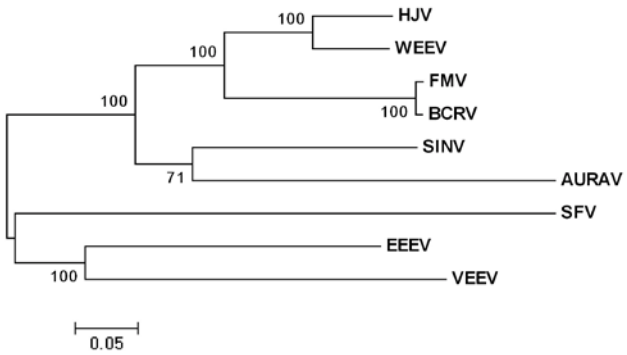
Figure 3.4. Evolutionary relationship of HJV to selected alphaviruses based on the amino acid sequences of the nonstructural and structural proteins. (A) Simplot analysis of the nonstructural and structural protein sequences of HJV, EEEV, WEEV, and SINV. Lines represent the percent similarity to HJV. A schematic representation of the protein organization of HJV is shown to scale above the graph. Sequences used to construct similarity plots were HJV 585-01 (FJ827631), WEEV (AF214040), EEEV (EF151502), and SINV (NC_001547). (B) Neighbor-joining phylogenies of the nonstructural (nsP1-nsP2-nsP3-nsP4) (left) and structural (C-E3-E2-6K-E1) (right) polyproteins. Evolutionary distances are in units of the number of amino acid substitutions per site. Numbers at each node represent bootstrap values (2000 replicates). Sequences used to construct the nonstructural polyprotein phylogenetic tree were HJV 585-01 (FJ827631), WEEV (AF214040), EEEV (EF151502), VEEV (AAD14558), SINV (NC_001547), and SFV (CAB62256). Sequences used to construct the structural polyprotein phylogenetic tree were HJV 585-01 (FJ827631), HJV B-230 (AAO33323), WEEV (AF214040), FMV (AAO33321), BCRV (AAO33319), SINV (NC_001547), AURAV (NP_632024), EEEV (EF151502), and VEEV (AAD14553). [HJV: *Highlands J virus*; WEEV: *western equine encephalitis virus*; EEEV: *eastern equine encephalitis virus*; VEEV: *Venezuelan equine encephalitis virus*; SINV: *Sindbis virus*; SFV: *Semliki Forest virus*; FMV: *Fort Morgan virus*; BCRV (not an ICTV abbreviation): Buggy Creek virus; AURAV: *Aura virus*].

A


nsP1
nsP2
nsP3
nsP4
C
E3
E2
6K
E1
WEEV
SINV
EEEV
Similarity Score
Amino acid position

B


HJV
WEEV
EEEV
VEEV
SINV
SFV
0.05


HJV
WEEV
FMV
BCRV
SINV
AURAV
SFV
EEEV
VEEV
0.05

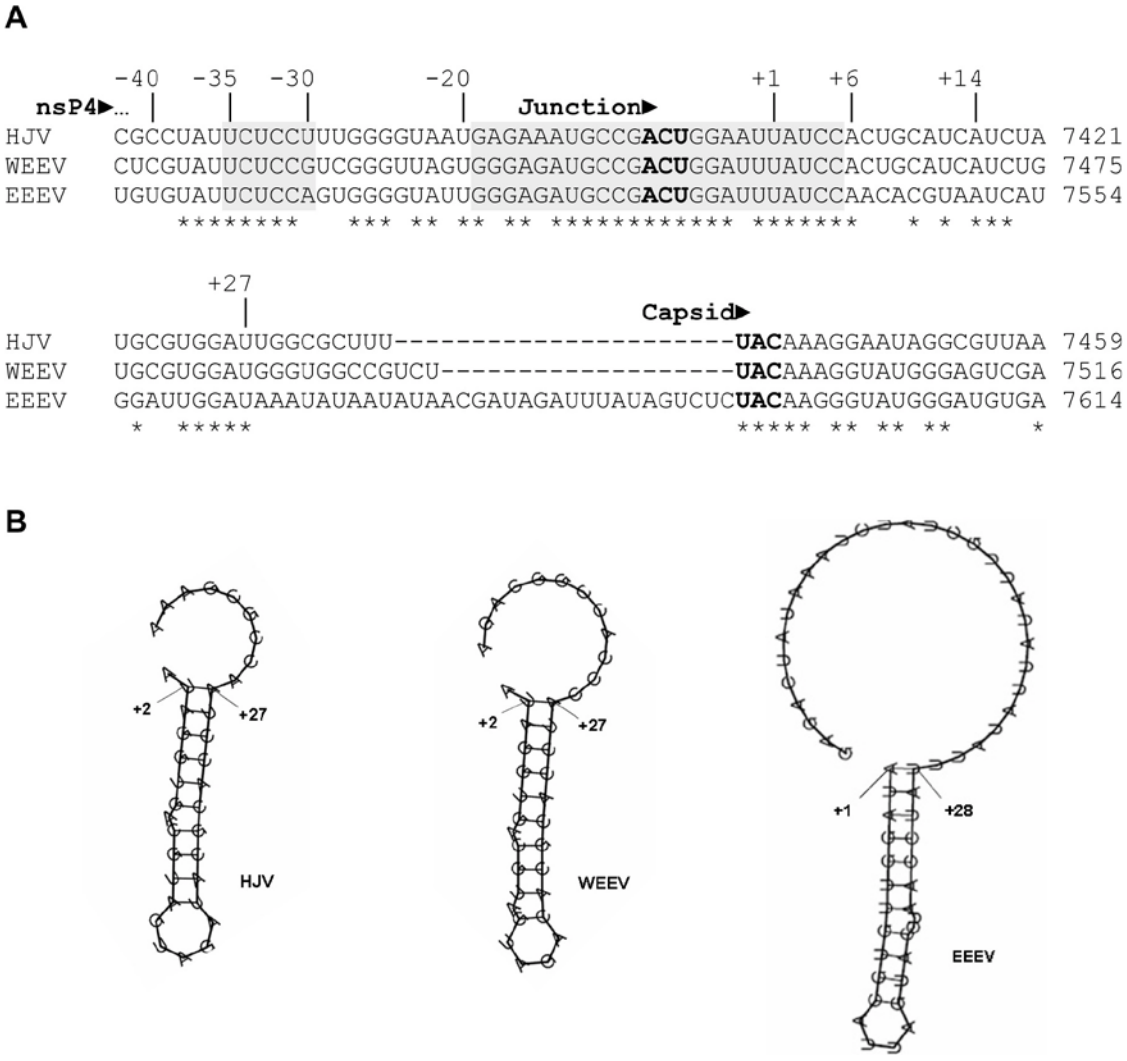
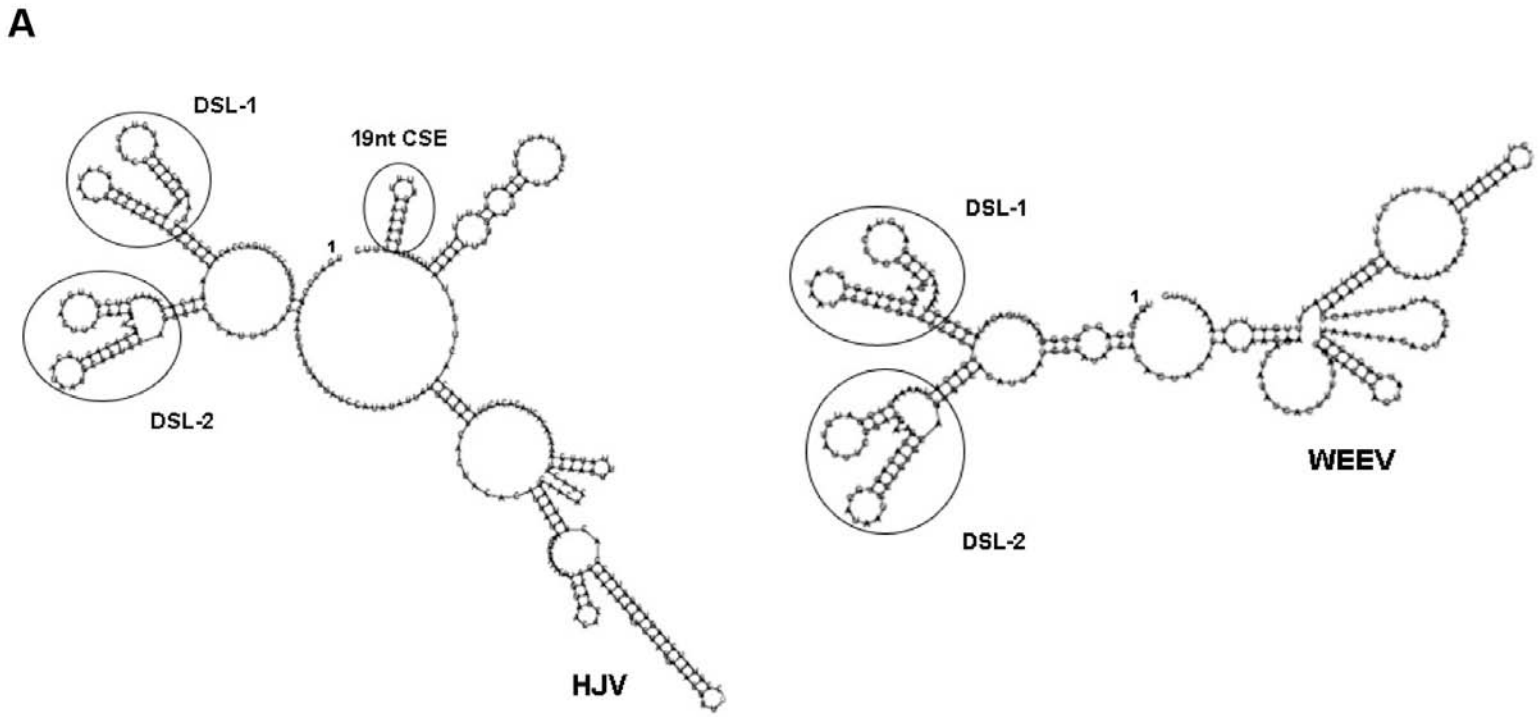
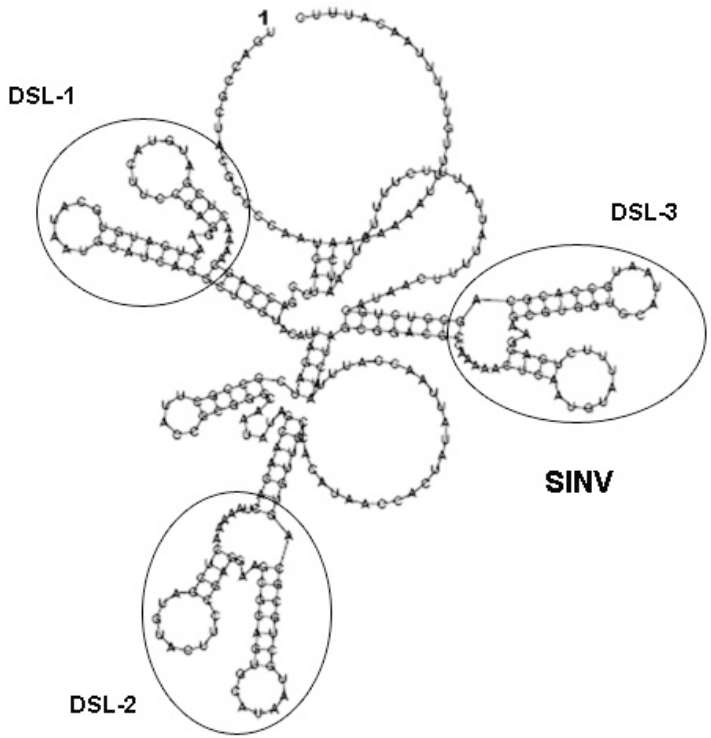
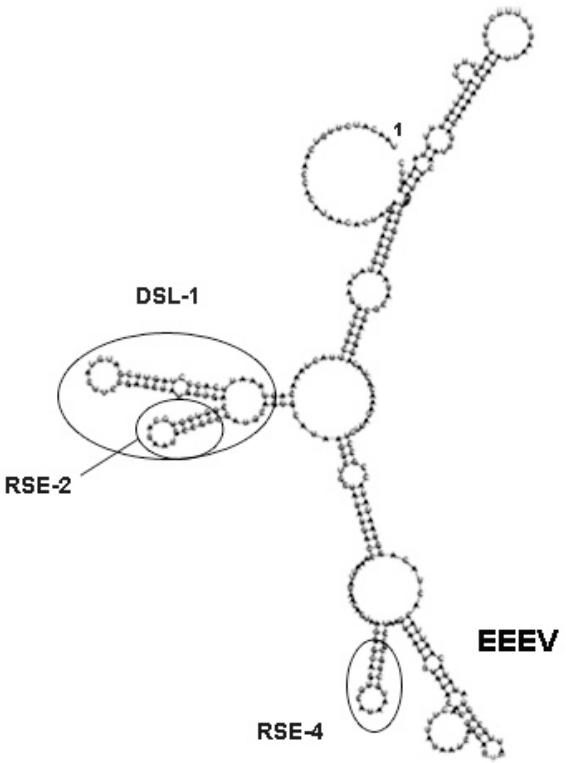


Fig. 3.5. Comparative analysis of the subgenomic UTR of HJV, WEEV, and EEEV in negative-sense orientation. (A) Nucleotide alignment of the subgenomic UTR and promoter region. Nucleotide positions (barring position +27) important in the transcription of the subgenomic RNA, relative to the initiation site (+1), are indicated above the alignment as per Wielgosz *et al.* (2001). Numbering to the right of the alignment indicates the respective genomic positions for each virus. Asterisks denote identity. The complements of the termination codon for nsP4 and the start codon for the capsid are in bold, with the beginning of the junction UTR and capsid indicated with arrowheads. The two regions in the subgenomic promoter that are conserved among the alphaviruses (-35 to -30 and -19 to +5) are shaded. The two additional regions required for high levels of transcription of the subgenomic RNA (-40 to -20 and +6 to +14) are noted. (B) Predicted secondary structures of the subgenomic RNA (in positive-sense) from position +1 to the end of the subgenomic UTR. Positions that are incorporated into the 5' terminal stem loop of the subgenomic RNA for each virus are noted.

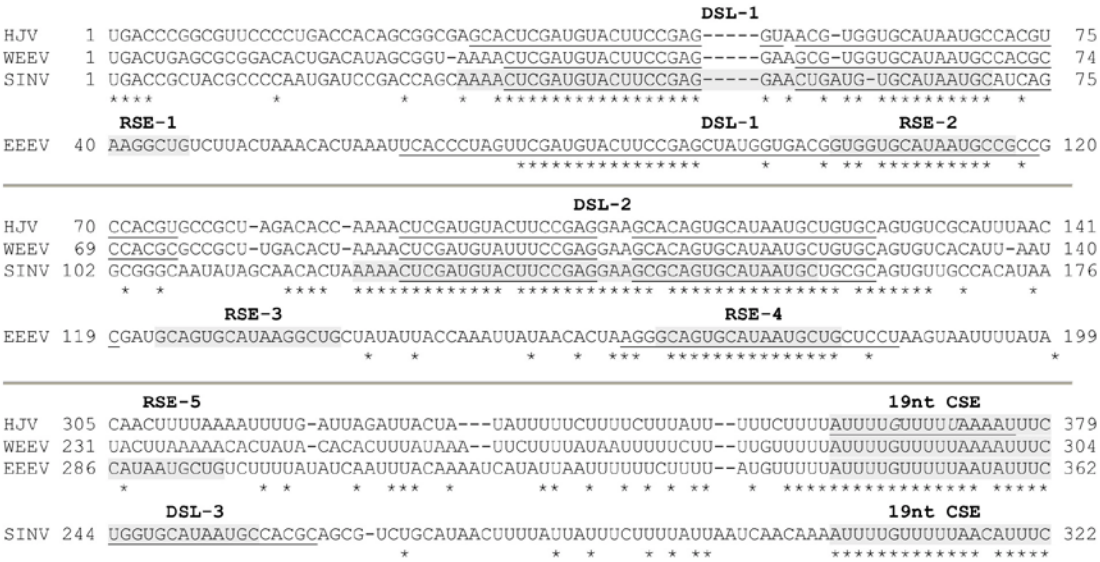
Fig. 3.6. Predicted secondary structure and nucleotide alignment of the 3' UTRs of HJV, WEEV, EEEV, and SINV. (A) The putative secondary structures of the 3' UTR of HJV, WEEV, SINV, and EEEV are shown. Conserved repeat sequence elements (RSEs) that form double stem-loop (DSL) structures are circled. Three single stem loop (SL) structures, the 3' terminal SL encompassing the 19-nucleotide (nt) conserved sequence element (CSE) in HJV, as well as EEEV RSE-2 and -5, are also circled. The first nucleotide (1) in the 3' UTR is noted. (B) Pairwise alignment of the nucleotide sequences of the 3' UTR of HJV, WEEV, EEEV, and SINV. The alignments are discontinuous and separated in order to highlight RSEs common to all four viruses. Asterisks denote nucleotide identity between the three or four viruses as shown in the upper and lower portion of each panel, respectively. The three 40-nt RSEs of SINV (structurally predicted to form DSLs) are shaded; note that the third 40-nt SINV RSE (DSL-3) is truncated for alignment purposes. Underlined portions in each alignment represent sequences that are predicted to be incorporated into each DSL or SL, starting at the base of each individual stem. Gaps between each SL in the DSL are not underlined (e.g., GAA). The five 16-nt RSEs of EEEV (GYRGYGYAUAAKGCYG) are shaded; EEEV RSE-1 and -5 are truncated for alignment purposes. Note that the EEEV RSE encompasses nucleotides 27-40 of the SINV RSE. The shaded area covering the 3' terminal 19 nucleotides represents the 19-nt CSE; the underlined portion of the shaded area covering nucleotides 361-376 in HJV represents the predicted terminal SL structure, with the mismatched G-U base-pairing italicized.


A
DSL-1
19nt CSE
1
DSL-2
HJV
DSL-1
1
WEEV
DSL-2


1
DSL-1
DSL-3
SINV
DSL-2


1
DSL-1
RSE-2
EEEV
RSE-4

B


DSL-1
HJV 1 UGACCCGGCGUUCCCCUGACCACAGCGGCGAGCACUCGAUGUACUUCCGAG-----GUAACG-UGGUGCAUAAUGCCACGU 75
WEEV 1 UGACUGAGCGCGGACACUGACAUAGCGGU-AAAACUCGAUGUACUUCCGAG-----GAAGCG-UGGUGCAUAAUGCCACGC 74
SINV 1 UGACCGCUACGCCCCAAUGAUCCGACCAGCAAAACUCGAUGUACUUCCGAG-----GAACUGAUG-UGCAUAAUGCAUCAG 75
RSE-1
DSL-1
RSE-2
EEEV 40 AAGGCUGUCUUACUAAACACUAAAUUCACCCUAGUUCGAUGUACUUCCGAGCUAUGGUGACGGUGGUGCAUAAUGCCGCCG 120
DSL-2
HJV 70 CCACGUGCCGCU-AGACACC-AAAACUCGAUGUACUUCCGAGGAAGCACAGUGCAUAAUGCUGUGCAGUGUCGCAUUUAAC 141
WEEV 69 CCACGCGCCGCU-UGACACU-AAAACUCGAUGUAUUUCCGAGGAAGCACAGUGCAUAAUGCUGUGCAGUGUCACAUU-AAU 140
SINV 102 GCGGGCAAUAUAGCAACACUAAAAACUCGAUGUACUUCCGAGGAAGCGCAGUGCAUAAUGCUGCGCAGUGUUGCCACAUAA 176
RSE-3
RSE-4
EEEV 119 CGAUGCAGUGCAUAAGGCUGCUAUAUUACCAAAUUAUAACACUAAGGGCAGUGCAUAAUGCUGCUCCUAAGUAAUUUUAUA 199
RSE-5
19nt CSE
HJV 305 CAACUUUUAAAAUUUUG-AUUAGAUUACUA---UAUUUUUCUUUUCUUUAUU--UUUCUUUUAUUUUGUUUUUAAAAUUUC 379
WEEV 231 UACUUAAAAACACUAUA-CACACUUUAUAAA--UUCUUUUAUAAUUUUUCUU--UUGUUUUUAUUUUGUUUUUAAAAUUUC 304
EEEV 286 CAUAAUGCUGUCUUUUAUAUCAAUUUACAAAAUCAUAUUAAUUUUUUCUUUU--AUGUUUUUAUUUUGUUUUUAAUAUUUC 362
DSL-3
19nt CSE
SINV 244 UGGUGCAUAAUGCCACGCAGCG-UCUGCAUAACUUUUAUUAUUUCUUUUAUUAAUCAACAAAAUUUUGUUUUUAACAUUUC 322

Acknowledgements

The authors would like to thank Drs. Robert Tesh and Scott Weaver at the Center for Biodefense and Emerging Infectious Diseases, University of Texas Medical Branch, Galveston, Texas, for providing HJV B-230 and for helpful suggestions regarding the manuscript, respectively. Funds were provided through sponsorship from the fish and wildlife agencies of Alabama, Arkansas, Florida, Georgia, Kansas, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Puerto Rico, South Carolina, Tennessee, Virginia, and West Virginia; through the Federal Aid to Wildlife Restoration Act (50 Stat. 917) and Grant Agreement 06ERAG0005, Biological Resources Division, U.S. Geological Survey, U.S. Department of the Interior; and through Cooperative Agreements 0596130032CA and 0696130032CA, Veterinary Services, Animal and Plant Health Inspection Service, U.S. Department of Agriculture.

References

Andreadis, T. G., Anderson, J. F., Tirrell-Peck, S. J., 1998. Multiple isolations of eastern equine encephalitis and Highlands J viruses from mosquitoes (Diptera: Culicidae) during a 1996 epizootic in southeastern Connecticut. J. Med. Entomol. 35, 296-302.

Berben-Bloemheuvel, G., Kasperaitis, M. A., van Heugten, H., Thomas, A. A., van Steeg, H., Voorma, H. O., 1992. Interaction of initiation factors with the cap structure of chimaeric mRNA containing the 5'-untranslated regions of Semliki Forest virus RNA is related to translational efficiency. Eur. J. Biochem. 208, 581-587.

Blom, N., Gammeltoft, S., Brunak, S., 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. 294, 1351-1362.

Brault, A. C., Powers, A. M., Chavez, C. L., Lopez, R. N., Cachón, M. F., Gutierrez, L. F., Kang, W., Tesh, R. B., Shope, R. E., Weaver, S. C., 1999. Genetic and antigenic diversity among eastern equine encephalitis viruses from North, Central, and South America. Am. J. Trop. Med. Hyg. 61, 579-586.

Calisher, C. H, Karabatsos, N., Lazuick, J. S., Monath, T. P., Wolff, K. L., 1988. Reevaluation of the western equine encephalitis antigenic complex of alphaviruses (family Togaviridae) as determined by neutralization tests. Am. J. Trop. Med. Hyg. 38, 447-452.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., Thompson, J. D., 2003. Multiple sequence alignment with the CLUSTAL series of programs. Nucleic Acids Res. 31, 3497-3500.

Cilnis, M. J., Kang, W., Weaver, S. C., 1996. Genetic conservation of Highlands J viruses. Virology 218, 343-351.

Dalrymple, J. M., Young, O. P., Eldridge, B. F., Russell, P. K., 1972. Ecology of arboviruses in a Maryland freshwater swamp. III. Vertebrate hosts. Am. J. Epidemiol. 96, 129-140.

Darsie, R. F., Jr., Ward, R. A., 2005. Identification and geographical distribution of the mosquitoes of North America, north of Mexico. University Press of Florida, Gainesville, Florida.

Dé, I., Fata-Hartley, C., Sawicki, S. G., Sawicki, D. L., 2003. Functional analysis of nsP3 phosphoprotein mutants of Sindbis virus. J. Virol. 77, 13106-13116.

Deresiewicz, R. L., Thaler, S. J., Hsu, L., Zamani, A. A., 1997. Clinical and neuroradiographic manifestations of eastern equine encephalitis. N. Engl. J. Med. 336, 1867-1874.

Edman, J. D., Timperi, R., Werner, B., 1993. Epidemiology of eastern equine encephalitis in Massachusetts. J. Fla. Mosq. Control Assoc. 64, 84-96.

Eleazer, T. H., Hill, J. E., 1994. Highlands J virus-associated mortality in chukar partridges. J. Vet. Diagn. Invest. 6, 98-99.

Englund, J. A., Breningstall, G. N., Heck, L. J., Lazuick, J. S., Karabatsos, N., Calisher, C. H., Tsai, T. F., 1986. Diagnosis of western equine encephalitis in an infant by brain biopsy. Pediatr. Infect. Dis. 5, 382-384.

Fenner, F., Gibbs, E. P. J., Murphy, F.A., Rott, R., Studdert, M.J., White, D.O. (ed.), 1993. Veterinary Virology. Academic Press, Orlando, Florida.

Ficken, M. D., Wages, D. P., Guy, J. S., Quinn, J. A., Emory, W. H., 1993. High mortality of domestic turkeys associated with Highlands J virus and eastern equine encephalitis virus infections. Avian Dis. 37, 585-590

Frolov, I., Schlesinger, S., 1996. Translation of Sindbis virus mRNA: analysis of sequences downstream of the initiating AUG codon that enhance translation. J. Virol. 70, 1182-1190.

Frolov, I., Hardy, R., Rice, C. M., 2001. Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. RNA 7, 1638-1651.

Galbraith, S. E., Sheahan, B. J., Atkins, G. J., 2006. Deletions in the hypervariable domain of the nsP3 gene attenuate Semliki Forest virus virulence. J. Gen. Virol. 87, 937-947.

Garmashova, N., Gorchakov, R., Volkova, E., Paessler, S., Frolova, E., Frolov, I., 2007. The Old World and New World alphaviruses use different virus-specific proteins for induction of transcriptional shutoff. J. Virol. 81, 2472-2484.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., Bairoch, A., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784-3788.

George, J., Raju, R., 2000. Alphavirus RNA genome repair and evolution: molecular characterization of infectious Sindbis virus isolates lacking a known conserved motif at the 3' end of the genome. J. Virol. 74, 9776-9785.

Guy, J. S., Ficken, M. D., Barnes, H. J., Wages, D. P., Smith, L. G., 1993. Experimental infection of young turkeys with eastern equine encephalitis virus and highlands J virus. Avian Dis. 37, 389-395.

Guy, J. S., Barnes, H. J., Smith, L. G., 1994. Experimental infection of young broiler chickens with eastern equine encephalitis virus and Highlands J virus. Avian Dis. 38, 572-582.

Hardy, W. R., Strauss, J. H., 1989. Processing the nonstructural polyproteins of Sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans. J. Virol. 63, 4653-4664.

Hahn, C. S., Lustig, S., Strauss, E. G., Strauss, J. H., 1988. Western equine encephalitis virus is a recombinant virus. Proc. Natl. Acad. Sci. U. S. A. 85, 5997-6001.

Hayes, C. G., Wallis, R. C., 1977. Ecology of western equine encephalomyelitis in the eastern United States. Adv. Virus Res. 21, 37-83.

Hayes, R. O., 1981. Eastern and western encephalitis, p. 29-57 *In* Beran, G. W. (ed.), CRC Handbook Series in Zoonoses, Section B: Viral Zoonoses, Volume I. CRC Press, Inc., Boca Raton, Florida.

Hofacker, I. L., 2003. Vienna RNA secondary structure server. Nucleic Acids Res. 31, 3429-3431.

Hoff, G. L., Bigler, W. J., Buff, E. E, Beck, E., 1978. Occurrence and distribution of western equine encephalomyelitis in Florida. J. Am. Vet. Med. Assoc. 172, 351-352.

Hong, E. M., Perera, R., Kuhn, R. J., 2006. Alphavirus capsid protein helix I controls a checkpoint in nucleocapsid core assembly. J. Virol. 80, 8848-8855.

Jennings W. L., Allen R. H., Lewis, A. L., 1966. Western equine encephalomyelitis in a Florida horse. Am. J. Trop. Med. Hyg. 15, 96-97.

Jennings, W. L., Kathein, R. A., Lewis, A. L., Bond, J. O., 1967. Encephalitis in horses in Hillsborough County, Florida, 1964. Am. J. Vet. Res. 28, 965-970.

Karabatsos, N., Lewis, A. L., Calisher, C. H., Hunt, A. R., Roehrig, J. T., 1988. Identification of Highlands J virus from a Florida horse. Am. J. Trop. Med. Hyg. 39, 603-606.

Kondig, J. P., Turell, M. J., Lee, J. S., O'Guinn, M. L., Wasieloski, L. P., Jr., 2007. Genetic analysis of South American eastern equine encephalomyelitis viruses isolated from mosquitoes collected in the Amazon Basin region of Peru. Am. J. Trop. Med. Hyg. 76, 408-416.

Kuhn, R. J., Hong, Z., Strauss, J. H., 1990. Mutagenesis of the 3' nontranslated region of Sindbis virus RNA. J. Virol. 64, 1465-1476.

LaStarza, M. W., Lemm, J. A., Rice, C. M., 1994. Genetic analysis of the nsP3 region of Sindbis virus: evidence for roles in minus-strand and subgenomic RNA synthesis. J. Virol. 68, 5781-5791.

Levis, R., Schlesinger, S., Huang, H. V., 1990. Promoter for Sindbis virus RNA-dependent subgenomic RNA transcription. J. Virol. 64, 1726-1733.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W., Ray, S. C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152-160.

McLean, R. G., Frier, G., Parham, G. L., Francy, D. B., Monath, T. P., Campos, E. G., Therrien, A., Kerschner, J., Calisher, C. H., 1985. Investigations of the vertebrate hosts of eastern equine encephalitis during an epizootic in Michigan, 1980. Am. J. Trop. Med. Hyg. 34, 1190-1202.

Mitchell, C. J., Monath, T. P., Sabattini, M. S., Daffner, J. F., Cropp, C. B., Calisher, C. H., Darsie, R. F. Jr., Jakob, W. L., 1987. Arbovirus isolations from mosquitoes collected during and after the 1982-1983 epizootic of western equine encephalitis in Argentina. Am. J. Trop. Med. Hyg. 36, 107-113.

Netolitzky, D. J., Schmaltz, F. L., Parker, M. D., Rayner, G. A., Fisher, G. R., Trent, D. W., Bader, D. E., Nagata, L. P., 2000. Complete genomic RNA sequence of western equine encephalitis virus and expression of the structural genes. J. Gen. Virol. 81:151-159.

Niesters, H. G., Strauss, J. H., 1990. Mutagenesis of the conserved 51-nucleotide region of Sindbis virus. J. Virol. 64, 1639-1647.

Ou, J. H., Strauss, E. G., Strauss, J. H., 1981. Comparative studies of the 3'-terminal sequences of several alphavirus RNAs. Virology 109, 281-289.

Ou, J. H., Trent, D. W., Strauss, J. H., 1982a. The 3'-non-coding regions of alphavirus RNAs contain repeating sequences. J. Mol. Biol. 156, 719-730.

Ou, J. H., Rice, C. M., Dalgarno, L., Strauss, E. G., Strauss, J. H., 1982b. Sequence studies of several alphavirus genomic RNAs in the region containing the start of the subgenomic RNA. Proc. Natl. Acad. Sci. U. S. A. 79, 5235-5239.

Ou J. H., Strauss, E. G., Strauss, J. H., 1983. The 5'-terminal sequences of the genomic RNAs of several alphaviruses. J. Mol. Biol.168, 1-15.

Owen, K. E., Kuhn, R. J., 1996. Identification of a region in the Sindbis virus nucleocapsid protein that is involved in specificity of RNA encapsidation. J. Virol. 70, 2757-2763.

Peränen, J., Kääriäinen, L., 1991. Biogenesis of type I cytopathic vacuoles in Semliki Forest virus-infected BHK cells. J. Virol. 65, 1623-1627.

Pfeffer, M., Proebster, B., Kinney, R. M., Kaaden, O. R., 1997. Genus-specific detection of alphaviruses by a semi-nested reverse transcription-polymerase chain reaction. Am. J. Trop. Med. Hyg. 57, 709-18.

Pfeffer, M., Kinney, R. M., 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., Shirako, Y., Strauss, E. G., Kang, W., Strauss, J. H., Weaver, S. C., 2001. Evolutionary relationships and systematics of the alphaviruses. J. Virol. 75, 10118-101131.

Powers, A. M., Aguilar, P. V., Chandler, L. J., Brault, A. C., Meakins, T. A., Watts, D., Russell, K. L., Olson, J., Vasconcelos, P. F., da Rosa, A. T., Weaver, S. C., Tesh, R. B., 2006. Genetic relationships among Mayaro and Una viruses suggest distinct patterns of transmission. Am. J. Trop. Med. Hyg. 75, 461-469.

Przelomski, M. M., O'Rourke, E., Grady, G. F., Berardi, V. P., Markley, H. G., 1988. Eastern equine encephalitis in Massachusetts: a report of 16 cases, 1970-1984. Neurology 38, 736-739.

Randolph, K. D., Vanhooser, S. L, Hoffman, M., 1994. Western equine encephalitis virus in emus in Oklahoma. J. Vet. Diagn. Invest. 6, 492-493.

Raju, R., Hajjou, M., Hill, K. R., Botta, V., Botta, S., 1999. In vivo addition of poly(A) tail and AU-rich sequences to the 3' terminus of the Sindbis virus RNA genome: a novel 3'-end repair pathway. J. Virol. 73, 2410-2419.

Reisen, W. K., Monath, T. P., 1988. Western equine encephalomyelitis, p. 89-137 *In* Monath, T.P. (ed.), The Arboviruses: Epidemiology and Ecology, Vol. V. CRC Press, Inc., Boca Raton, Florida.

Sawicki, D. L., Perri, S., Polo, J. M., Sawicki, S. G., 2006. Role for nsP2 proteins in the cessation of alphavirus minus-strand synthesis by host cells. J. Virol. 80, 360-371.

Scott, T. W., Weaver, S. C., 1989. Eastern equine encephalomyelitis virus: epidemiology and evolution of mosquito transmission. Adv. Virus Res. 37, 277-328.

Strauss, E. G., Rice, C. M., Strauss, J. H., 1983. Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. Proc. Natl. Acad. Sci. U. S. A. 80, 5271-5275.

Strauss, J. H., 1983. Recombination in the evolution of RNA viruses. In: Emerging Viruses. Morse, S. S. (Ed.). Oxford University Press, New York, New York, pp.241-251.

Strauss, J. H., Strauss, E. G., 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. 58, 491-562.

Strauss, J. H., Strauss, E. G., 1997. Recombination in alphaviruses. Semin. Virol. 8, 85-94.

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

Tomar, S., Hardy, R. W., Smith, J. L., Kuhn, R. J., 2006. Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity. J. Virol. 80, 9962-9969.

Trent, D. W., Grant, J. A., 1980. A comparison of New World alphaviruses in the western equine encephalomyelitis complex by immunochemical and oligonucleotide fingerprint techniques. J. Gen. Virol. 47, 261-282.

Vasiljeva, L., Merits, A., Auvinen, P., Kääriäinen, L., 2000. Identification of a novel function of the alphavirus capping apparatus. RNA 5'-triphosphatase activity of Nsp2. J. Biol. Chem. 275, 17281-17287.

Vihinen, H., Saarinen, J., 2000. Phosphorylation site analysis of Semliki forest virus nonstructural protein 3. J. Biol. Chem. 275, 27775-27783.

Vihinen, H., Ahola, T., Tuittila, M., Merits, A., Kääriäinen, L., 2001. Elimination of phosphorylation sites of Semliki Forest virus replicase protein nsP3. J. Biol. Chem. 276, 5745-5752.

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

Wang, Y. F., Sawicki, S. G., Sawicki, D. L., 1994. Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA. J. Virol. 68, 6466-6475.

Waters, J. R., 1976. An epidemic of western encephalomyelitis in humans-Manitoba, 1975. Can. J. Public Health 67, S1:28-32.

Weaver, S. C., Scott, T. W., Rico-Hesse, R., 1991. Molecular evolution of eastern equine encephalomyelitis virus in North America. Virology 182, 774-784.

Weaver S. C., Rico-Hesse R., Scott T. W., 1992. Genetic diversity and slow rates of evolution in New World alphaviruses. Curr. Top. Microbiol. Immunol. 176, 99-117.

Weaver, S. C., Hagenbaugh, A., Bellew, L. A., Netesov, S. V., Volchkov, V. E., Chang, G. J., Clarke, D. K., Gousset, L., Scott, T. W., Trent, D. W., 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., Hagenbaugh, A., Bellew, L. A., Gousset, L., Mallampalli, V., Holland, J. J., Scott, T. W., 1994. Evolution of alphaviruses in the eastern equine encephalomyelitis complex. J. Virol. 68, 158-169.

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

Weaver, S. C., Frey, T. K., Huang, H. V., Kinney, R. M., Rice, C. M., Roehrig, J. T., Shope, R. E., Strauss, E. G., 2005. Family Togaviridae p. 466-483 *In* Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., Ball, L. A. (ed.), Virus Taxonomy: Eighth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, California,

Wengler, G., Würkner, D., Wengler, G., 1992. Identification of a sequence element in the alphavirus core protein which mediates interaction of cores with ribosomes and the disassembly of cores. Virology 191, 880-888.

Williams, J. E., Young, O. P., Watts, D. M., Reed, T. J. 1971. Wild birds as eastern (EEE) and western (WEE) equine encephalitis sentinels. J. Wildl. Dis. 7, 188-94.

Wielgosz, M. M., Raju, R., Huang, H. V., 2001. Sequence requirements for Sindbis virus subgenomic mRNA promoter function in cultured cells. J. Virol. 75, 3509-3519.

RECOMBINATION, PART II:

THE EVOLUTIONARY GENETICS OF RECOMBINANT VIRUSES IN THE WESTERN EQUINE ENCEPHALITIS ANTIGENIC COMPLEX OF ALPHAVIRUSES

[1]Allison, A. B., and D. E. Stallknecht. To be submitted to *Infection, Genetics, and Evolution.*

Abstract

Western equine encephalitis virus (WEEV), *Highlands J virus* (HJV), and *Fort Morgan virus* (FMV) are the sole representatives of the WEE antigenic complex of the genus *Alphavirus*, family *Togaviridae,* that are endemic to North America. All three viruses are descended from a recombination event, which is believed to have occurred in South America, involving *eastern equine encephalitis virus* (EEEV) and a *Sindbis* (SIN)-like virus. This recombination event gave rise to a chimeric alphavirus which acquired the nonstructural genes and a single structural gene (capsid) from the parental EEEV, while the surface glycoproteins encoded in the structural polyprotein were obtained from the SIN-like virus. This ancestral recombinant then subsequently radiated to give rise to the present-day HJV, WEEV, and FMV. In the United States, these three recombinant viruses have adapted to different transmission cycles, and, therefore, have occupied ecological niches distinct from each other, although HJV circulates in a transmission cycle apparently identical to that of the North American lineage of EEEV. Previously, the complete nucleotide sequences of two of these recombinant viruses, HJV and WEEV, have been described. In this report, we present the genomic sequence of the third recombinant virus, FMV, thereby allowing for a comparative analysis of the full-length genomes between the three viruses (FMV, HJV, and WEEV). Biological, ecological, and genetic relationships between the recombinants are discussed, along with an alternate theory regarding the recombination event and the diversification of HJV, WEEV, and FMV from the ancestral recombinant.

Introduction

The genus *Alphavirus* within the family *Togaviridae* is comprised of 29 arthropod-borne viruses that have a worldwide distribution (Weaver *et al*., 2005). Of these 29 viruses, four are endemic to North America: *eastern equine encephalitis virus* (EEEV), *western equine encephalitis virus* (WEEV), *Highlands J virus* (HJV), and *Fort Morgan virus* (FMV) (Weaver *et al*., 1997). Two additional alphaviruses found in the western United States, Buggy Creek virus and the recently described Stone Lakes virus, are regarded as variants of FMV (Hopla *et al*. 1993; Powers *et al*., 2001; Brault *et al*., 2009). WEEV, HJV, and FMV are unique among the alphaviruses in that they are recombinants; all three viruses are descendants of an ancestral alphavirus that was derived from a recombination event, which is believed to have occurred in the neotropics of South America, between a *Sindbis* (SIN)-like virus and EEEV (Strauss and Strauss, 1997). As the surface glycoproteins (E1 and E2) of the recombinant were obtained from the SIN-like virus, WEEV, HJV, and FMV are therefore antigenically related (i.e., in neutralization tests) to SINV rather than EEEV (Calisher *et al*., 1988). Along with *Aura virus* (AURAV) and *Whataroa virus* (WHAV), WEEV, HJV, FMV, and SINV collectively constitute the WEE antigenic complex (Weaver *et al*., 1997).

Other than the three recombinants, AURAV is the only other WEE complex member that is found in the New World, having been isolated from *Culex* and *Aedes* species of mosquitoes in Brazil and Argentina (Causey *et al*., 1963; Rumenapf *et al*., 1994; Rumenapf *et al*., 1995). As AURAV is endemic to South America and is related to SINV, it initially provided an attractive candidate as the putative SIN-like parental virus of the recombinants. However, as it has been demonstrated that SINV is more closely

related to WEEV than it is to AURAV (Rumenapf *et al*., 1995), and as AURAV and SINV are believed to have diverged prior to the estimated time frame of the recombination event (Weaver *et al*., 1997), is suggestive that EEEV did not recombine with AURAV, but rather with a virus more closely related to the present-day SINV. Currently, SINV is no longer found in the New World, but is widely distributed throughout Eurasia, Africa, and Australasia, presumably as a result of transoceanic bird migration (Griffin, 2007). On the other hand, EEEV is found only throughout the New World and represents the sole species constituting the EEE antigenic complex, although it has recently been suggested that North and South American lineages of EEEV should be reclassified as two distinct species (Arrigo *et al*., 2010).

In North America, EEEV, WEEV, HJV, and FMV all circulate in transmission cycles involving passerine birds as amplifying hosts and hematophagous arthropods as vectors; however, FMV is unique among the four viruses in that the normal invertebrate vector is the cimicid swallow bug (*Oeciacus vicarious*), rather than a mosquito species [i.e., *Culiseta* (*Cs*.) *melanura*: EEEV and HJV; *Culex* (*Cx*.) *tarsalis*: WEEV] (Calisher *et al*., 1980). The distribution of each of the recombinant alphaviruses in the United States is, for the most part, spatially discrete, and is essentially defined by the geographical range of their respective enzootic vectors. HJV is primarily confined to eastern states along the Gulf and Atlantic seaboard and the Great Lakes region (Cilnis *et al*., 1996), while WEEV and FMV are endemic throughout most of the western United States (Reisen and Monath, 1988; Pfeffer *et al*., 2006), although FMV is much more focally distributed than WEEV due to its unique transmission cycle (as outlined below). Interestingly, both HJV and EEEV share apparently identical transmission cycles in

North America and thus concomitantly share nearly identical geographical ranges (Scott and Weaver, 1989). Although lineages of WEEV and EEEV exist in Central and South America (Srihongse and Galindo, 1967; Mitchell *et al*., 1987; Weaver *et al*., 1994; Weaver *et al*., 1997; Brault *et al*., 1999), FMV and HJV have not been isolated outside of the United States. Additionally, unlike EEEV and WEEV, neither FMV nor HJV are normally associated with disease in mammalian hosts (Hayes and Wallis, 1977; Calisher *et al*., 1980; Englund *et al*., 1986; Przelomski *et al*., 1988), although all four viruses are avian pathogens to varying degrees (Ficken *et al*., 1993; Randolph *et al*., 1994; Huyvaert *et al*., 2008).

Fort Morgan virus was first isolated by Hayes *et al*. (1977) from *O. vicarius* in eastern Colorado. Since its initial description, FMV (along with Buggy Creek virus and Stone Lakes virus) has been reported from a number of additional western and central states including Nebraska, Oklahoma, Texas, North Dakota, South Dakota, Washington, and California (Calisher *et al*., 1980; Pfeffer *et al*., 2006; Hopla *et al*., 1993; Padhi *et al*., 2008; Brault *et al*., 2009). The primary vertebrate amplifying host for FMV are cliff swallows (*Petrochelidon pyrrhonata*), and to a lesser extent, house sparrows (*Passer domesticus*), with the latter being inadvertently involved in transmission as they are known to occupy cliff swallow nests (Calisher *et al*., 1980). As *O. vicarius* is a sedentary ectoparasite that strictly blood-feeds on cliff swallows (and house sparrows which parasitize cliff swallow nests), this likely, in part, restricts the geographical range of FMV to more discrete endemic foci relative to that of other North American alphaviruses, which are vectored endemically by motile mosquito species which may feed on multiple avian species (*Cs. melanura*) or are more catholic in host preference (*Cx. tarsalis*) (Loye,

1985). That *Cx. tarsalis* and *Cx. pipiens* have been demonstrated to be refractory to FMV infection following intrathoracic inoculation, and *Aedes (Ae.) albopictus* cell cultures do not support replication of FMV (Calisher *et al*., 1980), is suggestive that FMV has undergone considerable selection pressure upon adaptation to *O. vicarius* and, therefore, may be more genetically divergent from the mosquito-borne recombinant alphaviruses (HJV and WEEV) in both the structural and nonstructural polyprotein sequences.

The first alphavirus demonstrated to be a recombinant was WEEV (Hahn *et al*., 1988). Sequence analysis of WEEV suggested that it was derived from a double cross-over event that occurred (either simultaneously or separately) within E3 and the 3' UTR, such that the 5' two-thirds of the genome, along with the terminal end of the 3' UTR, were derived from EEEV, while the C terminal region of E3 to the upstream portion of the 3' UTR were derived from a *Sindbis* (SIN)-like virus (see Fig. 3.7) (Hahn *et al*., 1988; Weaver *et al*., 1993). This finding provided not only a genetic basis for WEEV having the encephalitic properties of EEEV, while being antigenically closely related to SINV, but also suggested that the *cis*-acting elements contained within both the 5' and 3' ends of EEEV may have been a necessary prerequisite for the efficient replication, and, hence, survival of the ancestral recombinant (Hahn *et al*., 1988; Strauss and Strauss, 1997). It is believed that the ancestral virus of all recombinant WEE complex viruses emerged in South America approximately 1,300 to 1,900 years ago (Weaver *et al*., 1997). This recombinant alphavirus then diversified and gave rise to the ancestors of the FMV and WEEV-HJV groups. WEEV and HJV diverged from one another an estimated 650 years ago, with WEEV subsequently diversifying into several independent lineages (Weaver *et al*., 1997). Whether WEEV, HJV, or FMV is the most ancestral recombinant,

or the ancestral virus which gave rise to all three is extinct, is speculative. As certain WEEV lineages include isolates from both South and North America, indicating transportation of these genotypes occurs readily between the two continents (Weaver *et al*., 1997), this might suggest that the extant virus which gave rise to WEEV, HJV, and FMV could have been a South American variety of WEEV or WEE-like virus. However, as both EEEV and HJV use *Cs. melanura* as an enzootic vector and passerine birds as vertebrate hosts, and as passerine birds are the vertebrate hosts for SINV, it is plausible that the recombination event between EEEV and a SIN-like virus gave rise to an ancestral HJV that radiated to produce WEEV and FMV.

Previously, the complete nucleotide sequences of two of the recombinants, WEEV and HJV, along with representative progenitors of the parental viruses, EEEV and SINV, have been described (Shirako *et al*., 1991; Weaver *et al*., 1993; Netolitzky *et al*., 2000; Allison and Stallknecht, 2009). Comparative genomic analysis of HJV and WEEV disclosed that they were genetically very similar to one another, sharing a 75 and 87% identity at the nucleotide and amino acid level, respectively (Allison and Stallknecht, 2009). Although the structural polyprotein (C-E3-E2-6K-E1) and 3' UTR of FMV CM4-146, along with the cognate region in numerous isolates of the variant Buggy Creek virus, have been sequenced (Pfeffer *et al*., 2006; Padhi *et al*., 2008), the complete genomic portion of FMV that was acquired from EEEV (i.e., the 5' UTR to the E3 cross-over site) has not been described previously, thus precluding a comparative genomic analysis of the three recombinant viruses (FMV, HJV, and WEEV). In this report, we present the complete nucleotide sequence of FMV. Comparative biological, ecological, and genetic

relationships of the three recombinant viruses are discussed, along with an alternate theory on the evolutionary history of the recombinant viruses.

Materials and Methods

Viruses

FMV isolate CM4-146, originally isolated from *O. vicarius* in Colorado in 1973, was obtained from the American Type Culture Collection (ATCC) (lot no. 1907186). CM4-146 was additionally passaged once in Vero cells for generation of stock virus used for genomic sequencing and cell culture assays. For *in vitro* studies, in addition to FMV CM4-146, other recombinant WEE serocomplex viruses analyzed were 1) WEEV SW-99, obtained from a pool of *Cx. tarsalis* from Arizona in 1999, 2) HJV 585-01, isolated from the brain of a red-tailed hawk (*Buteo jamaicensis*) from Georgia in 2001, 3) HJV 744-01, a small plaque variant isolated from the brain of a barred owl (*Strix varia*) from Georgia in 2001, and 4) HJV isolate B-230, the prototype strain of HJV recovered from the blood of a blue jay (*Cyanocitta cristata*) from Florida in 1960. Passage histories were P2 in Vero cells for HJV 585-01, HJV 744-01, and WEEV SW-99; P7 in suckling mouse brain(6)/Vero cells(1) for HJV B-230; and P8 in Vero cells for FMV CM4-146.

Comparative cell line growth kinetics of FMV, HJV, and WEEV

The *in vitro* host ranges of FMV, HJV, and WEEV were assessed in the following mammalian, avian, reptilian, fish, and mosquito cell lines: 1) A-549 (human lung carcinoma), 2) C6/36 [Asian tiger mosquito (*Aedes albopictus*) larvae], 3) CRFK [cat

(*Felis catus*) kidney cortex], 4) DF-1 [chicken (*Gallus gallus*) embryo fibroblasts], 5) FHM [Fathead minnow (*Pimephales promelas*) connective tissue and muscle epithelium], 6) HEK (human embryonic kidney), 7) LLC-MK2 (rhesus monkey (*Macaca mulatta*) kidney epithelium), 8) N2a [mouse (*Mus musculus*) neuroblastoma], 9) PDE [Pekin duck (*Anas platyrhynchus domesticus*) embryo fibroblasts], 10) QNR/D [Japanese quail (*Coturnix coturnix japonica*) neuroretinal cells], 11) QNR/K2 (Japanese quail Müller cells), 12) Tb 1 Lu [Mexican free-tailed bat (*Tadarida brasiliensis*) lung epithelium], 13) TH-1 [Eastern box turtle (*Terrapene carolina*) heart fibroblasts], 14) Vero [African green monkey (*Cercopithecus aethiops*) kidney epithelium], and 15) VH-2 [Russell's viper (*Vipera russelli*) heart fibroblasts]. All cell lines were obtained from the ATCC and were grown in minimum essential medium (MEM) supplemented with 2.2g/l $NaHCO_3$, 5-10% fetal bovine serum (FBS), 400 units/ml penicillin, 400µg/ml streptomycin, and 1µg/ml amphotericin (Sigma). Cells were maintained at 28 °C or 37 °C in a humidified atmosphere containing 5% CO_2. For infection studies, cells in a 12-well plate format (2mLs/well) were inoculated with 10^3 plaque forming units (PFU) of each virus and harvested daily for 6 days (28 °C) or at 12-hr time points for 3 days (37 °C). Titrations were performed on Vero cells overlaid with 1% gum tragacanth/1X MEM supplemented with 3% FBS and 400 units/ml penicillin, 400µg/ml streptomycin, and 1µg/ml amphotericin. Cultures were inactivated on day 2-4 post-adsorption with 10% buffered formalin and stained with 0.25% crystal violet for plaque visualization. Dilutions in which 20-100 plaques could be counted (when applicable) were used in determining titers ($\log_{10}$ PFU/mL).

In vitro persistence in quail neurons

During *in vitro* experiments, it was recognized that a subset of QNR/D cells (but not QNR/K2 cells) infected with HJV, WEEV, or FMV survived an initial lytic infection and persistently produced virus (characterized by a small plaque phenotype) in the absense of any cytopathology. As opposed to the QNR/K2 cell line, which is composed of Müller cells (i.e., the primary macroglial cell of the neuroretina), the QNR/D cell line is comprised of amacrine and ganglion cells and is therefore strictly neuronal in origin (Pessac *et al*., 1983). QNR/D cells, maintained at 37 °C with 5% CO_2, were infected with 10^2 PFUs of HJV 585-01, HJV 744-01, HJV B-230, FMV CM4-146, or WEEV SW-99 and culture supernatant was removed and assayed for virus production at 3- to 6-day intervals for 100 days. For HJV B-230, the effect of virus-specific antibody on persistence was additionally analyzed, whereby infected cells were incubated with a 1:10 dilution of HJV B-230 hyperimmune mouse ascitic fluid (ATCC) or 2µg/ml of a HJV B-230 monoclonal antibody, clone 2Q1968 (U. S. Biological Inc.) for a 9-day period from days 60-69 post-infection. At day 100 post-infection, HJV B-230 was harvested from culture supernatant and the genome of the small plaque phenotype was sequenced as described below and compared to the initial inoculum to determine the mutations that occurred during the course of infection. In order to determine if the nucleotide and amino acid changes responsible for the small plaque phenotype were uniform between two different isolates, the genome of HJV 744-01, a naturally-occurring small plaque phenotype (i.e., small plaques were observed from clarified supernatant of brain homogenate), was additionally sequenced.

Genomic sequencing

Viral RNA was extracted from stock FMV using a QIAamp® Viral RNA Mini kit (Qiagen) according to the manufacturer's instructions. In order to amplify the portion of the FMV genome encompassing the nonstructural protein genes, degenerate primers were designed from conserved regions of a CLUSTALW alignment of genomic sequences of HJV (FJ827631), EEEV (AY705240), and WEEV (NC003908). Gaps in the genome were subsequently filled in using FMV-specific primers based on sequence obtained from the degenerate primers. To verify the full-length nucleotide sequence of the structural genes, primers were developed based on the previously available structural polyprotein gene sequence of FMV CM4-146 (NCBI accession AF339475). A list of all primers used in sequencing of the FMV genome may be obtained from the authors. The 5' and 3' ends of the FMV genome were sequenced by RACE using a commercial kit (First Choice® RLM-RACE Kit, Ambion). RACE amplicons were cloned using a PCR Cloning^{Plus}® Kit (Qiagen) and purified using a QIAprep® Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. Sequencing of RT-PCR products or RACE clones was performed using a 3100 Genetic Analyzer (Applied Biosystems, Inc.). The complete genome of FMV CM4-146 was constructed from multiple overlapping contigs using Sequencher 4.1.4. (Gene Codes Corporation). The genomes for HJV isolates 744-01 and B-230 (day 100 post-infection in QNR/D cells) were sequenced as previously described (Allison and Stallknecht, 2009).

RNA and protein sequence analysis

FMV nucleotide and amino acid identities to HJV, WEEV, EEEV, and SINV (Ockelbo virus) were determined by CLUSTALW alignments using the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) server (http://www.ebi.ac.uk/Tools/clustalw2/index.html) (Chenna *et al*., 2003). All pairwise comparisons were made using the following NCBI accession numbers: FMV (GQ281603), HJV (FJ827631), WEEV (AF214039/40), EEEV (EF151502), and SINV (M69205). Predicted minimum energy secondary structures of FMV were generated using the RNAfold program using default parameters (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) (Hofacker, 2003). The molecular weights and isoelectric points of the FMV nonstructural proteins were determined using the Compute pI/Mw tool on the ExPASy server (http://www.expasy.ch/tools/pi_tool.html) (Gasteiger *et al*., 2003). Phylogenetic trees were generated with a CLUSTALW alignment of the complete nucleotide sequences of representative alphaviruses using the Molecular Evolutionary Genetics Analysis (MEGA) program (Tamura *et al*., 2007). Similarity plots were generated in the Simplot program (Lole *et al*., 1999) using the 2-parameter (Kimura) distance model in a sliding window of 200 nucleotides moved in steps of 20 nucleotides along the alignment. In addition to the similarity plots, to determine areas of recombination, the FMV genome was analyzed against representative alphavirus genomes using the bootscanning option in the Simplot program or the Recombination Analysis Tool server (http://jic-bioinfo.bbsrc.ac.uk/dicks/software/RAT/index.html) using default parameters (Etherington *et al*., 2004).

Results and Discussion

Genetic relationships of the recombinant WEEV serocomplex viruses

Genomic sequencing of FMV was undertaken in order to perform direct genetic comparisons of the full-length genomes between the three viruses (FMV, WEEV, HJV) known to be derived from the recombination event between EEEV and a SIN-like virus. As the structural polyprotein (C-E3-E2-6K-E1) and 3' UTR of FMV CM4-146, accounting for ~3.9kB, has previously been sequenced (Pfeffer *et al*., 1998), these genomic regions will not be discussed in any detail. Additionally, portions of nsP1 and nsP4 of FMV CM4-146 have also been sequenced (Pfeffer *et al*., 1997), although these only accounted for ~12% (0.9kB) of the nonstructural polyprotein sequence. Excluding the 5' cap nucleotide and the 3' poly(A) tail, the genome of FMV was 11, 381 nucleotides in length. The FMV genome was the shortest among the recombinants, being 127 or 145 nucleotides less than either WEEV or HJV, respectively, although this truncation was primarily due to the short 3' UTR of FMV (Fig. 3.7). Over the entire genome, FMV shared a 69% nucleotide and 78% amino acid identity to both HJV and WEEV. In comparison, HJV and WEEV share a 75 and 87% identity at the nucleotide and protein level, respectively (Allison and Stallknecht, 2009). Pairwise comparisons of all nine genes and proteins (nsP1-nsP2-nsP3-nsP4-C-E3-E2-6K-E1) between FMV, HJV, and WEEV indicated that, except for the 6K amino acid sequence, HJV and WEEV shared higher nucleotide and amino acid identities to each another than either did to FMV (Table 3.2); thus, WEEV and HJV are more similar genetically to one another than either

are to FMV. A genetic comparison of FMV, HJV, WEEV, along with EEEV and SINV, is shown in Table 3.2.

The five full-length gene sequences of the recombinants that were acquired from EEEV include the genes present in the nonstructural polyprotein (nsP1, nsP2, nsP3, and nsP4) and the capsid protein gene. Overall, the nonstructural polyprotein of FMV shared an 80% amino acid identity to both WEEV and HJV, and a 78% identity to EEEV. When comparing each of the nonstructural and capsid protein amino acid sequences of all three recombinants against EEEV, nsP4 was the most conserved (87-88%), followed by the capsid (84-88%), nsP1 (80-86%), nsP2 (81-84%), and nsP3 (61-69%) (Table 3.2).

The 5' UTR of FMV was 46 nucleotides long, with nucleotides 2-30 forming a predicted stem loop (SL) structure (SL-1; not shown), as previously demonstrated for HJV and WEEV (Allison and Stallknecht, 2009). The 51-nt conserved sequence element (CSE) of FMV, located within nsP1, covered nucleotide positions 135-186 and formed a double stem loop (SL-3 and SL-4) that was most similar to EEEV in secondary structure predictions as it contained the nucleotides G^{168} and C^{182}, and therefore did not have the G^{168}-G^{182} mismatch in SL-4 characteristic of HJV and WEEV (Allison and Stallknecht, 2009). Overall, the 5' 186 nucleotides of the FMV genome, encompassing the two CSEs (i.e., SL-1 and SL-3/SL-4), shared an 88% nucleotide identity to HJV and an 84% nucleotide identity to EEEV and WEEV.

nsP1 (in concert with the RNA triphosphatase activity of nsP2; see below) is known to regulate capping of the genomic and subgenomic RNA, as it contains both guanine-7-methyltransferase and guanylyltransferase activity (Lampio *et al*., 1999). nsP1 residues essential for capping activities, H^{37}, D^{63}, and D^{89} (Ahola *et al*., 1997), were

conserved between FMV, HJV, WEEV, and EEEV. Additionally, nsP1 is known to associate with cytoplasmic membranes during replication, which consequently activates nsP1's transferase activities (Ahola *et al*., 1999). Association of nsP1 with cytoplasmic membranes is mediated by an amphipathic peptide sequence in the central portion of nsP1. The putative nsP1 membrane binding peptide sequence for FMV was ^{245}-GSTIYTEDRSLL*E*SWHLPNV-^{265}, with conserved residues critical for binding (Y^{249}, R^{253}, L^{256}, W^{259}), as determined by lethal mutagenesis in Semliki Forest virus (SFV), underlined (Spuul *et al*., 2007). Interestingly, the single residue of the binding peptide sequence that was unique to FMV (E^{257}; italicized) versus HJV, WEEV, and EEEV (R^{257}), when mutated in SFV (R257E), very closely mimicked wild-type binding levels, suggesting that a E for R substitution in SFV did not adversely effect peptide binding to membranes (Spuul *et al*., 2007). Although palmitoylation (post-translational addition of palmitic acids to cysteine residues) has a pronounced effect on how tightly nsP1 binds to membranes when the protein is expressed alone, its exact role in the viral life cycle is less understood, but it may be involved in morphogenesis or pathogenicity (Laakkonen *et al*., 1996; Aloha *et al*., 2000). The nsP1 palmitoylation site of alphaviruses is conserved and may consist of 1 to 3 consecutive cysteine residues, depending upon the virus (Ahola *et al*., 2000). Similar to *Venezuelan equine encephalitis virus* (VEEV), FMV (like HJV, WEEV, and EEEV) contains the dual cysteine motif, ^{416}-GCCW-^{419}, as the putative palmitoylation site.

nsP2 is the largest nonstructural protein and contains multiple enzymatic functions, including 1) nucleoside triphosphatase (NTPase) (i.e., couples nucleotide hydrolysis to a viral replicative component and/or the inhibition of host cell DNA

replication), 2) RNA helicase (i.e., unwinds RNA secondary structures and/or replicative intermediates), 3) RNA triphosphatase (i.e., removes the 5' γ-phosphate of the mRNA to allow for the addition of the cap nucleotide), and 4) protease (i.e., processes the nonstructural polyprotein) activities (Hardy and Strauss, 1989; Rikkonen *et al*., 1994; Rikkonen, 1996; Vasiljeva *et al*., 2000; Merits *et al*., 2001). nsP2 is also known to temporally regulate positive- and negative-strand synthesis (Sawicki and Sawicki, 1993; De *et al*., 1996; de Groot *et al*., 1990), as well as the transcription of the subgenomic mRNA (Suopanki *et al*., 1998). Additionally, nsP2 is believed to be involved in decreasing interferon production, inhibiting cellular transcription, and inducing cytopathology (Frolova *et al*., 2002; Garmashova *et al*., 2006), and thus has many diverse replicative and non-replicative functions during viral infection.

The two nsP2 nucleoside triphosphate-binding motifs, ^{186}GVPGSG<u>K</u>S^{193} and ^{252}DEAF^{255}, denoted as A and B, respectively, were conserved among FMV, HJV, WEEV, and EEEV, including the K^{192} residue in motif A (underlined) deemed critical for helicase, NTPase, and RNA triphosphatase activity (Vasiljeva *et al*., 2000; Sawicki *et al*, 2006). The nsP2 active site residues, C^{477} and H^{546}, responsible for the cleavage of the nonstructural polyprotein and shown to be invariant among members of *Alphavirus*, was also conserved in FMV, although FMV was unique among all alphaviruses in that it contained an arginine for lysine substitution at the third position following the catalytic cysteine (^{477}CWA<u>R</u>^{480}) (Russo *et al*., 2006). During infection, approximately half of nsP2 that is synthesized in the cytoplasm is translocated to the nucleus, although the function of nuclear nsP2 remains unclear (Peränen *et al*., 1990; Fazakerley *et al*., 2002; Tamm *et al*., 2008). With SFV, the nuclear localization signal (NLS) has been mapped to

residues ^{649}PRRRV^{653}, with R^{651} being essential for nuclear transport of nsP2. The NLS does not appear to be strictly conserved among the alphaviruses, as VEEV contains the NLS sequence ^{644}PGKMV^{648} (Montgomery and Johnston, 2007). The putative NLS sequences for FMV (^{644}QGKRV^{648}), WEEV (^{644}PGKKV^{648}), and HJV/EEEV (^{644}PGKRV^{648}), although varied, all contain a central basic (K) residue, potentially deemed critical for nuclear localization (Rikkonen, 1996; Montgomery and Johnston, 2007). Additionally, a nuclear export signal (NES) in nsP2 of VEEV has also been identified (L^{526} and L^{528}), suggesting that nsP2 cycles in an out of the nucleus during infection (Montgomery and Johnston, 2007). For FMV, HJV, and WEEV, the putative NES was ^{526}VDL^{528}, and thus contained a conserved hydrophobic residue (V^{526}), as well as the highly conserved leucine (L^{528}), previously recognized in other alphaviruses (Montgomery and Johnston, 2007); EEEV differed from the recombinants in that it has an isoleucine for leucine substitution at residue 528 (^{526}VDI^{528}).

All of the nonstructural proteins of FMV were the same length as in HJV, WEEV, and EEEV, except that the nsP3 protein of FMV (522aa) was intermediate in length between HJV (515aa) and WEEV (532aa), with the nsP3 protein of all of the recombinants being shorter than that of EEEV (559aa). nsP3, not surprisingly, was the least conserved protein (i.e., of those obtained from EEEV) between the four viruses, sharing amino acid identities of 61-73%. The N-terminal 160 amino acids of nsP3 in FMV, containing the macro domain which has been demonstrated to bind poly ADP-ribose and RNA, and has been shown to have phosphatase activity (Malet *et al*., 2009), shared a 73-76% amino acid identity to HJV, WEEV, and EEEV. All four viruses contained the 26 strictly conserved amino acids that were previously demonstrated

between the macro domains of selected New and Old World alphaviruses (and a coronavirus), in addition to the D^{10} residue involved in adenine binding and a positively-charged ^{111}-GIY-^{113} phosphate binding site (analogous to ^{112}-GIF-^{114} in VEEV) (Malet *et al*., 2009). Although the C-terminal region of nsP3 is the most variable coding region of the alphavirus genome, a number of short, conserved amino acid motifs were observed upstream of the nsP3/nsP4 cleavage site sequence in FMV, HJV, WEEV, and EEEV (^{316}-IPSP-^{319}, ^{368}-WSIPS-^{372}, ^{414}-QFLS-^{417}, ^{423}-PAPR-^{426}; ^{461}-PPGVAR-^{466}; numbering based on FMV) (Fig. 3.8), which may suggest that these motifs have some structural or functional role in the C-terminal domain. Similar to, HJV, WEEV, and EEEV, the nsP3 gene of FMV contains an opal termination codon (UGA) at genomic positions 5573-5575 followed by a C nucleotide at position 5576 (Strauss *et al*., 1983).

nsP4 is the RNA-dependent RNA polymerase and, in association with different cleaved and uncleaved configurations of the other nonstructural proteins, forms the replicase and transcriptase complexes (Shirako and Strauss, 1994; Strauss and Strauss 1994). The nsP4 of FMV, like the other recombinants and EEEV, contained the destabilizing N-terminal Y^1 and the catalytic core ^{465}-GDD-^{467} motif common to alphavirus RNA polymerases (de Groot *et al*., 1991; Tomar et al., 2006), as well as the double arginine residues ^{332}-RR-^{333} critical for binding to the subgenomic promoter (Li and Stollar, 2004). Although nsP4 was the most conserved nonstructural protein among FMV, HJV, WEEV, and EEEV (87-90% amino acid identity), residues 77-124, contained within the disordered N-terminal domain (Tomar *et al*., 2006), exhibited only a 52-56% amino acid identity. The nsP1/nsP2, nsP2/nsP3, and nsp3/nsP4 cleavage site sequences for FMV were EAGA/GSVE, EAGR/APAY, and RYEAGA/YIFS, respectively, which

were identical to that of EEEV, WEEV, and HJV (Strauss and Strauss, 1994; Allison and Stallknecht, 2009). Based on the cleavage site motifs, the theoretical isoelectric point and molecular weight of each of the FMV nonstructural proteins were: nsP1 (6.00; 59.8kDa), nsP2 (8.80; 89.0kDa), nsP3 (5.54; 57.4kDa), and nsP4 (6.32; 68.1kDa).

Similarity plot analysis of the FMV genome against both EEEV and SINV indicated that FMV exhibited a higher nucleotide identity to EEEV than SINV from the 5' UTR to the capsid, followed by a cross-over within E3 and a greater similarity to SINV in the remaining structural genes (Fig. 3.9), consistent with the single coding region recombination site first reported for WEEV (Hahn *et al*., 1988). Additional analysis of the FMV, HJV, and WEEV genomes against other complete alphavirus genomes using the Recombination Analysis Tool server (Etherington *et al.*, 2004) also demonstrated a single conclusive recombination site occurring within E3, suggesting that no additional recombinations occurred in the 5' two-thirds of the genome for any of the recombinant WEEV antigenic complex viruses (Fig. 3.10).

Phylogenetic analysis of the complete genomes of FMV, HJV, and WEEV with other selected alphaviruses demonstrated that WEE serocomplex viruses clustered independently into two distinct groupings, recombinant viruses (FMV, HJV, and WEEV) and those that are non-recombinant (SINV and AURAV) (Fig. 3.11). That the recombinant viruses formed a clade with EEEV, rather than with the non-recombinant WEE complex viruses, is likely dependent upon the fact that a greater extent of the genome of the ancestral recombinant was derived from EEEV (~8.3kB) than from the SIN-like virus (~3.2kB). Additionally, as the outer structural proteins (i.e., E1/E2, which were derived from the SIN-like virus), may, depending upon the virus, recognize

different host cell receptors and be under selection pressure from the immune response, this likely results in a greater rate of evolution relative to that of the more conserved (EEEV-derived) nonstructural proteins. Overall, the genomic phylogeny, as demonstrated previously with partial E1 sequences for all alphaviruses (Powers *et al*., 2001), was separated into two major lineages: the (1) EEE, VEE, and WEE antigenic complexes, and the (2) Barmah Forest, Middelburg, and Semliki Forest antigenic complexes, with the fish alphaviruses [*Salmon pancreas disease virus* (SPDV) and *Sleeping disease virus* (SDV)] clustering as an outgroup (Fig. 3.11).

Biological relationships of the recombinant WEEV serocomplex viruses

In vitro growth comparisons between FMV, HJV, and WEEV determined that each recombinant, for the most part, had similar growth properties and host range (Table 3.3). As the passage history for each of viruses tested differed (see *Material and Methods* section), how this aspect affects the ability of each virus to replicate in different cell lines is unknown, and therefore direct comparisons between each virus are intended only to indicate the relative susceptibility of each cell line; nevertheless, growth curves can provide a baseline indication of which host species and/or tissue type may be most permissive, or refractory, to alphavirus replication. All three viruses replicated in avian (quail, duck, chicken), mammalian (human, mouse, cat, monkey, bat), reptilian (viper, turtle) and fish (minnow) cells. Overall, fish connective tissue and muscle epithelium (i.e., FHM cells) and cat kidney cortex (i.e., CRFK cells) were the least permissive to replication, while most avian and mammalian cell lines supported high-level replication. However, one noticeable difference was the inability of FMV to infect cells of mosquito

origin (C6/36 cells derived from *Ae. albopictus*), as has previously been documented for FMV both *in vivo* (*Cx. tarsalis* and *Cx. pipiens*) and *in vitro* (*Ae. albopictus*) (Calisher *et al*., 1980; Brault *et al*., 2009). This suggests that one possible consequence of the adaptation of FMV to an aberrant transmission cycle involving *O. vicarious* may have been the loss of the ability to infect mosquitoes, which likely contributes to the focal nature of FMV transmission (i.e., cliff swallow nesting areas) in nature. However, it was also demonstrated that, contrary to the results with *Culex* species, FMV could infect *Cs. melanura in vivo* (Calisher *et al*., 1980), which may be suggestive of an ancestral trait (see *Evolutionary relationships* section). Nevertheless, additional studies, whereby C6/36 cells were inoculated at a higher m.o.i., led to a productive infection as demonstrated by the recovery of virus through multiple (five) passages (not shown), suggesting that the insusceptibility of FMV to mosquitoes or mosquito cells in previous experiments may have been dependent upon the size of the initial inoculum. Genetic characterization of the C6/36-adapted strain of FMV CM4-146 is underway.

During *in vitro* studies, cell lines infected with either HJV or FMV, and to a lesser extent, WEEV, resulted in the formation of small plaques under gum tragacanth. The small plaque phenotype could be abolished by the addition of the cationic polymer DEAE-dextran (100µg/ml) to the overlay, suggesting that positively-charged amino substitutions in the envelope glycoproteins (E1 or E2) interacting with anionic polysaccharides present in the gum tragacanth were responsible for the small plaque phenotype. Previously, it has been shown that small plaque variants of SINV, mediated by basic amino acid substitutions in E2, are cleared from circulation faster than large plaque phenotypes, resulting in an attenuation of virulence in mice (Brynes and Griffin,

2000). The proposed mechanism of this attenuation is the high avidity of small plaque variants for sulphated glycosaminoglycans [e.g., heparan sulfate (HS)], such that they are rapidly sequestered by tissues expressing HS (e.g., liver), resulting in a lower viremia level and limited systemic spread of the virus relative to large plaque types, ultimately resulting in reduced virulence (Byrnes and Griffin, 2000).

As small plaque variants are often suggested to be the result of cell culture adaptation, their relevance (or propensity to occur) during natural infection is largely unknown. HJV isolate 744-01, which was sequenced during this study and isolated from the brain of a barred owl, represents a naturally-occurring small plaque phenotype. Based on the experimental evidence derived from SINV, since basic E2 amino acid substitutions appear to attenuate the virus (in terms of viremia levels), it would appear such mutants would be removed by negative selection as they would not be transmitted efficiently between normal avian hosts and arthropod vectors, which by prerequisite, requires high viral titers in the blood. However, small plaque phenotypes may have some relevance in the pathogenesis (neurotropism) recognized with North American alphaviruses. It has recently been demonstrated that, in contrast to Old World alphaviruses (e.g., SINV), low passage North American alphaviruses (EEEV and WEEV) exhibit normal dependence on HS during infection in mice, and that such basic amino acids in E2 allow for increased infectivity of the viruses for cells of the central nervous system (CNS) (Ryman and Klimstra, 2008). For EEEV, which is a highly neurotropic and neurovirulent virus, these basic E2 amino acids are also present in non-passaged field isolates, suggesting that the HS-binding properties of low passage (small plaque) isolates is not an artifact of cell culture passage (Ryman and Klimstra, 2008). Therefore, basic

amino substitutions in the envelope glycoproteins that allow for efficient HS binding, rather then being attenuating (as observed with SINV), may confer neurotropism and neurovirulence in North American alphaviruses (Ryman and Klimstra, 2008). For pathogenic viruses like EEEV, this neurotropism is associated with neurovirulence and significant levels of mortality; however, for less pathogenic viruses like HJV or FMV, normal HS-binding properties could be associated with efficient infection of the CNS followed by clearance of the virus or, as suggested here, viral persistence.

During FMV, HJV, or WEEV infection of quail neurons (QNR/D), but not quail glial cells (QNR/K2), it was recognized that a subset of neurons survived the initial lytic phase and were persistently infected. For HJV B-230, the prototype HJV strain, persistence was monitored for 100 days; transition from a large plaque (initial inoculum) to the small plaque phenotype began within 24 hrs post-infection, with the small plaque phenotype becoming almost exclusive by day 24 (>99% of the population) (Fig. 3.12). Neurons infected with HJV B-230 maintained a persistent infection that was characterized by low-level virus production (~$\log_{10}$ 3-4 PFU/mL) until the end of the experiment (day 100) (not shown). All HJV isolates (585-01, 744-01, B-230), along with FMV CM4-146 and WEEV SW-99 (Fig. 3.13), maintained a persistent infection in QNR/D cells characterized by a small plaque phenotype and a lack of cytopathology.

The question of whether persistence of alphaviruses in vertebrate hosts is a normal part of their transmission cycle has been a matter of debate for many years (Levine *et al*., 1994). Persistent *in vivo* alphavirus infections of avian hosts are not undocumented, as WEEV has been isolated up to 10 months after experimental infection (Reeves *et al*., 1958). The blood isolation of WEEV 198 days post-inoculation in a

brown-headed cowbird (*Molothrus ater*) suggests that long-term persistence may be followed by reactivation of the virus as demonstrated by viremia. Additionally, the isolation of WEEV from the brain of a house finch (*Carpodacus mexicanus*) 245 days post-inoculation demonstrates that the CNS may serve as a potential site for persistence. In a report by Kandle *et al*. (1968) regarding arbovirus surveillance in New Jersey, both EEEV and HJV were isolated from wild birds during the winter months when mosquito activity was non-existent. As these were the first winter isolations of EEEV and HJV from birds in the state, avian surveillance was initiated the following year during the winter months, whereby six isolates were obtained: all were HJV. Although field evidence suggests that these alphaviruses may potentially be able to overwinter in avian hosts, this mode of annual transmission remains to be a theoretical and circumstantial hypothesis.

Phylogenetic evidence has suggested that EEEV and HJV are likely maintained through the winter in northern temperate regions in discrete locales, as the same genotypes appear to reemerge in the same areas over time (Cilnis *et al*., 1996; Weaver *et al*., 1994). Although overwintering of Buggy Creek virus in both the adult stages and eggs of *O. vicarius* has recently been reported (Brown *et al*., 2009a; Brown *et al*., 2009b), in addition to the demonstration of transovarial transmission of WEEV in nature (Fulhorst *et al*., 1994), a complete understanding of how North American alphaviruses overwinter in northern temperate regions is lacking. The primary scenarios by which alphaviruses could overwinter is by 1) maintenance in bird-mosquito cycles in subtropical locations (e.g., Florida) during the winter months, followed by reintroduction in a site-specific manner by migratory birds in the spring, or 2) persistence in some host,

whether it be in resident bird species, diapausing larval (EEEV or HJV) or adult (WEEV or FMV) insects, dormant poikilothermic hosts, or hibernating mammal species. It is likely that the transmission cycles of these alphaviruses are highly complex, such that it may be a combination of these or unrecognized overwintering mechanisms (which may additionally vary dependent upon geographical location) that contribute to their perpetuation in nature. Although alphaviruses have been documented to persist in post-mitotic neurons of murine origin, this is apparently the first report on the persistence of an alphavirus in a cell line derived from a normal (avian) host of the viruses under study (FMV, HJV, WEEV).

As aforementioned, although alphaviruses initiate lytic infections in most vertebrate cell lines, persistent infections may occur in post-mitotic neurons of mice or rats (Griffin, 1998). The age-dependent differential expression of cellular anti-apoptotic genes (e.g., bcl-2 and bcl-xL) in mature neurons, in comparison to immature neurons or non-terminally differentiated cell types, has been proposed as a potential mechanism for the conversion of a lytic (apoptotic) to a persistent alphavirus infection (Griffin *et al*., 1994; Griffin, 2005). Unlike immature neurons and most nucleated cells, post-mitotic neurons do not express class I major histocompatibility complex (MHC) molecules on their cell surface under resting conditions, nor do they express class I MHC when exposed to cytokine stimulation or viral infection (Joly *et al*., 1991). Therefore, as post-mitotic neurons are (for the most part) incapable of regeneration, and as class I-mediated cytotoxic T lymphocyte (CTL) cell lysis or apoptosis induction in neurons would result in the destruction of a virtually non-renewable cell pool, mature neurons provide a potential target for the persistence of encephalitic viruses.

The clearance of alphaviruses from infected neurons in immunocompetent animals is believed to be antibody-mediated (Griffin *et al*., 1997; Griffin *et al*., 2001). Monoclonal antibodies to E2 have been shown to restrict SINV gene expression in mature neurons in the absence of any exogenous antiviral cofactors (Levine *et al*., 1991; Ubol *et al*., 1995). Levine and Griffin (1992) proposed that the existence of an antibody-mediated non-cytolytic mechanism for viral clearance raises the possibility that, even in acute CNS infections, alphaviruses may persist in neurons for prolonged periods after the termination of a productive infection. If virus-specific antibody is then required to downregulate alphavirus gene expression, reinitiation of a productive infection may therefore be dependent upon the cessation of the antibody response. Levine *et al*. (1994) suggested that whether alphavirus infections in neurons is abortive (virus is cleared) or latent (virus persists and is subsequently reactivated) may be dependent on the magnitude of the antibody response.

Additionally, the viral and cellular mechanisms underlying persistent infections have been studied using alphaviruses or alphavirus replicons (i.e., virus genomes lacking structural genes). While wild-type EEEV replicons can readily establish persistent infections in vertebrate cell lines [as long as they are deficient in interferon (IFN)-α/β signaling], both SINV and VEEV require additional mutations in nonstructural genes in order to become noncytopathic (Frovola *et al*., 2002; Petrakova *et al*., 2005). In the case of both SINV replicons and standard SINV, a mutation occurring at the same exact amino acid position in nsP2 (726) was demonstrated (from independent studies) to initiate persistence, and thus it was concluded that persistence may be due to changes in nsP2 function due to this specific mutation (Dryga *et al*., 1997; Frovola *et al*., 2002).

Petrakova *et al*. (2005) additionally suggested that the putative trigger for initiating persistence (by alphavirus replicons) was low-level replication due to decreased nsP2 function, resulting in a weak inhibition of cellular transcription and translation and subsequent poor apoptosis induction. That is, cell functioning and cell survival was dependent upon the decreased fitness of replicons due to the mutation in nsP2. This may have some bearing on the neuronal persistence of the recombinant alphaviruses, as the observed decreased fitness of the viruses in neurons, relative to glial (QNR/K2) or non-CNS cell lines of avian origin (DF-1 and PDE cells) (Table 3.3), may facilitate the survival of a subset of neurons during the initial lytic phase of infection. However, as cell densities may have differed between the cell lines (i.e., QNR/D cells are at a lower density), ascribing a reduction in the replicative capacity of the recombinant alphaviruses in neurons as a reason for contributing factor for initiating persistence is only tentative. However, concomitant with the establishment of persistence is a progressive change in plaque phenotype, suggesting that persistence may be related to the selection of a mutant genotype(s).

In order to determine if incubation of infected QNR/D cells with HJV B-230 specific antisera would have an effect on virus infection, standard media was removed from days 60-69 post-infection and infected cells were incubated in media supplemented with HJV B-230 polyclonal or monoclonal antibody. With the polyclonal antisera treatment, no extracellular virus was detected on days 63, 66, or 69, and upon removal of the antisera, virus was detected again in the supernatant at levels equivalent to pre-treatment (not shown). Treatment of neurons with (non-neutralizing) monoclonal antibody had no effect on virus production. Although these findings may not be

surprising, it does potentially indicate that if neurons can be persistently infected with HJV (i.e., *in vivo*), virus-specific antibody may be able to neutralize any extracellular virus that is released, thereby inhibiting infection of any additional cells.

During SINV infection in mice, activated B-cells that enter into the CNS and encounter specific antigen may be selectively retained and secrete virus-specific antibody into the extracellular space of the CNS for up to one year (Tyor *et al*., 1992). Concomitantly, the gradual waning of antibody may then provide the trigger for the recrudescence of a productive infection due to the inability to neutralize nascent virus being released from persistently-infected neurons. This is of particular interest in relation to passerine birds (the amplifying hosts for all North American alphaviruses), which, in contrast to non-passerine avian or mammalian humoral immunity, alphavirus-specific antibody responses may be ephemeral. In a study of native Swedish birds experimentally infected with Ockelbo virus (a European subtype of SINV), 11/15 (69%) infected anseriforms (i.e., ducks, geese) that had antibodies at 1 month post-infection, maintained detectable antibodies for the remainder of the study (12 mo.) (Lundstrom and Niklasson, 1996). In contrast, only 2/13 (15%) of the passerine birds that had detectable antibodies at 1 month were seropositive when tested only 2 months later. The short-lived nature of alphavirus-specific antibodies in passerines was exemplified by field research conducted by Main *et al.* (1988) on free-ranging species in Massachusetts naturally-infected with EEEV and HJV. Eighteen of 22 (82%) recaptured black-capped chickadees (*Parus atricapillis*) that were seropositive for EEEV when initially sampled, had lost detectable neutralizing antibodies in 1 to 44 weeks (mean = 10.4 weeks). Likewise, 11 of 16 (69%)

chickadees lost detectable neutralizing antibody to HJV in 1 to 48 weeks (mean = 18.5 weeks) after a seropositive result.

As basic amino acid substitutions in the E2 glycoprotein are believed to be involved the initiation of a small plaque phenotype, and as mutations in nsP2 have been shown to be involved in initiating persistence of alphaviruses in vertebrate cell lines, a genetic analysis of the persistent small plaque HJV B-230 phenotype was undertaken. Comparison of the entire genomes of the initial inoculum and virus recovered at day 100 post-infection indicated five nucleotide changes, three of which occurred within nsP2, while the other two occurred within E2. Of these 5 nucleotide changes, three resulted in nonsynonymous amino acids substitutions: one in nsP2 (K222N) and two in E2 (D40V and E186K). To determine if one or more of the amino acid changes were responsible for the small plaque phenotype, the regions of interest were sequenced at different time points after the establishment of the persistent infection. At day 40 post-infection, the only amino acid substitution of the three previously noted was at residue E2-186 (E→K). The E186K substitution resulted in the displacement of an acidic/polar amino acid (E) with an amino acid with a basic/polar charge (K), consistent with the small plaque morphology induced by the interaction between the negatively-charged polysaccharides present in the overlay and the increased positive charge of E2 associated with the lysine substitution. Whether the E2-186 substitution is additionally involved with the induction of a persistent infection, or that persistence is solely a function of the intrinsic properties of post-mitotic neurons or other factors such as the formation of defective-interfering particles or interferon production (Eaton and Hapel, 1976; Meinkoth and Kennedy, 1980; Weiss *et al*., 1980), remains to be determined.

In order to assess if the genetic changes associated with the small plaque phenotype in HJV were uniform or varied among different isolates, the genome of a naturally-occurring small plaque HJV isolate (744-01) was sequenced. Comparison of the nonstructural and structural protein sequences of HJV 744-01 against two standard (large) plaque isolates, HJV 585-01 and HJV B-230, disclosed a total of 15 amino acid changes between the three isolates. Of these 15 substitutions, 5 were unique to HJV 744-01: nsP2-555 (M→ I), nsP3-149 (A→ T), E2-33 (N→ K), E2-70 (E→ G), and E2-283 (V→I). As valine and isoleucine have very similar side chain charges and hydropathy indices, it is unlikely that the E2-283 substitution has an effect on the induction of the small plaque phenotype. However, as seen with the day100 QNR/D variant (E2-E186K), 744-01 had a lysine substitution, occurring at E2-N33K. Additionally, a glycine for glutamate substitution occurred at position E2-70. As both of these substitutions (E2-E70G and E2-N33K) result in a reduction in the negative charge of the protein, it is likely that both contribute to the small plaque phenotype of HJV 744-01. Thus, although the small plaque morphology of HJV isolates appears to be dependent upon basic amino acid substitutions in E2, different residues appear to confer the phenotype in different isolates of HJV. Whether these changes affect the neurotropism of the virus or the ability to initiate persistence *in vivo* remains to be seen. Additionally, whether the nucleotide changes associated with the persistent small plaque phenotype in WEEV and FMV are also determined by analogous basic amino acid substitutions in E2 remains to be determined.

Evolutionary relationships of the recombinant WEEV serocomplex viruses

The geographical origin of the alphaviruses has been a matter of debate for many years and there are a number of reviews considering theories on both New and Old World origins (Powers *et al*., 2001; Weaver *et al*., 1997; Gould *et al*., 2009). For simplicity, both the origin of the *Alphavirus* genus and the origin of the recombination event leading to the radiance of HJV, WEEV, and FMV are assumed to have occurred in the New World. However, although it is believed that the recombination event occurred in the neotropics of South America, an alternative theory, as proposed here, is that the recombination between EEEV and the SINV-like virus may have occurred in North America, after the establishment of the North American lineage of EEEV.

The fact that both EEEV and HJV use *Cs. melanura* solely as an enzootic vector and, hence, occupy nearly identical geographical ranges, while other North American WEEV serocomplex recombinants occupy disparate geographical regions (i.e., western United States) and utilize alternative vectors (i.e., *Cx. tarsalis*, WEEV; *O. vicarius*, FMV), may be suggestive that the recombination event between EEEV and the SIN-like virus might have occurred in *Cs. melanura*. As *Cs. melanura* is restricted primarily to acid-water swamps from New Brunswick south to eastern Texas (Darsie and Ward, 2005), this would imply that the recombination event likely occurred in the eastern United States. If the recombination event occurred in South America, this would suggest that both EEEV and HJV were independently transported (or dually transported in a single infected avian host) to North America and both subsequently adapted to an enzootic cycle involving *Cs. melanura*. Currently, there an estimated 174 different species of mosquitoes in the United States (Darsie and Ward, 2005); the likelihood that

two alphaviruses, one of which is the parental virus of the other, were independently transported from South America and subsequently adapted to the same mosquito species and transmission cycle is low. However, it should be noted that *Tonate virus*, a strain of VEEV, was recovered from the original pool of the FMV isolate sequenced in this study (Monath *et al*., 1980), suggesting that alphaviruses may be intermittently transported from South America to the United States and that dual infections in a single host may occur.

As the median ranges on the divergence of North from South American lineages of EEEV are estimated to be approximately 1,600-2,300 years ago (Arrigo *et al*., 2010), while the recombination event between EEEV and a SIN-like virus is postulated to have occurred 1,300 to 1,900 years ago (Weaver *et al*., 1997), this could theoretically place the recombination event after the transportation of EEEV to North America. The recombination of EEEV and the SIN-like virus after the adaptation of EEEV to *Cs. melanura* in North America is consistent with the concomitant adaptation of HJV to the same habitat-specialized enzootic vector. The lack of reports regarding the isolation of HJV from South America also supports this hypothesis. Additionally, that fact that HJV, WEEV, and FMV show a greater degree of similarity to North American EEEV than they do to the South American lineages, may also suggest that the ancestral recombinant was derived from an existing North American strain of EEEV. It may also be of interest that the normal transmission cycle of SINV in the Old World involves passerine birds and *Culiseta*/*Culex* species of mosquitoes, and therefore may be more 'ecologically' likely to recombine with a North American lineage of EEEV that is maintained in a cycle involving passerine birds and a strictly ornithophilic mosquito species (*Cs. melanura*); on

the other hand, South American lineages of EEEV are apparently maintained in more diverse cycle(s) involving a number of catholic *Cx. (Melanoconion)* species and, based on the independent evolution of multiple geographically-defined lineages, may be maintained enzootically in ground-dwelling mammalian hosts (Srihongse and Galindo, 1967; Walder *et al*., 1984; Arrigo *et al*., 2010).

If the ancestral recombinant was a HJV-like virus that arose in *Cs. melanura* in the eastern United States, WEEV would have then likely diverged from HJV when it moved westward via bird migration and occupied a new ecological niche in which it could be stably maintained in an enzootic cycle involving passerine birds and a non-*Culiseta* species of mosquito, such as *Cx. tarsalis.* FMV may have then diverged from WEEV (or HJV) by adapting to an *Oe. vicarius*-cliff swallow cycle in the western United States. Although phylogenetic analysis does not suggest that WEEV is the progenitor of FMV (Fig. 3.11), the phylogenetic disparity between FMV and WEEV (or HJV) may be due in part to the adaptation of FMV to a unique transmission cycle involving *O. vicarius*. Adaptation to an aberrant, non-motile arthropod vector that is not closely related to mosquitoes may have resulted in significant genetic changes in order to infect, replicate in, and subsequently be transmitted efficiently by, *O. vicarious*, resulting in the independent evolution and subsequent increased genetic divergence observed in FMV relative to HJV or WEEV.

Although the theory that the recombination event occurred in neotropics is well-established, the close genetic relationship of HJV and North American EEEV, coupled with the identical transmission cycles of the two viruses involving a single highly habitat-specialized mosquito species, *Cs. melanura*, as an enzootic vector, may suggest that the

recombination event could have possibly occurred in North America and that HJV is most similar to, or is, the ancestral descendant of the recombination event from which WEEV and FMV radiated. However, this theory is not intended to supplant current views of alphavirus origins, but rather provide an alternative hypothesis to be considered regarding the origins of the recombinant viruses in the WEEV serocomplex.

Table 3.2. Comparative nucleotide (upper) and amino acid (lower) identities of the genomes of the recombinant western equine encephalitis antigenic complex viruses (FMV, HJV, and WEEV) and descendants of the parental viruses (EEEV and SINV).

		FMV	HJV	WEEV	EEEV	SINV
nsP1	FMV	▪	75	74	73	61
	HJV	84	▪	80	75	54
	WEEV	83	93	▪	75	53
	EEEV	80	86	85	▪	54
	SINV	57	59	59	59	▪
nsP2	FMV	▪	71	70	70	56
	HJV	83	▪	75	72	56
	WEEV	82	89	▪	73	55
	EEEV	81	84	84	▪	58
	SINV	56	55	55	55	▪
nsP3	FMV	▪	66	65	44	35
	HJV	63	▪	69	54	37
	WEEV	64	73	▪	66	33
	EEEV	61	69	65	▪	31
	SINV	36	41	39	37	▪
nsP4	FMV	▪	72	73	73	66
	HJV	87	▪	74	73	61
	WEEV	87	90	▪	74	67
	EEEV	87	88	88	▪	65
	SINV	74	73	73	73	▪
C	FMV	▪	74	70	71	44
	HJV	86	▪	78	75	50
	WEEV	84	90	▪	74	56
	EEEV	84	88	88	▪	51
	SINV	50	52	52	50	▪

		FMV	HJV	WEEV	EEEV	SINV
E3	FMV	▪	56	61	59	41
	HJV	61	▪	75	51	47
	WEEV	56	90	▪	57	48
	EEEV	61	56	53	▪	47
	SINV	51	53	53	39	▪
E2	FMV	▪	62	64	54	62
	HJV	63	▪	72	47	66
	WEEV	66	85	▪	52	64
	EEEV	43	45	45	▪	55
	SINV	60	69	69	44	▪
6K	FMV	▪	73	77	46	66
	HJV	74	▪	78	47	60
	WEEV	83	81	▪	55	72
	EEEV	41	40	43	▪	49
	SINV	65	58	65	38	▪
E1	FMV	▪	69	71	52	67
	HJV	76	▪	77	57	68
	WEEV	77	89	▪	58	70
	EEEV	51	51	50	▪	58
	SINV	73	76	76	52	▪
G[a]	FMV	▪	69	69	65	60
	HJV	78	▪	75	66	61
	WEEV	78	87	▪	68	62
	EEEV	70	73	73	▪	57
	SINV	58	61	60	53	▪

a = complete genomic (G) nucleotide and protein (i.e., nonstructural/structural polyprotein) sequences.

Table 3.3. In vitro host range of recombinant western equine encephalitis antigenic complex viruses.

Cell line	HJV	WEEV	FMV
A-549	5.28 (6)[a]	4.74 (6)	8.10 (5)
C6/36	8.11 (5)	8.06 (5)	0
CRFK	4.68 (3)	3.47 (3)	5.62 (3)
DF-1	7.98 (4)	7.39 (5)	8.51 (4)
FHM	3.84 (2)	3.71 (2)	4.10 (2)
HEK	5.26 (5)	7.41 (6)	6.33 (5)
LLC-MK2	6.12 (5)	5.77 (5)	7.41 (5)
N2a	7.92 (6)	7.95 (6)	9.00 (5)
PDE	8.02 (4)	7.04 (4)	8.72 (4)
QNR/D	5.45 (6)	4.47 (6)	5.58 (6)
QNR/K2	7.50 (3)	7.62 (6)	7.98 (3)
Tb 1 Lu	6.86 (5)	7.32 (4)	7.45 (4)
TH-1	7.45 (3)	7.74 (3)	6.62 (5)
Vero	7.83 (6)	8.02 (3)	8.78 (4)
VH-2	7.10 (3)	6.98 (3)	8.11 (2)

a = Maximum titer reached in each cell line at 28 °C over a 6-day growth period is shown, with day at which the maximum titer was reached in parentheses.

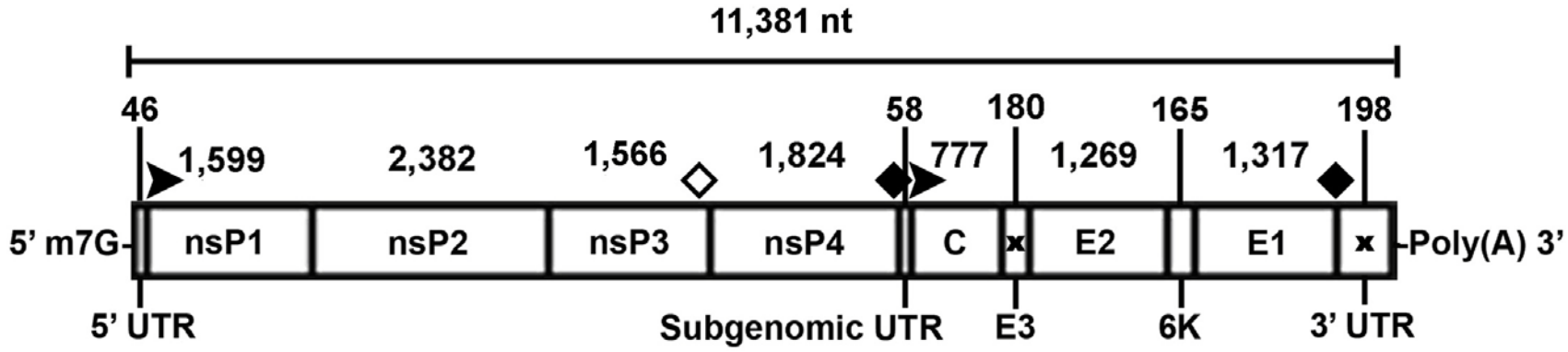


Figure 3.7. Recombinant alphavirus genome schematic. The genome of FMV is shown, which was 11,381 nucleotides (nt) in length. The length of each of the nine genes and three untranslated regions (UTRs) is shown above the schematic. The start and stop codons for the both the nonstructural (nsP1, nsP2, nsP3, nsP4) and structural (C, E3, E2, 6K, E1) polyproteins are indicated with arrowheads and solid diamonds, respectively. The location of the nsP3 readthrough opal termination codon is indicated by an open diamond. The approximate E3 and 3' UTR cross-over sites reported for the ancestral recombinant of WEEV, HJV, and FMV are indicated with an "X".

```
FMV     FRLRAAKKEQFSVCSSFHLPKYRITGVQKIQCSKPVIFSGIAPPAVHPRKYTAIISERT- 332
HJV     FRLRAAKKEQFAVCSSFPLPKYRITGVQKLQCSKPVLFSGIVPPAVHPRRYATTTIEVA- 332
WEEV    YRLRSAKKEQFAVCSSFLLPKYRITGVQKLQCSKPVLFSGVVPPAVHPRKYAEIILETP- 331
EEEV    YRLRMAKNEQFAVCSSFQLPKYRITGVQKIQCSKPVIFSGTVPPAIHPRKFASVTVEDTP 333
        :*** **:***:***** **********:******:*** .***:***:::    * .

FMV     --TPEPTEIFSIEEP---PNVIPSP-TQFLDYGAESLCFDNAVTTTGDSALSLCSSDGAS 387
HJV     ---TSPIEQRPLREAPPVPARIPSP-ISNRTISVESLLSVG-------TQSASISWDLHE 381
WEEV    ---PPPATTTVICEP-TVPERIPSP-VISRAPSAESLLSLGGVSFSSSATRSSTAWSDYD 386
EEEV    VVQPERLVPRRPAPPVPVPARIPSPPCTSTNGSTTSIQSLGEDQSASASSGAEISVDQVS 393
           .          .   *  ****       .. *:   .       :  :  : .  .

FMV     ET----------------------------PEDLSVRRTVSTWSIPSATGFEIKEEA--- 415
HJV     Q----------------------------VITADVHRDAESSAWSIPSASGFEV-LPS--- 410
WEEV    RR-------------------------FVVTADVH-QANTSTWSIPSAPGLDVQLPSDVT 420
EEEV    LWSIPSATGFDVRTSSSLSLEQSTFPTMVVEAEIH-ASQGSLWSIPSITGSETRVPS--- 452
                                       ::      * ***** .* :    :

FMV     -------------TEDEHVYIAEYDQRDYSNVTEILLEFSRAPVQFLSDFKPIPAPRSIR 462
HJV     ----PTPSHS---ISNISWEEVSADCTRERCVADIMQDFRSAPFQFLSDYKPIPAPRSR- 462
WEEV    DSHWSIPSASGFEVRTPSVQDLTAECAKPRGLAEIMQDFNTAPFQFLSDYRPVPAPRRR- 479
EEEV    ----PPSQGSRHSTPSVSASHTSVDLITFDSVAEILEDFSRSPFQFLSEIKPIPAPRTR- 504
                              :      :::*: :*  :*.****: :*:****

FMV     PKICPVPAPRTKVTGPSFGASTLQYSKVYERPPGVARAISEAELDAYIQQQLNXRYEAGA 522
HJV     ----PTPAPRSTVSAPPIPKPR---KIVYRQPPGVARSISEAELDEYIRQHTNXRYEAGA 515
WEEV    ----PIPSPRSTASAPPVPKPR---RTKYQQPPGVARAISEAELDEYIRQHSNXRYEAGA 532
EEEV    ----VNNMSRSADTIKPIPKPRKC-QVKYTQPPGVARAISAAEFDEFVRRHSNXRYEAGA 559
               .*:  :  ..  .       * :******:** **:* ::::: ********
```

Figure 3.8. Partial amino acid alignment of nsP3 between FMV, HJV, WEEV, and EEEV, showing the lack of conservation in the C-terminal region. The degree of truncation of the three recombinant WEE complex viruses relative to the parental EEEV is shown to the right of the alignment. The readthrough opal termination codon adjacent the RYEAGA/YIFS nsP3/nsP4 cleavage site sequence is highlighted in red and denoted as an “X”.

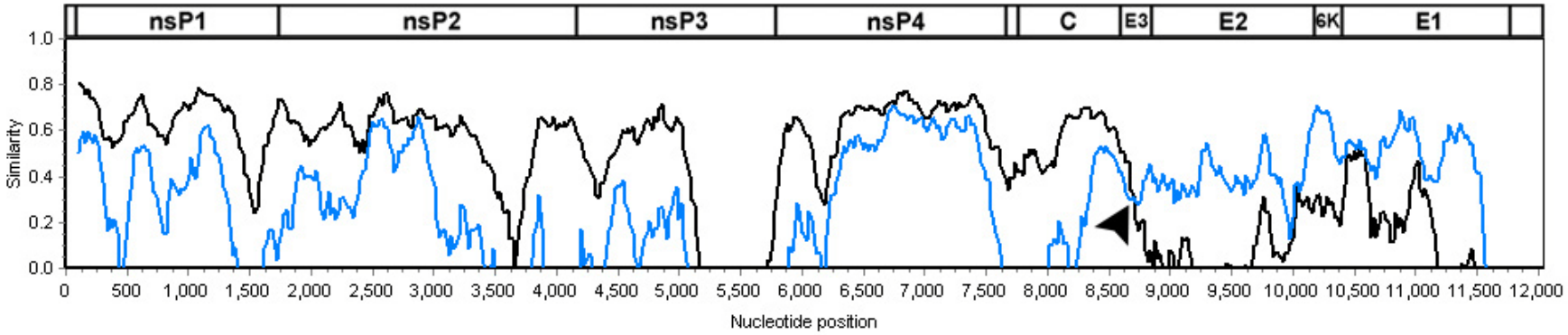


Figure 3.9. Genomic relatedness of FMV to descendants of the parental viruses of the recombination event. A schematic of the FMV genome is shown above the graph. Nucleotide similarity of EEEV (black line) and SINV (blue line) to the FMV genome is shown. The approximate location of the coding region recombination cross-over site between EEEV and the SIN-like virus is indicated with an arrowhead. Genomes used in the Simplot analysis were FMV (GQ281603), EEEV (EF151502), and Ockelbo virus, a subtype of SINV (M69205).

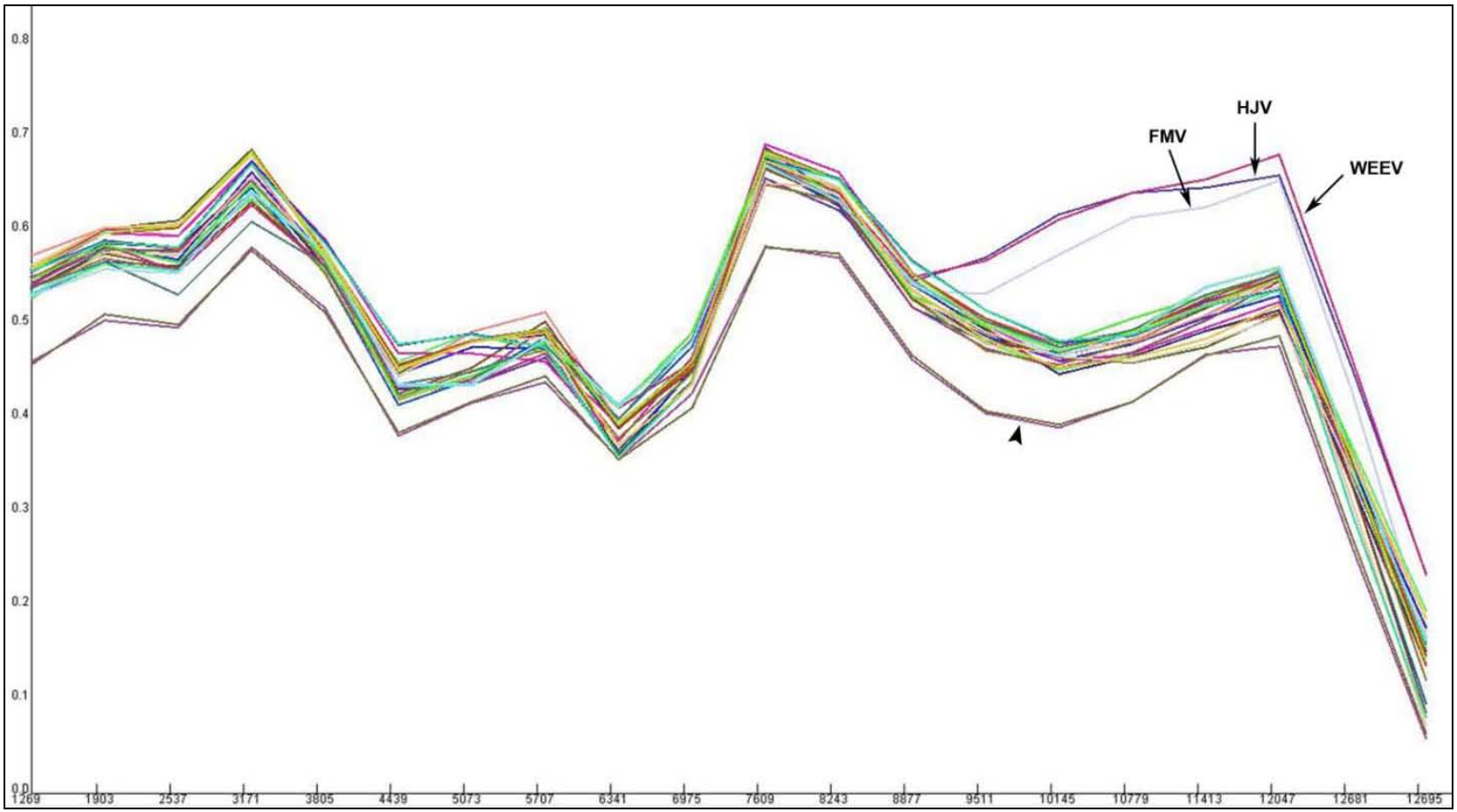
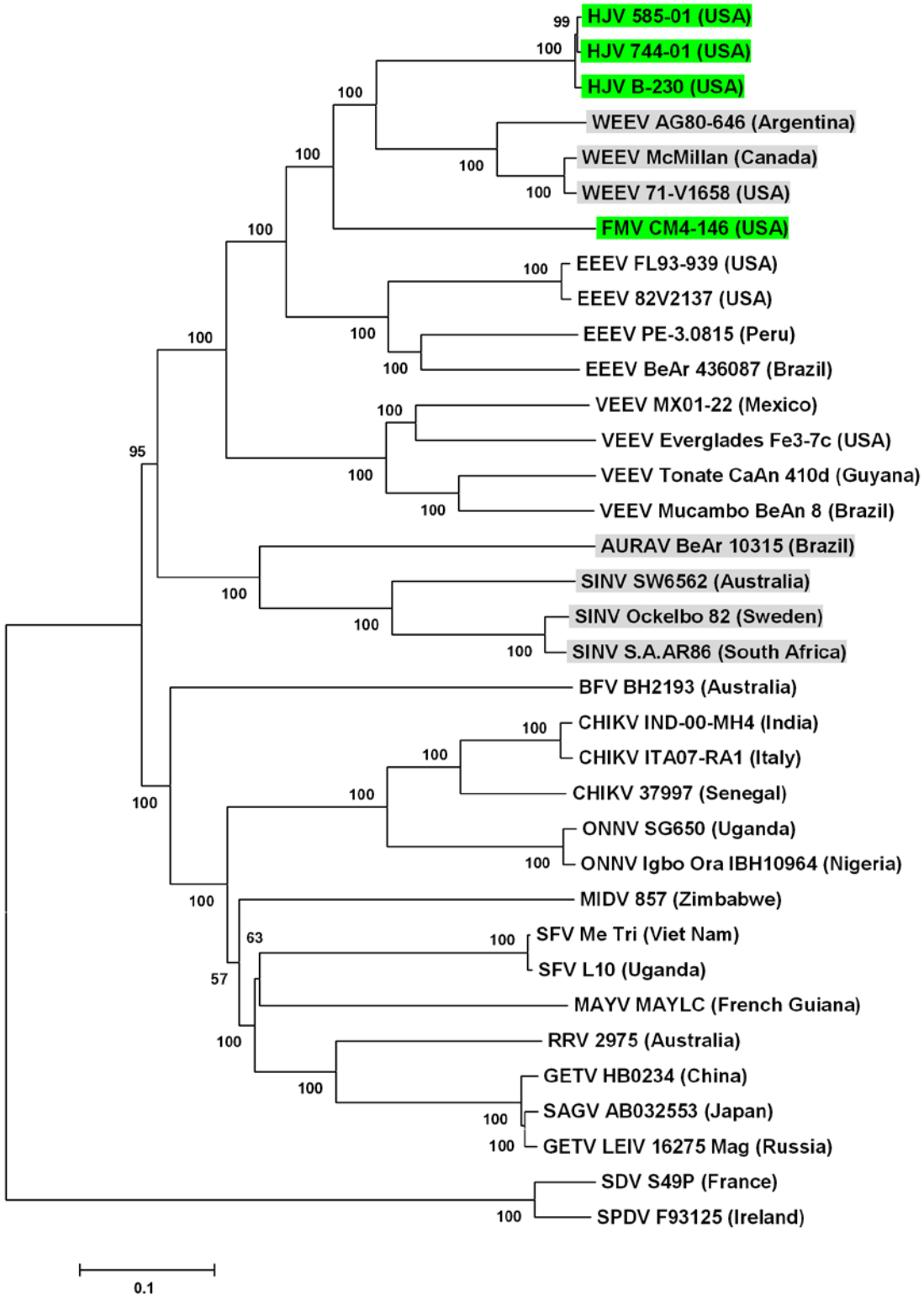


Figure 3.10. Recombination analysis of alphavirus genomes (see *Figure 3.11* for list of viruses used in the analysis). Lines represent the genetic similarity (X-axis) of each virus to SINV (Ockelbo virus). The nucleotide position relative to the SINV genome is shown on the Y-axis. All SIN-like or related viruses (i.e., AURAV) were removed from the analysis for clarity. The two lower lines indicating the greatest degree of genetic dissimilarity (arrowhead) represent the two fish alphaviruses, SPDV and SDV. The breakpoint for FMV, HJV, and WEEV occurs approximately over nucleotide position 8877.

Figure 3.11. Evolutionary relationships of the WEE serocomplex recombinant viruses with selected alphaviruses based on complete genomic sequences. Viruses are represented by their ICTV abbreviation, followed by strain/isolate designation (when available) and country of origin. Groupings of the recombinant (upper) and non-recombinant (lower) WEE serocomplex viruses are highlighted in gray, with the recombinant viruses whose genomes were sequenced as part of this research shown in green. The phylogeny was generated by neighbor-joining analysis. Bootstrap values were determined using 2000 replicates and are listed at each node. Branch lengths are drawn to scale. The tree was calculated using the Kimura 2-parameter method and evolutionary distances are represented as the number of nucleotide substitutions per site. Sequences used in the analysis were: *Aura virus* (AURAV) BeAr 10315 (AF126284); *Barmah Forest virus* (BFV) BH2193 (BFU73745); *Chikungunya virus* (CHIKV) ITA07-RA1 (EU244823), CHIKV IND-00-MH4 (EF027139), and CHIKV 37997 (AY726732); *Eastern equine encephalitis virus* (EEEV) FL93-939 (EF151502), EEEV 82V2137 (U01034), EEEV PE-3.0815 (DQ241303), and EEEV BeAr 436087 (EF151503); *Fort Morgan virus* (FMV) CM4-146 (NC_013528); *Getah virus* (GETV) LEIV 16275 Mag (EF631998) and GETV HB0234 (EU015062); *Highlands J virus* (HJV) 585-01 (NC_012561), HJV 744-01 (GU167952), and HJV B-230 (GQ227789); *O'nyong-nyong virus* (ONNV) SG650 (AF079456) and ONNV Igbo Ora strain IBH10964 (AF079457); *Mayaro virus* (MAYV) MAYLC (DQ001069); *Middelburg virus* (MIDV) 857 (EF536323); *Ross River virus* (RRV) 2975 (GQ433360); *Sagiyama virus* (SAGV, not an ICTV abbreviation; currently classified as a subtype of RRV) (AB032553); *Salmon pancreas disease virus* (SPDV) F93125 (AJ316244); *Semliki forest virus* (SFV) L10 (AY112987) and SFV Me Tri (EU350586); *Sindbis virus* (SINV) SW6562 (AF429428), SINV Ockelbo strain Edsbyn 82-5 (M69205), and SIN-like virus S.A.AR86 (ACU38305); *Sleeping disease virus* (SDV) S49P (AJ316246); *Venezuelan equine encephalitis virus* (VEEV) MX01-22 (AY823299), VEEV Tonate CaAn 410d Guyana (AF075254), VEEV Everglades Fe3-7c (AF075251), and VEEV Mucambo BeAn 8 (AF075253); *Western equine encephalitis virus* (WEEV) McMillan (GQ287640), WEEV 71-V1658 (GQ287645), and WEEV AG80-646 (GQ287646).


HJV 585-01 (USA)
HJV 744-01 (USA)
HJV B-230 (USA)
WEEV AG80-646 (Argentina)
WEEV McMillan (Canada)
WEEV 71-V1658 (USA)
FMV CM4-146 (USA)
EEEV FL93-939 (USA)
EEEV 82V2137 (USA)
EEEV PE-3.0815 (Peru)
EEEV BeAr 436087 (Brazil)
VEEV MX01-22 (Mexico)
VEEV Everglades Fe3-7c (USA)
VEEV Tonate CaAn 410d (Guyana)
VEEV Mucambo BeAn 8 (Brazil)
AURAV BeAr 10315 (Brazil)
SINV SW6562 (Australia)
SINV Ockelbo 82 (Sweden)
SINV S.A.AR86 (South Africa)
BFV BH2193 (Australia)
CHIKV IND-00-MH4 (India)
CHIKV ITA07-RA1 (Italy)
CHIKV 37997 (Senegal)
ONNV SG650 (Uganda)
ONNV Igbo Ora IBH10964 (Nigeria)
MIDV 857 (Zimbabwe)
SFV Me Tri (Viet Nam)
SFV L10 (Uganda)
MAYV MAYLC (French Guiana)
RRV 2975 (Australia)
GETV HB0234 (China)
SAGV AB032553 (Japan)
GETV LEIV 16275 Mag (Russia)
SDV S49P (France)
SPDV F93125 (Ireland)
0.1

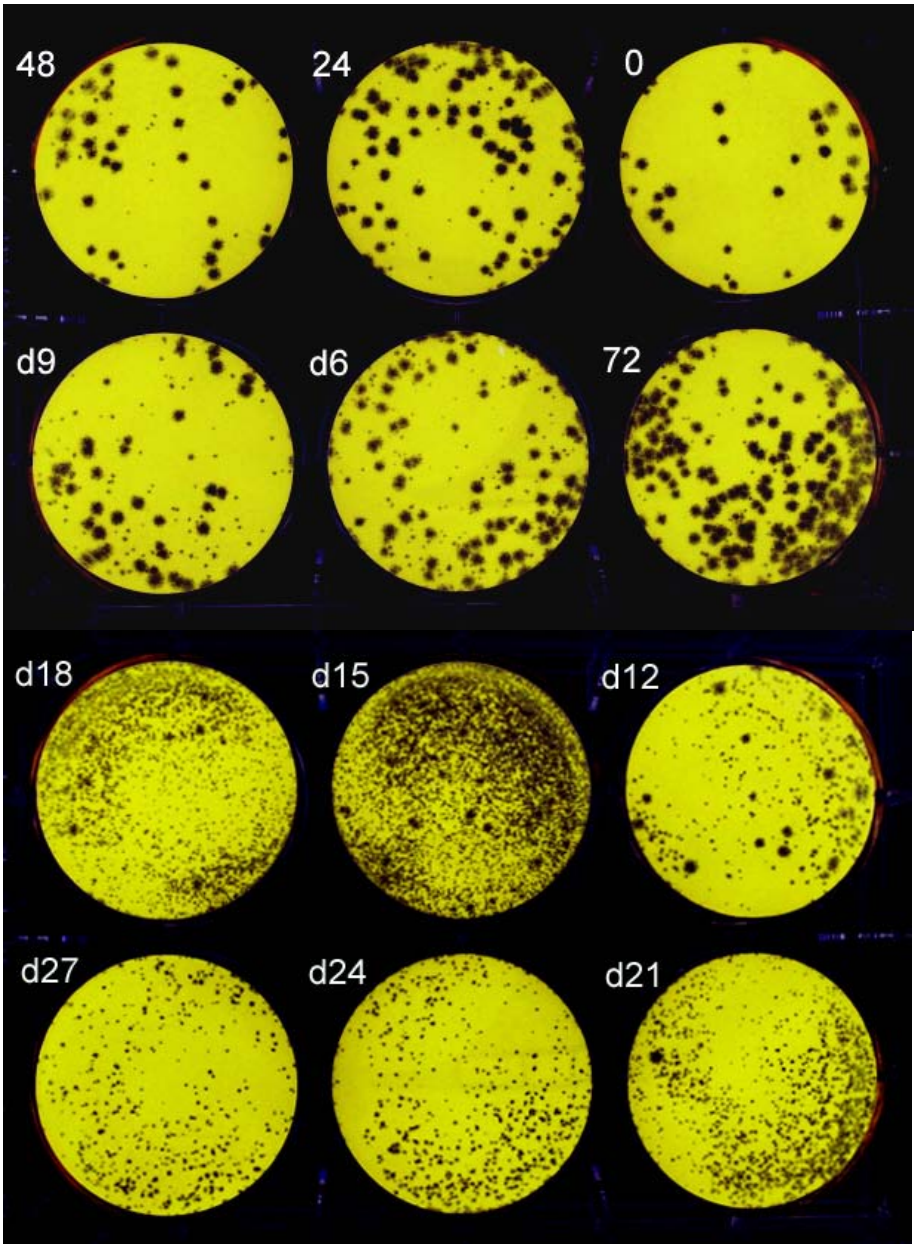


Figure 3.12. Selection for the small plaque HJV B-230 phenotype over the course of infection in quail neurons. The initial inoculum (0 hours) represents HJV B-230 prior to passage in QNR/D cells. Plaque morphologies at 24 hour intervals up to 72 hours post-infection, followed by 3-day intervals up to day 27, are shown.

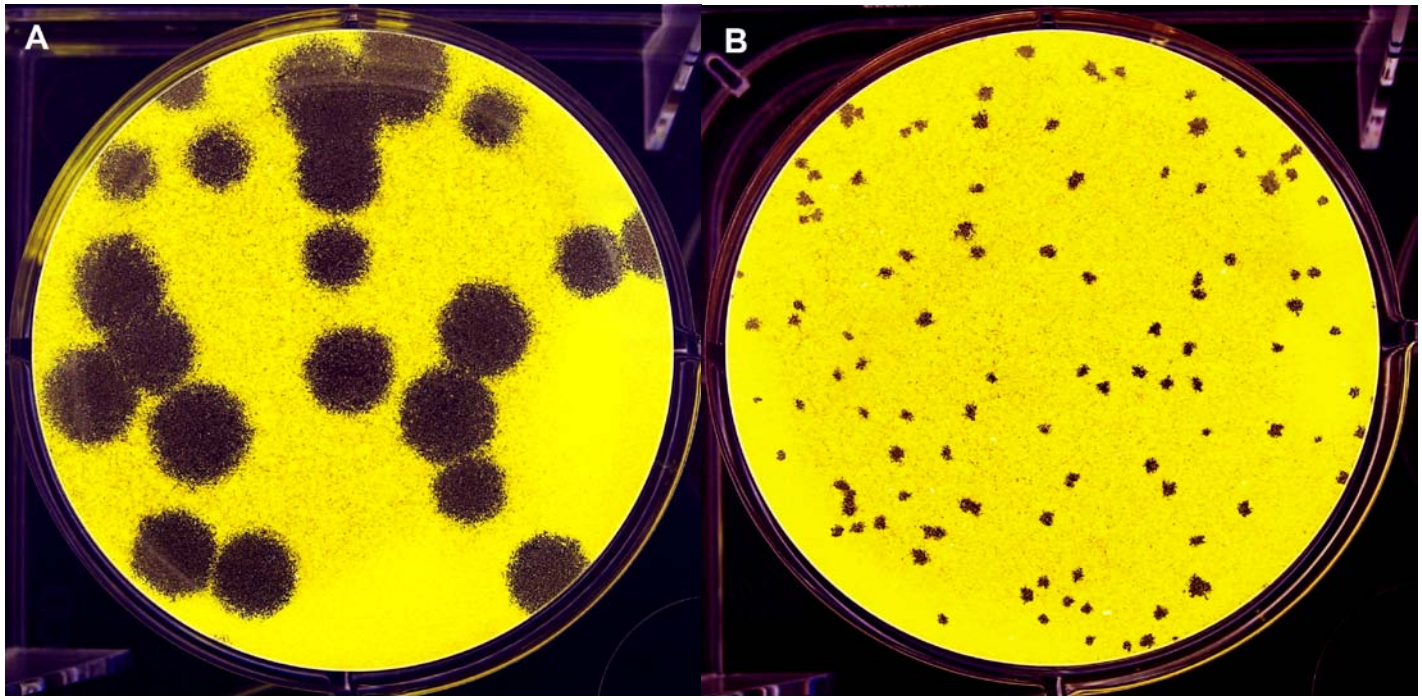


Figure 3.13. Plaque morphologies of WEEV SW-99 prior to and after passage in quail neurons. Plaque titrations of WEEV SW-99 (A) prior to passage in QNR/D cells (stock virus) and (B) at day 24 post-infection in QNR/D cells. Plates were fixed and stained ~60 hours post-adsorption.

Acknowledgements

Funds were provided through sponsorship from the fish and wildlife agencies of Alabama, Arkansas, Florida, Georgia, Kansas, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Puerto Rico, South Carolina, Tennessee, Virginia, and West Virginia; through the Federal Aid to Wildlife Restoration Act (50 Stat. 917) and Grant Agreement 06ERAG0005, Biological Resources Division, U.S. Geological Survey, U.S. Department of the Interior; and through Cooperative Agreements 0596130032CA and 0696130032CA, Veterinary Services, Animal and Plant Health Inspection Service, U.S. Department of Agriculture.

References

Ahola, T., Laakkonen, P., Vihinen, H., Kääriäinen, L., 1997. Critical residues of Semliki Forest virus RNA capping enzyme involved in methyltransferase and guanylyltransferase-like activities. J. Virol. 71, 392-397.

Ahola, T., Lampio, A., Auvinen, P., Kääriäinen, L. 1999. Semliki Forest virus mRNA capping enzyme requires association with anionic membrane phospholipids for activity. EMBO J. 18, 3164-3172.

Ahola, T., Kujala, P., Tuittila, M., Blom, T., Laakkonen, P., Hinkkanen, A., Auvinen, P., 2000. Effects of palmitoylation of replicase protein nsP1 on alphavirus infection. J. Virol. 74, 6725-6733.

Allison, A. B., Stallknecht, D. E., 2009. Genomic sequencing of Highlands J virus: a comparison to western and eastern equine encephalitis viruses. Virus Res. 145, 334-340.

Arrigo, N. C., Adams, A. P, Weaver, S. C., 2010. Evolutionary patterns of eastern equine encephalitis virus in North versus South America suggest ecological differences and taxonomic revision. J. Virol. 84, 1014-1025.

Brault, A. C., Powers, A. M., Chavez, C. L., Lopez, R. N., Cachón, M. F., Gutierrez, L. F., Kang, W., Tesh, R. B., Shope, R. E., Weaver, S. C., 1999. Genetic and antigenic diversity among eastern equine encephalitis viruses from North, Central, and South America. Am. J. Trop. Med. Hyg. 61, 579-586.

Brault, A. C., Armijos, M. V., Wheeler, S., Wright, S., Fang, Y., Langevin, S., Reisen, W. K., 2009. Stone Lakes virus (family Togaviridae, genus Alphavirus), a variant of Fort Morgan virus isolated from swallow bugs (Hemiptera: Cimicidae) west of the Continental Divide. J. Med. Entomol. 46, 1203-1209.

Brown, C. R., Moore, A. T., Knutie, S. A., Komar, N., 2009a. Overwintering of infectious Buggy Creek virus (Togaviridae: Alphavirus) in Oeciacus vicarius (Hemiptera: Cimicidae) in North Dakota. J. Med. Entomol. 46, 391-394.

Brown, C. R., Moore, A. T., Young, G. R., Padhi, A., Komar, N., 2009b. Isolation of Buggy Creek virus (Togaviridae: Alphavirus) from field-collected eggs of Oeciacus vicarius (Hemiptera: Cimicidae). J. Med. Entomol. 46, 375-379.

Byrnes, A. P., Griffin, D. E., 2000. Large-plaque mutants of Sindbis virus show reduced binding to heparan sulfate, heightened viremia, and slower clearance from the circulation. J. Virol. 74, 644-651.

Calisher, C. H., Monath, T. P., Muth, D. J., Lazuick, J. S., Trent, D. W., Francy, D. B., Kemp, G. E., Chandler, F. W. 1980. Characterization of Fort Morgan virus, an alphavirus

of the western equine encephalitis virus complex in an unusual ecosystem. Am. J. Trop. Med. Hyg. 29, 1428-1440.

Calisher, C. H, Karabatsos, N., Lazuick, J. S., Monath, T. P., Wolff, K. L., 1988. Reevaluation of the western equine encephalitis antigenic complex of alphaviruses (family Togaviridae) as determined by neutralization tests. Am. J. Trop. Med. Hyg. 38, 447-452.

Causey O. R., Casals, J., Shope, R. E., Udomsakdi, S. 1963. Aura and Una, two new Group A arthropod-borne viruses. Am. J. Trop. Med. Hyg. 12, 777-781.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., Thompson, J. D., 2003. Multiple sequence alignment with the CLUSTAL series of programs. Nucleic Acids Res. 31, 3497-3500.

Cilnis, M. J., Kang, W., Weaver, S. C., 1996. Genetic conservation of Highlands J viruses. Virology 218, 343-351.

Darsie, R. F., Jr., Ward, R. A., 2005. Identification and geographical distribution of the mosquitoes of North America, north of Mexico. University Press of Florida, Gainesville, Florida.

Dé, I., Sawicki, S. G., Sawicki, D. L., 1996. Sindbis virus RNA-negative mutants that fail to convert from minus-strand to plus-strand synthesis: role of the nsP2 protein. J. Virol. 70, 2706-2719.

de Groot, R. J., Hardy, W. R., Shirako, Y., Strauss, J. H. 1990. Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo. EMBO J. 9, 2631-2638.

de Groot R. J., Rümenapf, T., Kuhn, R.J., Strauss, E. G., Strauss, J. H. 1991. Sindbis virus RNA polymerase is degraded by the N-end rule pathway. Proc. Natl. Acad. Sci. U. S. A. 88, 8967-8971.

Dryga, S. A., Dryga, O. A., Schlesinger, S., 1997. Identification of mutations in a Sindbis virus variant able to establish persistent infection in BHK cells: the importance of a mutation in the nsP2 gene. Virology 228, 74-83.

Eaton, B. T., Hapel, A. J., 1976. Persistent noncytolytic togavirus infection of primary mouse muscle cells. Virology 72, 266-271.

Englund, J. A., Breningstall, G. N., Heck, L. J., Lazuick, J. S., Karabatsos, N., Calisher, C. H., Tsai, T. F., 1986. Diagnosis of western equine encephalitis in an infant by brain biopsy. Pediatr. Infect. Dis. 5, 382-384.

Etherington, G. J., Dicks, J., Roberts, I. N., 2004. Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. Bioinformatics 21, 278-281.

Fazakerley, J. K., 2002. Pathogenesis of Semliki Forest virus encephalitis. J. Neurovirol. Suppl. 2, 66-74.

Ficken, M. D., Wages, D. P., Guy, J. S., Quinn, J. A., Emory, W. H., 1993. High mortality of domestic turkeys associated with Highlands J virus and eastern equine encephalitis virus infections. Avian Dis. 37, 585-590

Frolova, E. I., Fayzulin, R. Z., Cook, S. H., Griffin, D. E., Rice, C. M., Frolov, I., 2002. Roles of nonstructural protein nsP2 and alpha/beta interferons in determining the outcome of Sindbis virus infection. J. Virol. 76, 11254-11264.

Fulhorst, C. F., Hardy, J. L., Eldridge, B. F., Presser, S. B., Reeves, W. C., 1994. Natural vertical transmission of western equine encephalomyelitis virus in mosquitoes. Science 263, 676-678.

Garmashova, N., Gorchakov, R., Frolova, E., Frolov, I. 2006. Sindbis virus nonstructural protein nsP2 is cytotoxic and inhibits cellular transcription. J. Virol. 80, 5686-5696.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., Bairoch, A., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784-3788.

Gould, E. A., Coutard, B., Malet, H., Morin, B., Jamal, S., Weaver, S., Gorbalenya, A., Moureau, G., Baronti, C., Delogu, I., Forrester, N., Khasnatinov, M., Gritsun, T., de Lamballerie, X., Canard, B., 2009. Understanding the alphaviruses: Recent research on important emerging pathogens and progress towards their control. Antiviral Res. (in press).

Griffin, D. E., Levine, B., Ubol, S., Hardwick, J. M., 1994. The effects of alphavirus infection on neurons. Ann. Neurol. 35, S23-27.

Griffin, D. E., Levine, B., Tyor, W., Ubol, S., Despres, P., 1997. The role of antibody in recovery from alphavirus encephalitis. Immunol. Rev. 159, 155-161.

Griffin, D. E., 1998. A review of alphavirus replication in neurons. Neurosci. Biobehav. Rev. 22, 721-733.

Griffin, D. E., Ubol, S., Despres, P., Kimura, T., Byrnes, A., 2001. Role of antibodies in controlling alphavirus infection of neurons. Curr. Top. Microbiol. Immunol. 260, 191-200.

Griffin, D. E., 2005. Neuronal cell death in alphavirus encephalomyelitis. Curr. Top. Microbiol. Immunol. 289, 57-77.

Griffin, D. E., 2007. Alphaviruses, p. 1023-1067 *In* Knipe, D. M., Howley, P. M., Griffin, D. E., Martin, M. A., Lamb, R. A., Roizman, B., Straus, S. E. (ed.), Fields Virology, Vol. 1, Fifth ed. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.

Hahn, C. S., Lustig, S., Strauss, E. G., Strauss, J. H., 1988. Western equine encephalitis virus is a recombinant virus. Proc. Natl. Acad. Sci. U. S. A. 85, 5997-6001.

Hardy, W. R., Strauss, J. H., 1989. Processing the nonstructural polyproteins of Sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans. J. Virol. 63, 4653-4664.

Hayes, C. G., Wallis, R. C., 1977. Ecology of western equine encephalomyelitis in the eastern United States. Adv. Virus Res. 21, 37-83.

Hayes, R. O., Francy, D. B., Lazuick, J. S., Smith, G. C., Gibbs, E. P. J., 1977. Role of the cliff swallow bug (Oeciacus vicarius) in the natural cycle of a western equine encephalitis-related alphavirus. J. Med. Entomol. 14, 257-262.

Hofacker, I. L., 2003. Vienna RNA secondary structure server. Nucleic Acids Res. 31, 3429-3431.

Hopla, C. E., Francy, D. B., Calisher, C. H., Lazuick, J. S., 1993. Relationship of cliff swallows, ectoparasites, and an Alphavirus in west-central Oklahoma. J. Med. Entomol. 30, 267-272.

Huyvaert, K. P., Moore, A. T., Panella, N. A., Edwards, E. A., Brown, M. B., Komar, N., Brown, C. R., 2008. Experimental inoculation of house sparrows (Passer domesticus) with buggy creek virus. J. Wildl. Dis. 44, 331-340.

Joly, E., Mucke, L., Oldstone, M. B., 1991. Viral persistence in neurons explained by lack of major histocompatibility class I expression. Science 253, 1283-1285.

Kandle, R. P., Goldfield, M., Altman, R., Sussman, O., Gusciora, W. R., 1968. Arbovirus studies in New Jersey: A progress report. New Jersey Mosq. Extermin. Assoc. 55, 14-19.

Laakkonen, P., Ahola, T., Kääriäinen L., 1996. The effects of palmitoylation on membrane association of Semliki forest virus RNA capping enzyme. J. Biol. Chem. 271, 28567-28571.

Lampio, A., Ahola, T., Darzynkiewicz, E., Stepinski, J., Jankowska-Anyszka, M., Kääriäinen, L., 1999. Guanosine nucleotide analogs as inhibitors of alphavirus mRNA capping enzyme. Antiviral Res. 42, 35-46.

Levine, B., Hardwick, J. M., Trapp, B. D., Crawford, T. O., Bollinger, R. C., Griffin, D. E., 1991. Antibody-mediated clearance of alphavirus infection from neurons. Science 254, 856-860.

Levine, B., Hardwick, J. M., Griffin, D. E., 1994. Persistence of alphaviruses in vertebrate hosts. Trends Microbiol. 2, 25-28.

Levine, B., Griffin, D. E., 1992. Persistence of viral RNA in mouse brains after recovery from acute alphavirus encephalitis. J. Virol. 66, 6429-6435.

Li, M. L., Stollar, V., 2004. Identification of the amino acid sequence in Sindbis virus nsP4 that binds to the promoter for the synthesis of the subgenomic RNA. Proc. Natl. Acad. Sci. U. S. A. 101, 9429-9434.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W., Ray, S. C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152-160.

Loye, J. E., 1985. The life history and ecology of the cliff swallow bug, Oeciacus vicarious (Hemiptera: Cimicidae). Ent. Med. et Parasitol. 23, 133-139.

Lundstrom, J. O., Niklasson, B., 1996. Ockelbo virus (Togaviridae: Alphavirus) neutralizing antibodies in experimentally infected Swedish birds. J. Wild. Dis. 32, 87-93.

Main, A. J., Anderson, K. S., Maxfield, H. K., Rosenau, B., Oliver, C., 1988. Duration of alphavirus neutralizing antibody in naturally infected birds. Am. J. Trop. Med. 38, 208-217.

Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E. A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A. E., de Lamballerie, X., Canard, B., 2009. The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. J. Virol. 83, 6534-6545.

Meinkoth, J., Kennedy, S. I., 1980. Semliki forest virus persistence in mouse L929 cells. Virology 100, 141-155.

Merits, A., Vasiljeva, L., Ahola, T., Kääriäinen, L., Auvinen, P., 2001. Proteolytic processing of Semliki Forest virus-specific non-structural polyprotein by nsP2 protease. J. Gen. Virol. 82, 765-773.

Mitchell, C. J., Monath, T. P., Sabattini, M. S., Daffner, J. F., Cropp, C. B., Calisher, C. H., Darsie, R. F. Jr., Jakob, W. L., 1987. Arbovirus isolations from mosquitoes collected during and after the 1982-1983 epizootic of western equine encephalitis in Argentina. Am. J. Trop. Med. Hyg. 36, 107-113.

Monath, T. P., Lazuick, J. S., Cropp, C. B., Rush, W.A., Calisher, C. H., Kinney, R. M., Trent, D. W., Kemp, G. E., Bowen, G. S., Francy, D. B., 1980. Recovery of Tonate virus ("Bijou Bridge" strain), a member of the Venezuelan equine encephalomyelitis virus complex, from Cliff Swallow nest bugs (Oeciacus vicarius) and nestling birds in North America. Am. J. Trop. Med. Hyg. 29, 969-983.

Montgomery, S. A., Johnston, R. E., 2007. Nuclear import and export of Venezuelan equine encephalitis virus nonstructural protein 2. J. Virol. 81, 10268-10279.

Netolitzky, D. J., Schmaltz, F. L., Parker, M. D., Rayner, G. A., Fisher, G. R., Trent, D. W., Bader, D. E., Nagata, L. P., 2000. Complete genomic RNA sequence of western equine encephalitis virus and expression of the structural genes. J. Gen. Virol. 81, 151-159.

Padhi, A., Moore, A. T., Brown, M. B., Foster, J. E., Pfeffer, M., Gaines, K. P., O'Brien, V. A., Strickler, S. A., Johnson, A. E., Brown, C. R., 2008. Phylogeographical structure and evolutionary history of two Buggy Creek virus lineages in the western Great Plains of North America. J. Gen. Virol. 89, 2122-2131.

Peränen, J., Rikkonen, M., Liljeström, P., Kääriäinen, L., 1990. Nuclear localization of Semliki Forest virus-specific nonstructural protein nsP2. J. Virol. 64, 1888-1896.

Pessac, B., Girard, A., Romey, G., Crisanti, P., Lorinet, A. M., Calothy, G., 1983. A neuronal clone derived from a Rous sarcoma virus-transformed quail embryo neuroretina established culture. Nature 302, 616-618.

Petrakova, O., Volkova, E., Gorchakov, R., Paessler, S., Kinney, R. M., Frolov, I., 2005. Noncytopathic replication of Venezuelan equine encephalitis virus and eastern equine encephalitis virus replicons in mammalian cells. J. Virol. 79, 7597-7608.

Pfeffer, M., Proebster, B., Kinney, R. M., Kaaden, O. R., 1997. Genus-specific detection of alphaviruses by a semi-nested reverse transcription-polymerase chain reaction. Am. J. Trop. Med. Hyg. 57, 709-718.

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

Pfeffer, M., Foster, J. E., Edwards, E. A., Brown, M. B., Komar, N., Brown, C. R., 2006. Phylogenetic analysis of Buggy Creek virus: evidence for multiple clades in the Western Great Plains, United States of America. Appl. Environ. Microbiol. 72, 6886-6893.

Przelomski, M. M., O'Rourke, E., Grady, G. F., Berardi, V. P., Markley, H. G., 1988. Eastern equine encephalitis in Massachusetts: a report of 16 cases, 1970-1984. Neurology 38, 736-739.

Powers, A. M., Brault, A. C., Shirako, Y., Strauss, E. G., Kang, W., Strauss, J. H., Weaver, S. C., 2001. Evolutionary relationships and systematics of the alphaviruses. J. Virol. 75, 10118-101131.

Randolph, K. D., Vanhooser, S. L, Hoffman, M., 1994. Western equine encephalitis virus in emus in Oklahoma. J. Vet. Diagn. Invest. 6, 492-493.

Reisen, W. K., Monath, T. P., 1988. Western equine encephalomyelitis, p. 89-137 *In* Monath, T. P. (ed.), The Arboviruses: Epidemiology and Ecology, Vol. V. CRC Press, Inc., Boca Raton, Florida.

Reeves, W. C., Hutson, G. A., Bellamy, R. E., Scrivani R. P., 1958. Chronic latent infections of birds with western equine encephalomyelitis virus. Proc. Soc. Exp. Med. 97, 733-736.

Rikkonen, M., Peränen, J., Kääriäinen, L., 1994. ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2. J. Virol. 68, 5804-5810.

Rikkonen, M., 1996. Functional significance of the nuclear-targeting and NTP-binding motifs of Semliki Forest virus nonstructural protein nsP2. Virology 218, 352-361.

Rümenapf, T., Strauss, E. G., Strauss, J.H., 1994. Subgenomic mRNA of Aura alphavirus is packaged into virions. J. Virol. 68, 56-62.

Rümenapf, T., Strauss, E. G., Strauss, J. H., 1995. Aura virus is a New World representative of Sindbis-like viruses. Virology 208, 621-633

Russo, A. T., White, M. A., Watowich, S. J., 2006. The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. Structure 14, 1449-1458.

Ryman, K. D., Klimstra, W. B., 2008. Host responses to alphavirus infection. Immunol. Rev. 225, 27-45.

Sawicki D. L, Sawicki S. G., 1993. A second nonstructural protein functions in the regulation of alphavirus negative-strand RNA synthesis. J. Virol. 67, 3605-3610.

Sawicki, D. L., Perri, S., Polo, J. M., Sawicki, S. G., 2006. Role for nsP2 proteins in the cessation of alphavirus minus-strand synthesis by host cells. J. Virol. 80, 360-371.

Scott, T. W., Bowen, G. S., Monath, T. P., 1984. A field study on the effects of Fort Morgan virus, an arbovirus transmitted by swallow bugs, on the reproductive success of cliff swallows and symbiotic house sparrows in Morgan County, Colorado, 1976. Am. J. Trop. Med. Hyg. 33, 981-991.

Scott, T. W., Weaver, S. C., 1989. Eastern equine encephalomyelitis virus: epidemiology and evolution of mosquito transmission. Adv. Virus Res. 37, 277-328.

Shirako, Y., Niklasson, B., Dalrymple, J. M., Strauss, E. G., Strauss, J. H., 1991. Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses. Virology 182, 753-764.
Shirako, Y., Strauss, J. H., 1994. Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus-strand RNA synthesis. J. Virol. 68, 1874-1885.

Spuul, P., Salonen, A., Merits, A., Jokitalo, E., Kääriäinen, L., Ahola, T., 2007. Role of the amphipathic peptide of Semliki forest virus replicase protein nsP1 in membrane association and virus replication. J. Virol. 81, 872-883.

Srihongse, S., Galindo, P., 1967. The isolation of eastern equine encephalitis virus from Culex (Melanoconion) taeniopus Dyar and Knab in Panama. Mosq. News 27, 74-76.

Strauss, E. G., Rice, C. M., Strauss, J. H., 1983. Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. Proc. Natl. Acad. Sci. U. S. A. 80, 5271-5275.

Strauss, J. H., Strauss, E. G., 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. 58, 491-562.

Strauss, J. H., Strauss, E. G., 1997. Recombination in alphaviruses. Semin. Virol. 8, 85-94.

Suopanki, J., Sawicki, D. L., Sawicki, S. G., Kääriäinen, L., 1998. Regulation of alphavirus 26S mRNA transcription by replicase component nsP2. J. Gen. Virol. 79, 309-319.

Tamm, K., Merits, A., Sarand, I., 2008. Mutations in the nuclear localization signal of nsP2 influencing RNA synthesis, protein expression and cytotoxicity of Semliki Forest virus. J. Gen. Virol. 89, 676-686.

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

Tomar S., Hardy, R. W., Smith, J. L., Kuhn, R. J., 2006. Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity. J. Virol. 80, 9962-9969.

Tyor, W. R., Wesselingh, S., Levine, B., Griffin, D. E., 1992. Long term intraparenchymal Ig secretion after acute viral encephalitis in mice. J. Immunol. 149, 4016-4020.

Ubol, S., Levine, B., Lee, S. H., Greenspan, N. S., Griffin D. E. 1995. Roles of immunoglobulin valency and the heavy-chain constant domain in antibody-mediated

downregulation of Sindbis virus replication in persistently infected neurons. J. Virol. 69, 1990-1993.

Vasiljeva, L., Merits, A., Auvinen, P., Kääriäinen, L., 2000. Identification of a novel function of the alphavirus capping apparatus. RNA 5'-triphosphatase activity of Nsp2. J. Biol. Chem. 275, 17281-17287.

Walder, R., Suarez, O. M., and Calisher, C. H. 1984. Arbovirus studies in the Guarija region of Venezuela: Activities of eastern equine encephalitis and Venezuelan equine encephalitis viruses during an interepizootic period. Am. J. Trop. Med. Hyg. 33, 699-707.

Weaver, S. C., Hagenbaugh, A., Bellew, L. A., Netesov, S. V., Volchkov, V. E., Chang, G. J., Clarke, D. K., Gousset, L., Scott, T. W., Trent, D. W., 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., Hagenbaugh, A., Bellew, L. A., Gousset, L., Mallampalli, V., Holland, J. J., Scott, T. W., 1994. Evolution of alphaviruses in the eastern equine encephalomyelitis complex. J. Virol. 68, 158-169.

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

Weaver, S. C., Frey, T. K., Huang, H. V., Kinney, R. M., Rice, C. M., Roehrig, J. T., Shope, R. E., Strauss, E. G., 2005. Family Togaviridae, p. 466-483 *In* Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., Ball, L. A. (ed.), Virus Taxonomy: Eighth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, California.

Weiss, B., Rosenthal, R., Schlesinger, S., 1980. Establishment and maintenance of persistent infection by Sindbis virus in BHK cells. J. Virol. 33, 463-474.

CHAPTER 4

REASSORTMENT:

DESCRIPTION OF A NOVEL REASSORTANT EPIZOOTIC HEMORRHAGIC DISEASE VIRUS CONTAINING RNA SEGMENTS DERIVED FROM BOTH EXOTIC AND ENDEMIC SEROTYPES[1]

[1]Allison, A. B., V. H. Goekjian, A. C. Potgieter, W. C. Wilson, D. J. Johnson, P. P. C. Mertens, and D. E. Stallknecht. 2010. *Journal of General Virology* 91:430-439. Reprinted here in unabridged format with permission of publisher.

Abstract

Epizootic hemorrhagic disease virus (EHDV) is a *Culicoides*-transmitted orbivirus that infects domestic and wild ruminants and is provisionally distributed throughout Africa, North America, Australia, East Asia, and the Middle East. Historically, of the seven proposed serotypes of EHDV, only EHDV-1 and EHDV-2 have been reported from North America. In 2006, EHDV isolates were recovered from moribund or dead white-tailed deer (*Odocoileus virginianus*) in Indiana and Illinois that could not be identified as either EHDV-1 or EHDV-2 by virus neutralization tests or by serotype-specific RT-PCR. Additional serological and genetic testing identified the isolates as EHDV-6, a serotype that, although originally described from Australia, has recently been recognized as an emerging pathogen of cattle in Morocco, Algeria, and Turkey. In 2007-2009, EHDV-6 was isolated from white-tailed deer in Missouri, Kansas, Texas, and Michigan, suggesting that the virus is capable of overwintering and that it may become, or already is, endemic in a geographically widespread region of the United States. Genetic characterization of the virus indicates that it is a reassortant, such that the outer capsid proteins determining serotype specificity (VP2 and VP5) are derived from exotic EHDV-6, while the remaining structural and nonstructural proteins were apparently obtained from indigenous EHDV-2 (Alberta).

Introduction

Epizootic hemorrhagic disease virus (EHDV) is a serogroup of the genus *Orbivirus*, family *Reoviridae*, which infects domestic and free-ranging ruminants. EHDV is arthropod-borne and is maintained in nature in a ruminant-*Culicoides* biting midge cycle. According to current International Committee on the Taxonomy of Viruses (ICTV) designations, there are 10 recognized serotypes/strains of EHDV: serotypes 1 through 8, EHDV-318, and *Ibaraki virus* (IBAV) (Mertens *et al*., 2005). However, based on recent genetic and phylogenetic analyses of the outer capsid proteins responsible for serotype specificity (VP2 and VP5) for all prototype strains, the current classification of the EHDV serogroup has been proposed to be condensed to seven serotypes, with the inclusion of EHDV-3 into EHDV-1, EHDV-318 into EHDV-6, and IBAV into EHDV-2 (Anthony *et al*., 2009).

In North America, two serotypes, EHDV-1 (New Jersey strain), first authoritatively described by Shope *et al*. (1955) during a large-scale die-off of white-tailed deer in the northeastern United States, and EHDV-2 (Alberta strain), originally isolated in southern Alberta, Canada, in 1962 (Chalmers *et al*., 1964) and later characterized as being serologically distinct from EHDV-1 (Barber and Jochim, 1975), are known to be endemic. Three serotypes have been recognized in Africa: EHDV-3 (proposed to be EHDV-1) and EHDV-4, which were initially isolated from *Culicoides* spp. in Nigeria in 1967 and 1968, respectively (Lee *et al*., 1974), and EHDV-318 (proposed to be EHDV-6), which was first isolated from cattle in Bahrain in 1983 and subsequently described from sentinel calves in the Sudan (Mohammed and Mellor, 1990; Rabenau *et al*., 1993; Anthony *et al*., 2009). EHDV serotypes 5, 7, and 8, along with the

prototype strain of EHDV-6 (CSIRO 753), were initially isolated from sentinel cattle herds in Queensland and the Northern Territory of Australia during 1977-1982 (St. George *et al*., 1983). Additionally, topotypes of EHDV-1 and EHDV-2 are present in Australia, which although genetically divergent from analogous North American serotypes, share common neutralization epitopes (Gould and Pritchard, 1991; Weir *et al*., 1997). IBAV, which is now recognized, as it has been previously suggested, as a topotype of EHDV-2, has been associated with periodic outbreaks in cattle in East Asia since its first description during a large-scale epizootic in Japan involving 40,000 cattle in 1959 (Omori *et al*., 1969; Bak *et al*., 1983; Liao *et al*., 1996; Uchinuno *et al*., 2003; Anthony *et al*., 2009).

Although all ruminants appear to be susceptible to EHDV infection, the clinical course of disease differs substantially depending upon the species infected. In a generalized scheme, domestic ruminants (i.e., cattle, sheep, and goats) infected with EHDV tend to develop inapparent to mild infections, with the latter regressing quickly without serious complications (MacLachlan and Osbourn, 2004). The anomaly to inapparent EHDV infection in cattle is IBAV, in which mortality rates as high as 10% have been documented during field outbreaks and clinical disease has been reproduced experimentally in calves (Omori *et al*., 1969). However, in recent years, the recognition of EHDV-associated morbidity and mortality in cattle from many parts of the world has increased dramatically. EHDV-7 was recently implicated in cattle outbreaks in Israel and, although mortality rates were low (<1%), morbidity rates ranged from 5-80%, with clinical signs reported in 105 different herds (Yadin *et al*., 2008). Additionally, morbidity and mortality associated with EHDV infection in cattle was reported from

Reunion Island in 2003 (serotype not disclosed; Bréard *et al*., 2004), Morocco and Algeria in 2006 (reported as EHDV-318/EHDV-9 and later reclassified as EHDV-6; Yadin *et al*., 2008; Anthony *et al*., 2009), and western Turkey in 2007 (EHDV-6; Temizel *et al*., 2009), suggesting that the geographical distribution and/or the recognition of clinical disease associated with EHDV infection (primarily serotype 6) is increasing.

During September-October 2006 in Indiana and Illinois, six virus isolates were recovered from moribund or dead white-tailed deer that were identified by RT-PCR as EHDV using serogroup-specific primers (i.e., they detect both EHDV-1 and EHDV-2); however, these isolates could not be neutralized with EHDV-1 or EHDV-2-specific antisera, nor could RNA be amplified with EHDV-1 or EHDV-2-specific primers directed against VP2, the gene predominately responsible for serotype specificity. Subsequent virus neutralization tests with exotic EHDV antisera, along with RT-PCR using primers directed against the VP2 gene of exotic EHDV serotypes, identified the viruses as EHDV-6, a serotype that, prior to 2006, had formerly only been described from Australia. In 2007, EHDV-6 was detected again in the Midwestern United States, this time in Missouri, suggesting the virus may be capable of overwintering in northern temperate areas (39.97° N latitude). In 2008 and 2009, EHDV-6 was recovered from Kansas, Texas, and Michigan, demonstrating that the virus is geographically widespread and suggesting that it may (or already has) become endemic in the United States. Moreover, the newly identified isolates appear to be all derived from the same reassortment event, such that the two surface antigens (VP2 and VP5) were obtained from EHDV-6, while multiple nonstructural (NS1, NS3) and structural (VP1, VP3, VP7) genes were derived from EHDV-2 (Alberta). As the virus has retained the serotype

specificity of EHDV-6, we tentatively propose the name EHDV-6 [Indiana (strain)] for the EHDV-6/EHDV-2 reassortant.

Methods

Case histories

EHDV-6 (Indiana) was isolated from 14 individual white-tailed deer on nine occasions during 2006-2009. Four cases (cases 1-4), involving six deer, occurred in Indiana and Illinois in 2006; one case (case 5), involving a single deer, occurred in Missouri in 2007; three cases (cases 6-8), involving five deer, occurred in Texas and Kansas in 2008; and a single case (case 9), involving two deer, occurred in Michigan in 2009 (Fig. 4.1). Clinical cases (CC) 304-06 and 311-09, shown as cases 3 and 9, respectively, in Fig. 4.1, represented the only submissions that were wild white-tailed deer; the remaining 11 isolates were obtained from captive deer herds. Depending upon the case, tissues samples received for virus isolation by the Southeastern Cooperative Wildlife Disease Study (SCWDS) included spleen, lung, liver, lymph node, and blood. Case histories are presented in chronological order.

Case 1: CC 286-06 - 22 September 2006 - in proximity of Newport, Vermillion County, Indiana - a 4-month-old male and adult doe were observed moribund by a private deer farm owner and the buck was subsequently euthanized and submitted by the Indiana Division of Fish and Wildlife to the Indiana Animal Disease Diagnostic Laboratory at Purdue University for necropsy.

Case 2: 06-29537 - 24 September 2006 - Milford, Iroquois County, Illinois - a 1.5-year-old doe was found by a private deer farm owner exhibiting apparent abdominal pain and salivating excessively prior to becoming immobile, although remaining coherent. The deer was found dead the following morning and was taken to the Veterinary Diagnostic Laboratory (VDL) at the University of Illinois at Urbana Champaign (UIUC) for necropsy.

Case 3: CC 304-06 - 6 October 2006 - in proximity of Newcastle, Henry County, Indiana - a 1.5-year-old, free-ranging doe was observed moribund by an archery hunter. The deer was subsequently euthanized by Indiana Department of Natural Resources personnel and local veterinarians.

Case 4: 06-30872-A, B, and C - 16-17 October 2006 - Milford, Iroquois County, Illinois - three 4-5 month-old captive male deer were found dead on the property of the submitter of case 2 and were subsequently sent to the VDL at the UIUC for necropsy.

Case 5: CC 433-07 - 5 September 2007 - Jamesport, Grundy County, Missouri - a 4-year-old doe was found dead by a private deer farm owner and sent to the VDL at the UIUC for necropsy.

Case 6: 08-750 - 30 April 2008 - San Diego, Duval County, Texas - a female deer (age unknown) was found moribund by a deer ranch owner exhibiting mucosal/nasal redness, shortness of breath, and open-mouth breathing, and died within 48 hours of exhibiting clinical signs.

Case 7: 08-814, 08-835, and 08-924 - 22-24 September 2008 - in proximity of Patillo, Palo Pinto County, Texas - four captive deer (sexes/ages unknown) were found dead by a ranch owner. Bluetongue virus serotype 12 (BTV-12), another exotic orbivirus

never previously detected in the United States, was isolated from the fourth submission (*08-812*).

Case 8: 08-63835 - 1 October 2008 - Haven, Reno County, Kansas - a juvenile male deer was found (suddenly) dead by a private deer farm owner and sent to the Kansas State Veterinary Diagnostic Laboratory for necropsy.

Case 9: CC 311-09-A and B - 24-25 September 2009 - Fenton, Livingston County, Michigan - Two free-ranging female deer, aged 0.5- and 4.5-years-old, were found dead by a private citizen and reported to the Michigan Department of Natural Resources. Necropsies demonstrated clinical lesions consistent with hemorrhagic disease including facial edema, edematous lungs, and massive hemorrhage in major organs. Estimates of 200 dead deer were reported from the area.

Virus isolation and L2 RNA segment purification

Virus isolation was performed on tissue and blood samples using cattle pulmonary artery endothelium (CPAE) cells [American Type Culture Collection (ATCC)]. CPAE cells were propagated in growth media [1X minimal essential media (MEM), 2.2g/l $NaHCO_{3,}$ 10% fetal bovine serum (FBS), 100 units/ml penicillin, 100µg/ml streptomycin, 0.25µg/ml amphotericin B] (Sigma).

For virus isolation from tissue, samples (~0.5cm^3) were mechanically homogenized in 650µl of virus isolation (VI) media [1X MEM, 2.2g/l $NaHCO_{3,}$ 20% FBS, 400 units/ml penicillin, 400µg/ml streptomycin, 1µg/ml amphotericin B]. Homogenized tissues were centrifuged (6700 x *g* for 10 min) to pellet debris and an

aliquot (100µl) of clarified supernatant was used to inoculate 3-day-old CPAE cell culture in a 12.5cm^2 flask or 12-well plate format.

For virus isolation from blood, heparinized whole blood (1ml) was washed in Dulbecco's phosphate buffered saline and cells were pelleted by light centrifugation (720 x *g* for 4 min). The wash step was repeated three times and an aliquot (100µl) of the final cell pellet was diluted 1:10 in VI media and then sonicated for ~20 sec at 5 watts using a VirSonic 100 Ultrasonic Cell Disrupter (VirTis). An aliquot of the sonicated solution (100µl) was then used to inoculate CPAE cell culture as per the tissue samples.

For wells exhibiting cytopathology, cell culture supernatant was harvested for 1) virus neutralization testing, 2) RNA extraction using a QIAamp® Viral RNA Mini kit (Qiagen) according to the manufacturer's instructions, and 3) stock virus. In order to gain sequence information of the VP2 gene, stock virus was inoculated into three 75cm^2 flasks of Madin-Darby bovine kidney cells (ATCC) and virus was precipitated on day 6 post-inoculation using polyethylene glycol as described by Killington *et al*. (1996). Extracted RNA was electrophoresed on a 1% agarose ethidium bromide (EtBr)-stained TAE gel and the L2 RNA segment was excised and purified from agarose using a QIAquick® Gel Extraction Kit (Qiagen) according to the manufacturer's instructions.

Virus serotyping

Virus neutralization tests against EHDV and bluetongue virus (BTV) serotypes currently considered indigenous to the United States (EHDV serotypes 1 and 2; BTV serotypes 2, 10, 11, 13, and 17) was performed as described previously (Abdy *et al*., 1999). For virus neutralization tests using exotic EHDV antiserum, the 2006 isolates

were sent to the National Veterinary Services Laboratory (NVSL), Ames, Iowa, USA. The isolates were then sent to the Institute for Animal Health (IAH), Pirbright Laboratory, Woking, UK, for confirmatory testing using VP2-specific primers directed against all EHDV serotypes.

RT-PCR and phylogenetic analysis

All EHDV isolates were analyzed by RT-PCR using serogroup (i.e., amplify both EHDV-1 and EHDV-2) and serotype [i.e., amplify EHDV-1, EHDV-2, or EHDV-6 (Indiana)] specific primers (Supplementary Table S4.1). Serogroup-specific primers to NS1, NS3, VP1, VP3, VP5, and VP7 listed in Supplementary Table S4.1 have not been tested on EHDV serotypes other than 1 and 2. Serotype-specific primers against VP2 (primary and nested) were developed for EHDV-1, EHDV-2, and EHDV-6 (Indiana). Serotype-specific primers against VP5 (primary only) were also made for EHDV-6 (Indiana). Decamer oligonucleotides (not shown) were originally used in order to obtain VP2 sequence for EHDV-6 (Indiana). Primers based on the sequence of the South African EHDV-6 M44/96 isolate (not shown) were originally used in order to obtain VP5 sequence for EHDV-6 (Indiana).

Single-tube RT-PCRs (50 µl) were set up using 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 0.1% Triton® X-100, 1.5 mM $MgCl_2$, 250 µM deoxynucleotide triphosphates, 0.5 µM of each primer, 2 U of AMV reverse transcriptase (Promega), 1 U of *Taq* DNA polymerase (Promega), and 5 µl of extracted RNA. Cycling parameters were reverse transcription at 42 °C for 20 min, followed by 40 cycles of denaturation at 94 °C for 1 min, annealing at 50 °C for 1 min, and extension at 72 °C for 1 min, with a final

extension step at 72 °C for 10 min. Amplicons were electrophoresed on a 2% agarose EtBr-stained TAE gel, excised and purified from agarose, cloned using a PCR Cloning^{Plus}® Kit (Qiagen), and recombinant plasmid was subsequently purified using a QIAprep® Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. Sequencing was performed using an Applied Biosystems Inc. 3100 Genetic Analyzer (Foster City, California, USA). Phylogenetic analysis of VP2, VP5, and VP7 amino acid sequences by neighbor-joining and maximum parsimony methods were conducted using the Molecular Evolutionary Genetics Analysis 4 (MEGA4) program (Tamura *et al*., 2007).

Results

Virus serotyping

All six 2006 isolates, identified as EHDV using serogroup-specific primers directed against the NS1 gene, were not neutralized by antisera to EHDV or BTV serotypes currently considered indigenous to the United States (EHDV serotypes 1 and 2; BTV serotypes 2, 10, 11, 13, and 17). Isolates were then sent to the NVSL for further virus neutralization testing using antisera to exotic EHDV serotypes. A significant reduction in cytopathology was observed in wells inoculated with the 2006 viruses and EHDV-6 antiserum (CSIRO 753); antisera to EHDV serotypes 3 (1), 4, 5, 7, and 8 did not neutralize the 2006 viruses. By virus neutralization, the viruses were therefore preliminarily serotyped as EHDV-6. The isolates were then sent to the IAH for confirmatory testing and were screened with serotype-specific primers directed against

the VP2 gene of all EHDV serotypes. RT-PCR products were amplified from RNA derived from the 2006 isolates using VP2-specific primers directed against EHDV-6 (primers not shown); no amplicons were generated using VP2 primers specific to other EHDV serotypes. Based on virus neutralization and VP2-specific RT-PCR, the 2006 isolates were determined to contain an L2 RNA segment derived from EHDV-6.

RT-PCR and phylogenetic analysis

RT-PCR directed against multiple RNA segments was performed on the EHDV isolates. RT-PCR products were amplified using serogroup-specific primers against NS1, NS3, VP1, VP3, and VP7 (Supplementary Table S4.1). CLUSTAL alignment (VP1) or BLAST analysis (NS1, NS3, VP3, and VP7) of the RT-PCR products indicated that the viruses were very similar in nucleotide and deduced amino acid sequence to North American serotypes, particularly EHDV-2 (Fig. 4.2 and Table 4.1). However, serotype-specific primers (both primary and nested) directed against the VP2 gene of EHDV-1 or EHDV-2 did not result in detectable amplicons when using RNA from the 2006 isolates as template. Additionally, two sets of degenerate VP5-specific primers that detect both EHDV-1 and EHDV-2 (Supplementary Table S4.1), did not amplify any of the isolates, suggesting that both of the outer capsid proteins (VP2 and VP5) of the reassortants were derived from EHDV-6.

Random cloning of the L2 segment of clinical case isolate 304-06 (hereafter referred as 304-06) resulted in the generation of two overlapping clones. The contig generated by these two clones was 1,014 nucleotides in length, resulting in a deduced open reading frame of 338 amino acids. CLUSTAL alignment of the 304-06 nucleotide

contig with the cognate VP2 sequences of EHDV-1 and EHDV-2 indicated an identity of 29% and 30%, respectively. Protein BLAST analysis of the translated contig revealed it contained a limited amino acid identity to EHDV serotypes currently present in GenBank: a 40% identity to EHDV-1, a 38% identity with Australian and North American EHDV-2, and a 37% identity to Japanese EHDV-2 (IBAV). CLUSTAL alignment of the VP2 338 amino acid contig from 304-06 versus the cognate region of available VP2 sequences of other serotypes [EHDV-1, EHDV-2 (including North American, Australian, and Japanese topotypes), and EHDV-6 (South African strain M44/96)] revealed a number of amino acid motifs (3 consecutive aa) [^{1}-TCYDL-^{5}, ^{183}-RAIW-^{186}, ^{156}-NIR-^{158}, ^{232}-CDR-^{234}, ^{256}-YKW-^{258}, ^{334}- DAY(V/I)D-^{338}] that appear to be conserved among the serotypes.

Alignment of the VP2 nucleotide and amino acid contig from 304-06 with a 1996 bovine EHDV-6 isolate from South Africa (M44/96) revealed a 73% and 79% identity, respectively, providing comparative genetic evidence to collaborate the serological data that EHDV-6 was the parental virus of the L2 RNA segment (Table 4.1). Additionally, nucleotide and amino acid comparisons of the complete coding sequence of VP5 between 304-06 and M44/96 indicated an 80% and 94% identity, respectively, confirming that both of the outer capsid proteins (VP2 and VP5) of 304-06 were derived from EHDV-6. In comparison, EHDV-1 and EHDV-2 VP5 amino acid sequences shared a 65% and 75% identity, respectively, to 304-06 (Table 4.1). Phylogenetic analysis of both of the outer capsid proteins (VP2 and VP5) demonstrated that 304-06 and M44/96 formed a distinct clade (with 100% bootstrap support) separate from other EHDV serotypes/topotypes (Fig. 4.3A and B).

In contrast, the VP7 sequence from 304-06 (975 nt) shared a 99% nucleotide identity to EHDV-2, while displaying an 89% nucleotide identity to EHDV-6 M44/96, confirming that the parental virus of the S7 RNA segment was EHDV-2 (Table 4.1). Phylogenetic analysis of the VP7 amino acid sequences for available EHDV serotypes grouped 304-06 with an EHDV-2 Alberta isolate from the United States (Fig. 4.3C). Interestingly, the EHDV-6 (Indiana) [304-06] and EHDV-2 (Alberta) VP7 sequences (211-91) from the United States clustered with the EHDV-6 isolate from Africa (M44/96) to form a distinct clade, separate from EHDV-2 topotypes from Australia and Asia. Sequence comparison of additional structural genes (VP1 and VP3) from 304-06 versus endemic serotypes and EHDV-6 also indicated that EHDV-2 was the likely parental virus of their corresponding RNA segments (Table 4.1).

Comparison of partial sequences of two nonstructural genes (NS1 and NS3) from 304-06 with the prototype strains of all EHDV serotypes demonstrated that 304-06 shared the highest nucleotide identity (97%) in both genes to EHDV-2 SV-124, isolated in Alberta in 1962 (Table 4.2). Additionally, the partial NS3 sequence (448 nt) for 304-06 shared an even higher nucleotide identity (99.8%) when compared against more recent EHDV-2 isolates from the Midwestern and central United States (Murphy *et al*., 2005), verifying that the S10 segment of 304-06 is derived from EHDV-2. When comparing the amino acid identities of NS1 and NS3 of 304-06 to other EHDV serotypes, the only serotype to display >90% identity in both proteins, other than endemic EHDV-1 and EHDV-2, was EHDV-6 M44/96 (97% in both NS1 and NS3) (Table 4.2).

Based on sequence data obtained from partial cloning of the L2 RNA segment from 304-06, two sets of primers were developed for diagnostic use in order to amplify

the VP2 gene of the reassortant viruses from clinical samples (designated as EHDV-6Ind-VP2; see Supplementary Table S4.1). EHDV-6 (Indiana) isolates from 2007-2009 were identified using these primer sets. The primer set, EHDV-6Ind-VP2-1242/1506, amplifies a 265bp fragment, while the second internal set, EHDV-6Ind-VP2-1280i/1479i, are nested primers that amplify a 200bp fragment; hence, these primers are useful in detecting RNA directly from tissue if virus isolation cannot be performed on clinical samples. Partial VP2 sequence comparisons of 304-06 with the single EHDV-6 (Indiana) isolate from 2007 (CC isolate 433-07) revealed a 98% (482/490) nucleotide and a 99% (162/163) amino acid identity (not shown). Seven of the eight nucleotide substitutions were synonymous third base transitions. The single non-synonymous nucleotide change resulted in a conservative substitution of basic amino acids (K→R) between 304-06 and 433-07, respectively. Additional analysis of other EHDV-6 (Indiana) isolates from 2006-2009 indicated similar nucleotide identities in the VP2 gene (≥98%) irrespective of geographical location. Sequence analysis of multiple genes from all 14 isolates suggests that each has the same reassortment configuration (i.e., L2 and M6 are derived from EHDV-6, while the remaining RNA segments are derived from EHDV-2; not shown), although these results are preliminary and require a more comprehensive analysis.

The EHDV-6 (Indiana) sequences from 304-06 were deposited in GenBank under the following accession numbers: NS1 (GQ387643), NS3 (GQ387644), VP1 (GQ387634), VP2 (GQ385939), VP3 (GQ387637), VP5 (GQ387638) and VP7 (GQ387641), as were sequences to EHDV-1 VP1 (GQ387635), EHDV-2 VP1 (GQ387636), and EHDV-2 VP5 (GQ387640).

Discussion

Sequence data obtained from the U.S. EHDV isolates indicates that they are all derived from the reassortment of at least two serotypes: EHDV-2 and EHDV-6. The L1, L3, M5, S7, and S10 RNA segments of the reassortant, encoding the VP1, VP3, NS1, VP7, and NS3 genes, respectively, are apparently derived from EHDV-2 (Alberta) (Table 4.1). As VP1, VP3, NS1, and NS3 are highly conserved between EHDV-1 and EHDV-2 (Fig. 4.2 and Table 4.1), ascribing EHDV-2 as the parental virus from which their respective segments were obtained is tentative. However, as partial sequencing of VP7 (975 nt) indicated that EHDV-6 (Indiana) shared a 78% and 99% nucleotide identity to EHDV-1 and EHDV-2, respectively (Table 4.1), it is presumed that EHDV-2 was likely the parental virus of other less variable RNA segments that share a very close identity between EHDV-1 and EHDV-2. Currently, other than the M5 (NS1) and S10 (NS3) sequences of the prototype EHDV strains recently reported by Wilson *et al*. (2009), there are no gene sequences for any other RNA segments for EHDV serotypes 3 (1), 4, 5, 6, 7, or 8 available in public databases to allow for comparison.

Based on genetic comparisons to a bovine EHDV-6 isolate from South Africa (M44/96), in conjunction with virus neutralization testing with exotic EHDV antisera, the L2 RNA segment (encoding the VP2 gene predominately responsible for serotype specificity and mammalian receptor binding/cell entry) of the 2006 isolates was derived from EHDV-6 (Mertens *et al*., 1989; Iwata *et al*., 1992; Hassan and Roy, 1999). Further genetic comparisons of the VP2 gene of all 14 isolates recovered from 2006-2009 indicated that they all share the same L2 RNA segment and are likely derived from the same reassortment event. Although the VP2 amino acid identity shared between 304-06

and M44/96 (covering approximately a third of the coding region of VP2) was only 79%, it is roughly equivalent to the divergence observed between topotypes of EHDV-2 (e.g., Australian and North American EHDV-2 VP2 amino acid sequences share a 73% identity) and is also suggestive that the North American and African strains of EHDV-6 did not recently diverge from one another (i.e., M44/96, or a virus very closely related to it, is not the immediate parental virus of 304-06). Additionally, the M6 RNA segment, encoding the VP5 gene, is also apparently derived from EHDV-6. The amino acid identity of the complete coding region of VP5 (527 aa) between 304-06 and M44/96 was relatively higher (94%) in comparison to that of VP2, indicative of the less selection pressure VP5 is under from the host immune response due to its presumed buried location in the outer capsid (Hewat *et al*., 1994; Iwata *et al*., 1991). The fact that VP2 and VP5 must interact directly with one other, albeit not extensively, to form the outer capsid in the related *bluetongue virus* (BTV) (Nason *et al*., 2004), and VP2, and to a lesser extent VP5, are more variable in amino acid sequence than other EHDV proteins, likely imparts some degree of functional and/or conformational constraints that may inhibit the survival of reassortants bearing chimeric outer surface proteins derived from two different EHDV serotypes. Additionally, reassortants containing outer capsid proteins derived from different parental viruses have been shown to be difficult to produce and isolate *in vitro*, suggesting that VP2/VP5 reassortants may be exceedingly rare in nature (Mertens, 1999).

Interestingly, Nason *et al*. (2004) demonstrated that, with BTV, the most extensive protein interactions were not between VP2 and VP5, but rather between VP7 and the outer capsid proteins. Comparative analysis of the nucleotide and amino acid sequence of VP7 between M44/96 and 304-06 (with VP7 being derived from EHDV-2

Alberta) demonstrated an 89 and 99% identity, respectively (Table 4.1). In comparison, the VP7 sequence of EHDV-1 is only 78 and 94% identical at the nucleotide and amino acid level to that of EHDV-2. Additionally, phylogenetic analysis of the VP7 amino acid sequence indicated that EHDV-2 Alberta, along with 304-06, formed a monophyletic group with the EHDV-6 isolate from South Africa, which was distinct from other EHDV-2 topotypes (Fig. 4.3C). This grouping of EHDV-6 and EHDV-2 Alberta VP7 sequences was supported by high bootstrap values and was also present when the sequences were compared using alternative (maximum parsimony) methods. That fact that the VP7 sequences of EHDV-2 Alberta and an African EHDV-6 isolate are more closely related to one another than EHDV-2 Alberta is to Australian and Asian EHDV-2 topotypes is intriguing and may also suggest that it was this direct compatibility of the protein-protein interactions between VP7 of EHDV-2 with that of VP2 and VP5 of the parental EHDV-6 that facilitated the viability and survival of these new reassortants in nature.

Although the extremely high nucleotide identity (99.8%) of the partial NS3 sequence from 304-06 indicates that EHDV-2 is the parental virus of the S10 segment, M44/96 and 304-06 shared a 97% amino acid identity in NS3, while other exotic serotypes, including the Australian EHDV-6 prototype (CSIRO 753), shared relatively lower identities with 304-06, ranging from 87-88% (Table 4.2). For BTV, NS3 has been demonstrated to play a functional role in viral egress from *Culicoides* cells by binding not only to cellular proteins involved in exocytosis, but also with VP2, thereby potentially acting as a bridging molecule between cellular export proteins and assembled virions (Beaton *et al*., 2002; Celma and Roy, 2009). Although the interaction between NS3 and VP2 may potentially be limited to short conserved motifs in both proteins

(Celma and Roy, 2009), the high amino acid identity of 304-06 NS3 to that of M44/96 NS3 (97%) in comparison to other serotypes (87-88%) may suggest that this could have facilitated the functional interaction between NS3 of EHDV-2 and VP2 of the parental EHDV-6 during viral egress in North American vectors (i.e., *Culicoides sonorensis*). In addition, as VP7, which exhibits a very high amino acid identity (99%) between M44/96 and EHDV-2 (Table 4.1), has been demonstrated to be the viral protein involved in receptor attachment and infection in *Culicoides* cells (Mertens *et al*., 1996; Xu *et al*., 1997), this not only potentially suggests that the structural viability of the reassortant (i.e., VP2/VP5 interaction with VP7) was facilitated by the overall compatibility of the proteins between EHDV-2 and EHDV-6, but its subsequent transmission (i.e., VP7 interaction with North American *Culicoides* receptors; NS3 interaction with VP2) may have also been dependent, in part, upon the compatibility of these protein interactions.

How the parental EHDV-6, and/or EHDV-6 (Indiana), entered into captive and free-ranging white-tailed deer populations in the United States is unknown. However, there are a number of theoretical mechanisms by which non-indigenous orbiviruses may potentially enter into new areas: 1) importation of infected, exotic cattle, and/or associated germplasm, 2) importation of infected, exotic game ruminants for private hunting facilities or zoological collections, 3) accidental entry of infected *Culicoides* species associated with exotic animal, or possibly agricultural, imports, 4) wind-blown infected *Culicoides* species from the Caribbean or Latin America, 5) illegal use of an unlicensed foreign-derived vaccine, or 6) contamination of therapeutics for use in ruminants.

Although EHDV-6 was historically only described from Australia, the distribution of this serotype may be geographically widespread, as the isolation of the virus from cattle in South Africa, as well as the outbreaks in Morocco, Algeria, and Turkey attest to. As the extent of knowledge of EHDV serotype diversity in many areas where EHDV is known or suggested to be endemic is extremely limited (e.g., South America or Africa), it is likely that the currently proposed geographical distribution of most EHDV serotypes is underestimated. Although nearly all cattle imported into the United States come from either Canada or Mexico (Hoar *et al*., 2004), the fact that EHDV infection in cattle is predominately subclinical, in conjunction with the inevitable mixing of animals from different geographical regions through trade, potentially facilitates the movement of EHDV throughout many areas of the world undetected. This same situation could be applied to exotic ruminants brought into the United States for zoological or private collections. Other than cattle, the duration of EHDV viremia in susceptible exotic ruminants that are currently imported into the United States has not been evaluated (Gibbs and Lawman, 1977; Abdy *et al*., 1999).

Despite the fact that adult *Culicoides* are generally short-lived, they can survive up to 90 days, during which time they may take multiple blood meals (Mellor *et al*., 2000), thereby providing a theoretical mechanism of EHDV entry into the United States through their inadvertent association with live animal, or possibly agricultural, imports. Additionally, the wind dispersal of *Culicoides* species has been suggested as a mechanism for the introduction of exotic orbiviruses from Indonesia into Australia (Parsonson and Snowdon, 1985), North Africa into the Mediterranean region, and subsequently, northern Europe (Gloster *et al*., 2007), and the Caribbean into the United

States (Sellers and Maarouf, 1988). Although serological evidence of EHDV-1 and/or EHDV-2 has been documented in Columbia, Guyana, Suriname, and the Caribbean, no comprehensive field studies have been undertaken in Latin America to ascertain if other described or undescribed serotypes exist in the region (Gumm *et al*., 1984; Homan *et al*., 1985).

Currently, the only commercially available licensed vaccine for EHDV is a live attenuated strain of IBAV (Ohashi *et al*., 1999). As EHDV can be a severe pathogen of white-tailed deer, the production of a vaccine for enzootic North American EHDV serotypes is a major concern to the private white-tailed deer farming industry in the United States. If unlicensed foreign vaccines to EHDV were to be produced, they may be conceived as being beneficial in protecting private collections of North American white-tailed deer. However, there is no evidence to suggest that unlicensed live EHDV vaccines, either foreign or domestically-derived, are currently being used in the United States. Additionally, there is the potential of inadvertent contamination of other ruminant vaccines or therapeutics, as EHDV-6 has been previously implicated as a bovine serum contaminant that was isolated from Chinese hamster ovary cells being used for the production of recombinant protein for human therapeutic use (Rabenau *et al*., 1993).

The reassortment event between EHDV-6 and EHDV-2 presumably took place in the United States, giving rise to EHDV-6 (Indiana). Whether parental EHDV-6 is also circulating in the United States is unknown. Alternatively, reassortment may have been a necessary prerequisite for the efficient transmission and consequential survival of parental EHDV-6 in North America. Whether other reassortants containing different gene segment constellations of the two parental viruses are also circulating is speculative and

may be difficult to detect if they contain only highly conserved (non-serotype specific) RNA segments from EHDV-6. The fact that an EHDV-6 isolate from South Africa (M44/96) shares ≥86% nucleotide identity in partial sequences of NS1, NS3, VP3 and VP7 with EHDV-2 (Alberta) reiterates this assumption. However, as aforementioned, since the geographical distributions of most EHDV serotypes are not well delineated and there is some serological evidence of EHDV-2 (Alberta) infection in cattle from Guyana and Columbia (Gumm *et al*., 1984; Homan *et al*., 1985), it is also plausible that the reassortment event may have occurred outside the United States (e.g., Central or South America).

Although the natural reassortment of different serotypes within an orbivirus serogroup (species) has been previously described (e.g., Collison and Roy, 1983), to our knowledge, this is the first report of the field isolation of a reassortant EHDV containing RNA segments derived from both indigenous and exotic serotypes. Reassortment has likely played an important role in orbivirus evolution, contributing to the genotypic and phenotypic diversity observed within each serogroup of the genus (Sugiyama *et al*., 1982; Samal *et al*., 1987a; Samal *et al*., 1987b; Stott *et al*., 1987; Brown *et al*., 1988; Maia and Osburn, 1992). However, the comparative effects that reassortment (i.e., versus parental viruses) has on virus-host interactions (e.g., viral fitness, host virulence, vector competency and transmissibility) has not been extensively studied and requires further research.

When EHDV-6 entered into the United States is also enigmatic. In reality, the parental EHDV-6 and/or EHDV-6 (Indiana) may have been circulating in the United States for many years prior to the detection of the latter. Other than analyzing tissue

samples from archived undiagnosed cases of hemorrhagic disease, the most supposable method for potentially ascertaining the time frame of introduction may be through serotype-specific serological testing of archived ruminant blood samples. Retrospective testing of banked serum from wild and domestic ruminant populations throughout the United States could possibly predict an approximate time frame (and/or location) of the introduction of EHDV-6, provided a comprehensive longitudinal serum bank exists and no other undescribed cross-reactive EHDV serotypes circulate in the United States.

A primary concern of the recognition of EHDV-6 (Indiana) in the United States is attempting to predict the potential impacts of its introduction on wild and domestic ruminant populations. Limited experimental infections with the Australian prototype of EHDV-6 (CSIRO 753) demonstrated that cattle were susceptible to infection as evidenced by a viremia, but clinical disease was not observed; similarly, experimentally–infected sheep became viremic and one of five animals developed mild lesions (fever, malaise, and hyperemia of skin and mucous membranes) of short duration, followed by complete recovery (St. George *et al*., 1983; Uren, 1986). At necropsy, no pathological changes were observed in cattle or sheep infected with EHDV-6. These results are similar as those obtained with experimental infections with EHDV-2, whereby, although a prolonged viremia was detectable in cattle, clinical abnormalities and pathological disease were absent (Abdy *et al*., 1999). Previous experimental infection of sheep with EHDV-2 resulted in low levels of viremia without clinical signs of disease (Fletch and Karstad, 1970; Gibbs and Lawman, 1977).

The underlying factors precipitating the recognition of sporadic cases of reported clinical disease in cattle infected with indigenous EHDV in the United States are

currently unknown (Abdy *et al*., 1999). Nevertheless, in areas where EHDV is endemic, clinical disease is uncommon in cattle and indigenous serotypes do not appear to be a serious concern to either cattle or sheep production. Although it appears unlikely that EHDV-6 (Indiana) will pose a more serious threat to the food animal industry in the United States, the effects that a novel reassortant has on host pathogenicity to North American domestic ruminant populations cannot be predicted and can only be preliminarily assessed by comprehensive *in vivo* pathogenicity testing.

Experimental studies with EHDV in white-tailed deer have indicated that previous exposure to one serotype may result in cross-protection when challenged with a second serotype (Gaydos *et al*., 2002). However, previous exposure to one serotype did not result in a decrease in viremia titer of the challenge serotype, suggesting that deer may still act as amplifying hosts for the superinfecting virus (Gaydos *et al*., 2002). Thus, EHDV-6 (Indiana) may still be able to be transmitted efficiently in deer populations that are immune to EHDV-1 and/or EHDV-2. In areas where EHDV-1 and EHDV-2 are not endemic, sporadic outbreaks of hemorrhagic disease induced by EHDV-6 (Indiana) may go unrecognized unless, as in these cases, virus isolation is attempted from clinical submissions. In areas where indigenous EHDV serotypes and white-tailed deer populations have apparently reached enzootic stability, such as in Texas, large-scale EHDV epizootics have not been previously documented (Stallknecht *et al*., 1996). If EHDV-6 (Indiana) enters into these populations, it is unknown whether the proposed innate resistance and/or acquired immunity to indigenous serotypes will be protective or, on the other hand, severe mortality events, as seen during epizootics in non-endemic regions, will result. If the latter is the case, the movement of EHDV-6 (Indiana) into

these areas might be detected due to the recognition of abnormal levels of white-tailed deer mortality. How EHDV-6 (Indiana) will affect white-tailed deer populations, along with other important game ruminants in the United States that are known to be susceptible to EHDV-induced clinical disease [e.g., mule deer (*Odocoileus hemionus hemionus*), pronghorn (*Antilocapra americana*)], remains to be determined. As in the case with domestic ruminants, the outcome of infection with a novel reassortant exhibiting exotic serotype-specific antigens in indigenous wild ruminant species is speculative until comprehensive experimental infections can be performed.

The recognition of EHDV-6 (Indiana) in the United States was predominately afforded through the presence of a sentinel species (white-tailed deer) that is highly susceptible to EHDV-induced clinical disease. The lack of reported EHDV-induced disease in deer populations outside of North America may consequently be partially due to species composition, reporting bias, and/or the relative population size. Therefore, in areas where large populations of ruminant species that are highly susceptible to clinical disease are absent, EHDV serotypes may circulate undetected unless a proactive surveillance system (e.g., sentinel cattle) is in place for their detection. Additionally, the recognition of EHDV-6 (Indiana) in the United States would have been missed by serogroup-based diagnostic protocols, underscoring the importance of serotyping EHDV isolates by virus neutralization and/or serotype-specific RT-PCR in order to ensure proper identification. Currently, the geographical distribution of EHDV-6 (Indiana) in the United States is unknown, but presumably stretches from at least southern Michigan west to Kansas, and south to Texas (Fig. 4.1). Continued hemorrhagic disease surveillance and virus isolation from clinical submissions, in addition to current and

retroactive serological testing of white-tailed deer or domestic ruminant populations from different geographical regions, will likely help better define the distribution of EHDV-6 (Indiana) in the United States.

Table 4.1. Comparative nucleotide and amino acid identities of selected nonstructural (NS) and structural (VP) genes of EHDV-6 (Indiana) [isolate CC 304-06] versus the cognate region in endemic North American serotypes (EHDV-1 and EHDV-2) and an exotic EHDV-6 isolate from South Africa (M44/96).

EHDV-6 (Indiana)	EHDV-1 (New Jersey)		EHDV-2 (Alberta)		EHDV-6 (South Africa)	
	nt	aa	nt	aa	nt	aa
NS1	96	98	97	97	89	97
NS3	97	100	99	100	86	97
VP1	96	98	96	97	88	95
VP2	29	40	30	38	73	79
VP3	96	99	98	100	89	99
VP5	64	65	69	75	80	94
VP7	78	94	99	100	89	99

Table 4.2. Nucleotide and amino acid comparison of the partial NS1 and NS3 sequences of EHDV-6 (Indiana) [isolate CC 304-06] versus the cognate region in the prototype strains of all EHDV serotypes.

				NS1		NS3	
Serotype	Strain/isolate	Location	Year[c]	nt	aa	nt	aa
EHDV-1	New Jersey	USA	1955	96	98	95	98
EHDV-2	SV-124	Canada	1962	97	97	97	98
EHDV-2	Ibaraki	Japan	1959	-[d]	-	75	87
EHDV-3[a]	Ib Ar 22619	Nigeria	1967	90	95	74	88
EHDV-4	Ib Ar 33853	Nigeria	1968	91	97	74	88
EHDV-5	CSIRO 157	Australia	1977	76	89	76	88
EHDV-6	CSIRO 753	Australia	1981	77	89	77	87
EHDV-6	M44/96	South Africa	1996	89	97	86	97
EHDV[b]	AY351654	Reunion Island	2003	-	-	86	97
EHDV-7	CSIRO 775	Australia	1981	77	89	74	87
EHDV-8	DPP 59	Australia	1982	79	92	75	88

[a]For clarity, the serotype designation for EHDV-3 is based on the current International Committee on the Taxonomy of Viruses classification (Mertens *et al*., 2005).
[b]Untyped EHDV; although VP2 sequence was obtained for EHDV isolates from the Reunion Island outbreak (Bréard *et al*., 2004), the sequence is currently not available in GenBank; hence, no serotype designation can be given. However, VP7 and NS3 amino acid sequences of the Reunion Island isolates share a 100% identity to EHDV-6 M44/96 from South Africa.
[c]Year of isolation of compared sequence is given; however, no isolate date on the Ibaraki NS3 sequence was given in GenBank and, as such, the date of the first description of IBAV is listed.
[d] = sequence not available.

Supplementary Table S4.1. EHDV serogroup and serotype-specific primers used in this study.

Primer name[a]	Sequence (5'-3')
EHDV-NS1-1320	AATGATGAATGGCAGGAGG
EHDV-NS1-1702	ACGATAGTTACGGTTTAGTTA
EHDV-NS3-187	GCAACTATGGCRCAAAAAG
EHDV-NS3-671	GGYGGTATGTCAAATCCA
EHDV-VP1-2564	ACAAACGATGGTGACTAAAG
EHDV-VP1-3063	AGAAATCTCCCAATGGTAAAT
EHDV-VP3-544	AGCATAGAGCGATGATACAGC
EHDV-VP3-934	CGAGGATTGGGTGAAACA
EHDV-VP5-403	GATTTAGATGAWGTGTATMRRTTCGCA
EHDV-VP5-701	GCRCTCATMCYTGCWATYTCCTGTAT
EHDV-VP5-1206	CTTTWTRGGRTTTGATCTKGA
EHDV-VP5-1606	CGTAWCGCATAAAYTCGG
EHDV-VP7-18	ATGGACACGATTGCWGCRAGAG
EHDV-VP7-1060	TAGGCRTTTTGDGCAACAG
EHDV-1-VP2-1763	CTAACATCAAATAGCCAAGA
EHDV-1-VP2-2216	GATAAAACTGTGGAAAATA
EHDV-1-VP2-1893i[b]	TATGTCGCTATGCTTTTTA
EHDV-1-VP2-2066i	AGATACTTCGGAGGTTCAGA
EHDV-2-VP2-49	CGCATAATAATGAAGAAGAG
EHDV-2-VP2-633	AGCGTTTTGTCACATCCTG
EHDV-2-VP2-170i	GCTGAGAACGATAGGAGACG
EHDV-2-VP2-436i	CTAATCCAACTCGCCGC
EHDV-6Ind-VP5-329[c]	GAAATAGAGGAAGAGGAAAAAGGAG
EHDV-6Ind-VP5-1119	ACATTGGGATCTTAAAACGCATCGC
EHDV-6Ind-VP2-1242	TAAATCAAGAAGAAGGAGACTC
EHDV-6Ind-VP2-1506	GCATCTTTTGTGAAATCGTAT
EHDV-6Ind-VP2-1280i	TTAGAGCCAATACTACGAG
EHDV-6Ind-VP2-1479i	ATACCACCTACACCTGATA

[a]Serogroup-specific and serotype-specific primers are indicated by the absense or presence of the numerical serotype designation after EHDV, respectively.
[b]"i" designates the primer pair as an internal (nested) set to be used in conjunction with the preceding primer pair in the table.
[c]EHDV-6 (Indiana) serotype-specific primers are designated as EHDV-6Ind-VP2/VP5-#, with the # representing the approximate nucleotide position based on the corresponding sequence in EHDV-6 M44/96 .

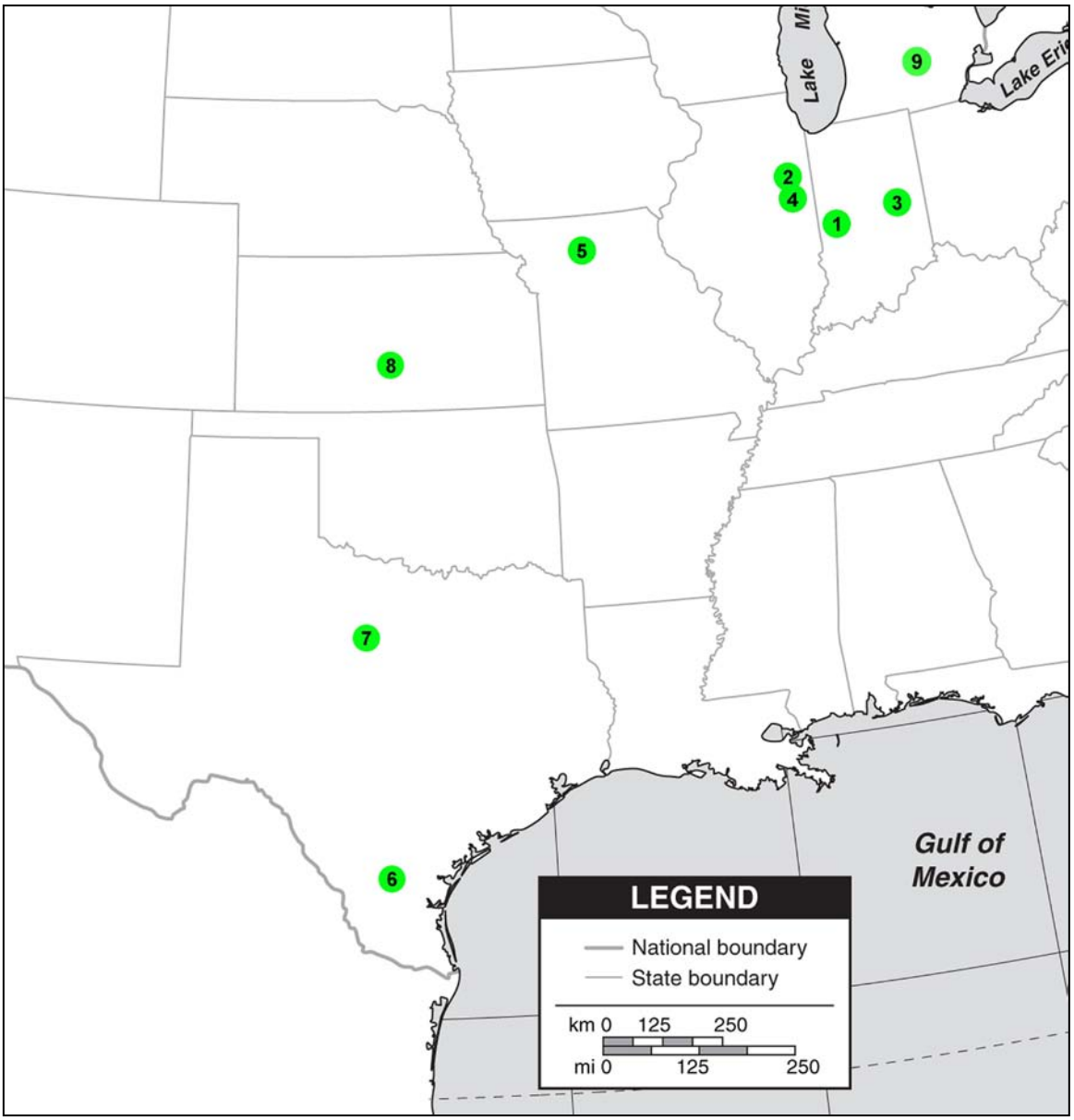
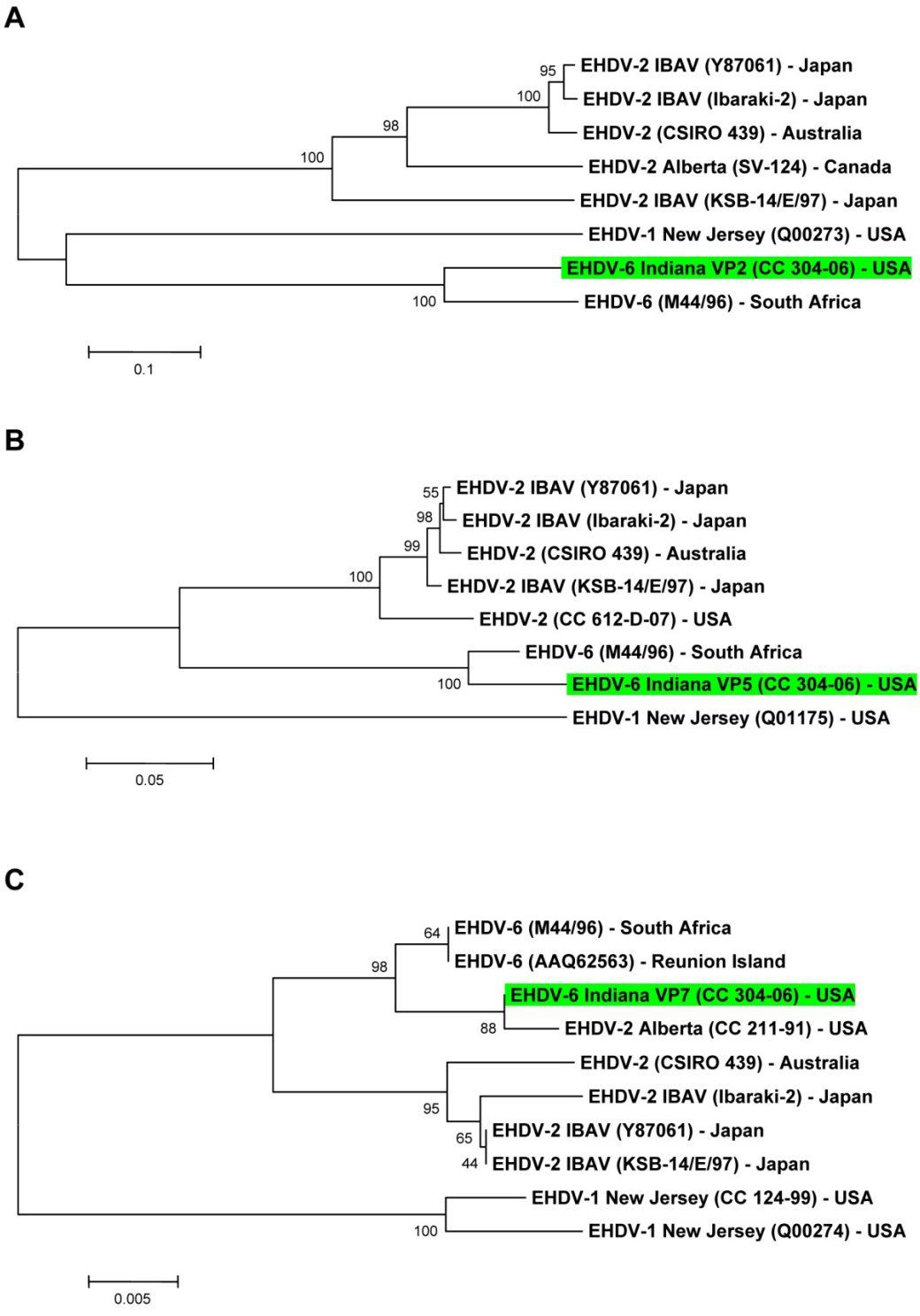


Figure 4.1. Geographical locations of EHDV-6 (Indiana) isolates recovered from free-ranging and captive white-tailed deer by the Southeastern Cooperative Wildlife Disease Study in the United States during 2006-2009. EHDV-6 (Indiana) isolations obtained during 2006, 2007, 2008, and 2009 are represented as cases 1-4 (Indiana and Illinois), case 5 (Missouri), cases 6-8 (Texas and Kansas), and case 9 (Michigan), respectively. See *Case histories* in *Methods* section for details.

```
EHDV-1 VP1  QGYVRSQVQTMVTKVSRGFCHDLAQLILMLKTTFIGAWKMKRTIKEEGIYRDRKFDSNEE 60
EHDV-2 VP1  QGYVRSQVQTMVTKVSRGFCHDLAQLILMLKTTFIGAWKMKRTIKEEGIYRDRKFNSNEE 60
EHDV-6 VP1  QGYVRSQVQTMVTKVSRGFCHDLAQLILMLKTTFIGAWKMKRTIKEEGIYRDRKFDSNEE 60
             ************************************************************:****

EHDV-1 VP1  DGYTLVMLRNPLALYVPIGWNGYGAHPVAINIVMTEEMFLDSMCIGNLDEVMRPILKIRG 120
EHDV-2 VP1  DGYTLVMLRNPLALYVPIGWNGYGAHPVAINIVMTEEMFLDSMCISNLDEVMRPILKIRG 120
EHDV-6 VP1  DGYTLVMLRNPLALYVPIGWNGYGAHPVAINIVMTEEMFLDSMCINSLDEVMRPVLKIRG 120
             ********************************************..*******:*****

EHDV-1 VP1  KIPPTWNETEADKRAIGSETKMSFFSKMARPAVQIALNNREIMEAVEHLPLGDFSPGKLS 180
EHDV-2 VP1  KIPPTWNETEADKRAIGSETKMSFFSKMARPAVQIALNNREIMEAVEHLPLGDFSPGKLS 180
EHDV-6 VP1  KIPPTWNETEADKRAIGSETKMSFFSKMARPAVQIALNNREIMEAVEHLPLGDFSPGKLS 180
             ************************************************************



EHDV-1 VP2  TCYDLSSRLKLRVIGDVDRHRRSMQNVLGR-VIHTGDPKIINRVNQIGSQQFIDRAIGPD 59
EHDV-2 VP2  TCYDLGDHIQLRTIGDVGPRPRDHVDVLGR-THPRGEKHIIRRYGG-DEIKTLTTSMSPD 58
EHDV-6 VP2  TCYDLLKQLELIVIGEIKPSTRERQNVITRQMIPIGTPEITNREPYKNKQTKIQAALGPK 60
             ***** .:::* .**::    *.  :*: *     *  .* .*    ..   :  ::.*.

EHDV-1 VP2  KFELKREIFDRLKALDVDVRKVIREEEASAELDEMGRRWMRDQNVNIVNDIIQSLVKKGS 119
EHDV-2 VP2  EFELKKKILNGDVAIGVEKRNLIKYSNEILQLDDIAASWIRSQNANDLEKIVALLERLGE 118
EHDV-6 VP2  VNELKKEIFSGKYGLKVKYVARLLDDPLITRLDVIAEEWMQRQSDGKVDELCDLLEAKGR 120
              ***::*:.   .: *.    :  .    .** :.  *:: *. . ::.:   *   *

EHDV-1 VP2  RSEKLAHRNEQGMQARFRRTIATNLRDQRQGKEVLNIRGTRGQPEEKKFAAVLLMTGCDI 179
EHDV-2 VP2  KDQKVEPHNSNDIRERFRRKLLQNLQ--KTDGEIRNIRNYHQQDATKRFAAVLIVTMCDT 176
EHDV-6 VP2  QIKMAGTSADYCKKARSRLHGVLKANLVKTTDEIGNIRAVRNENAGSILAAVLVVSACDS 180
             : :      .   : * *     : .  :   *: ***  : :   . :****::: **



EHDV-1 VP3  GAFILHNIPTKDHRGMEIADPEILGVDVKSILPVLTAEHRAMIQHVLDGAIIENGNIATR 60
EHDV-2 VP3  GAFILHNIPTKDHRGMEIADPEILGVDVKSILPVLTAEHRAMIQHVLDGAIIENGNIATR 60
EHDV-6 VP3  GAFILHNIPTKDHRGMEIADPEILGVDVKSILPVLTAEHRAMIQHVLDGAIIENGNIATR 60
             ************************************************************

EHDV-1 VP3  DVDVYLGACSESVYRIYNRLQGYIEAVQLEELRAAVTWLERLGKRKRMTFSQEFLTDFRR 120
EHDV-2 VP3  DVDVYLGACSESVYRIYNRLQGYIEAVQLEELRAAVTWLERLGKRKRMTFSQEFLTDFRR 120
EHDV-6 VP3  DVDVYLGACSESVYRIYNRLQGYIEAVQLEELRAAVTWLERLGKRKRMTFSQEFLTDFRR 120
             ************************************************************

EHDV-1 VP3  VDTIWVLALRLPANPRVIWDVPRCSIANLIMNIATCLPTGEYVSPNPRIASITLTQRITT 180
EHDV-2 VP3  ADTIWVLALRLPANPRVIWDVPRCSIANLIMNIATCLPTGEYVSPNPRIASITLTQRITT 180
EHDV-6 VP3  ADTIWVLALRLPANPRVIWDVPRCSIANLIMNIATCLPTGEYVSPNPRIASITLTQRITT 180
             .************************************************************
```

Figure 4.2. Comparative partial alignment of selected structural proteins [VP1, VP2, and VP3] between EHDV-1 (New Jersey), EHDV-2 (Alberta), and EHDV-6 (Indiana) [isolate CC 304-06]. VP1 is the RNA-dependent RNA polymerase; VP2 is the major surface antigen determining serotype specificity; VP3 is the capsid protein forming the core particle encasing the viral RNA and associated replicative enzymes. Asterisks denote identity; colons and periods represent conserved and semi-conserved amino acid substitutions, respectively.

Figure 4.3. Neighbor-joining phylogenies of VP2, VP5, and VP7 amino acid sequences of currently available EHDV serotypes. Length of the VP2, VP5, and VP7 sequences used in phylogenetic comparisons were 388, 527, and 325 amino acids, respectively. The particular viral protein used to construct each tree is highlighted in each of the EHDV-6 (Indiana) CC 304-06 sequences. Bootstrap confidence values were determined using 500 replicates. Amino acid distances were estimated using the Poisson correction method and are represented as the number of amino acid substitutions per site. Phylogenetic analyses using maximum parsimony methods yielded similar results. Geographical location of each EHDV isolate/strain is indicated. Isolate/strain designations are in parentheses, with GenBank accession numbers listed if isolate/strain designation is not given. GenBank VP2, VP5, and VP7 sequences used for comparisons were: EHDV-2/IBAV Y87061 (BAC20272, BAC20274, BAC20275); EHDV-2/Ibaraki-2 (BAC20268, BAC20270, BAC20271), EHDV-2/IBAV KSB-14/E/97 (BAC20276, BAC20278, BAC20279), EHDV-2 CSIRO 439 (BAC20280, BAC20281, AAA86263), EHDV-2 SV-124 (L33818), EHDV-2 CC 211-91 (AAP86688), EHDV-2 CC 612-D-07 (GQ387640), EHDV-1 CC 124-99 (AAP86721), EHDV-1 New Jersey (Q00273, Q01175, Q00274,), EHDV-6 M44/96 (GQ385940, GQ387639, GQ387642), and EHDV-6 Indiana CC 304-06 (GQ385939, GQ387638, GQ387641).


A
95
100
98
100
EHDV-2 IBAV (Y87061) - Japan
EHDV-2 IBAV (Ibaraki-2) - Japan
EHDV-2 (CSIRO 439) - Australia
EHDV-2 Alberta (SV-124) - Canada
EHDV-2 IBAV (KSB-14/E/97) - Japan
EHDV-1 New Jersey (Q00273) - USA
EHDV-6 Indiana VP2 (CC 304-06) - USA
100
EHDV-6 (M44/96) - South Africa
0.1
B
55
98
99
100
EHDV-2 IBAV (Y87061) - Japan
EHDV-2 IBAV (Ibaraki-2) - Japan
EHDV-2 (CSIRO 439) - Australia
EHDV-2 IBAV (KSB-14/E/97) - Japan
EHDV-2 (CC 612-D-07) - USA
EHDV-6 (M44/96) - South Africa
100
EHDV-6 Indiana VP5 (CC 304-06) - USA
EHDV-1 New Jersey (Q01175) - USA
0.05
C
64
98
EHDV-6 (M44/96) - South Africa
EHDV-6 (AAQ62563) - Reunion Island
EHDV-6 Indiana VP7 (CC 304-06) - USA
88
EHDV-2 Alberta (CC 211-91) - USA
EHDV-2 (CSIRO 439) - Australia
95
EHDV-2 IBAV (Ibaraki-2) - Japan
65
EHDV-2 IBAV (Y87061) - Japan
44
EHDV-2 IBAV (KSB-14/E/97) - Japan
EHDV-1 New Jersey (CC 124-99) - USA
100
EHDV-1 New Jersey (Q00274) - USA
0.005

Acknowledgements

Funds were provided through sponsorship from the fish and wildlife agencies of Alabama, Arkansas, Florida, Georgia, Kansas, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Puerto Rico, South Carolina, Tennessee, Virginia, and West Virginia; through the Federal Aid to Wildlife Restoration Act (50 Stat. 917) and Grant Agreement 06ERAG0005, Biological Resources Division, U.S. Geological Survey, U.S. Department of the Interior; and through Cooperative Agreements 0596130032CA and 0696130032CA, Veterinary Services, Animal and Plant Health Inspection Service, U.S. Department of Agriculture.

References

Abdy, M. J., Howerth, E. W., & Stallknecht, D. E. (1999). Experimental infection of calves with epizootic hemorrhagic disease virus. *Am J Vet Res* **5**, 621-626.

Anthony, S. J., Maan, S., Maan, N., Batten, C., Darpel, K. E., Sutton, G., Attoui, H., & Mertens, P. P. C. (2009). Genetic and phylogenetic analysis of the outer-coat proteins VP2 and VP5 of epizootic haemorrhagic disease virus (EHDV): comparison of genetic and serological data to characterise the EHDV serogroup. *Virus Res* **145**, 200-210.

Barber, T. L., & Jochim, M. M. (1975). Serotyping bluetongue and epizootic hemorrhagic disease virus strains. *Proc Am Assoc Vet Lab Diagn* **18**, 149-157.

Bak, U. B., Cheong, C. K., Choi, H. I., Lee, C.W., & Oh, H. S. (1983). An outbreak of Ibaraki disease in Korea. *Korean J Vet Res* **23**, 81-89.

Beaton, A. R, Rodriguez, J., Reddy, Y. K., & Roy, P. (2002). The membrane trafficking protein calpactin forms a complex with bluetongue virus protein NS3 and mediates virus release. *Proc Natl Acad Sci U S A* **99**, 13154-13159.

Bréard, E., Sailleau, C., Hamblin, C., Graham, S. D., Gourreau, J. M., & Zientara, S. (2004). Outbreak of epizootic haemorrhagic disease on the island of Réunion. *Vet Rec* **155**, 422-423.

Brown, S. E., Gonzalez, H. A., Bodkin, D. K., Tesh, R. B., & Knudson, D. L. (1988). Intra- and inter-serogroup genetic relatedness of orbiviruses. II. Blot hybridization and reassortment *in vitro* of epizootic haemorrhagic disease serogroup, bluetongue type 10 and Pata viruses. *J Gen Virol* **69**, 135-147.

Celma, C. C. P., & Roy, P. (2009). A viral nonstructural protein regulates bluetongue virus trafficking and release. *J Virology* **83**, 6806-6816.

Chalmers, G. A., Vance, H. N., & Mitchell, G. J. (1964). An outbreak of hemorrhagic epizootic disease in wild ungulates in Alberta. *J Wildl Dis* **15**, 1-6.

Collisson, E., & Roy, P. (1983). Analysis of the genomes of bluetongue virus serotype 10 vaccines and a recent BTV-10 isolate from Washington. *Am J Vet Res* **44**, 235-237.

Fletch A. L., & Karstad, L. H. (1971). Studies on the pathogenesis of experimental epizootic hemorrhagic disease of white-tailed deer. *Can J Comp Med* **35**, 224-229.

Gaydos, J. K., Davidson, W. R., Elvinger, F., Howerth, E. W., Murphy M., &. Stallknecht, D. E. (2002). Cross-protection between epizootic hemorrhagic disease virus serotypes 1 and 2 in white-tailed deer. *J Wildl Dis* **38**, 720-728.

Gibbs, E. P., & Lawman, M. J. (1977). Infection of British deer and farm animals with epizootic haemorrhagic disease of deer virus. *J Comp Pathol* **87**, 335-343.

Gloster, J., Mellor, P. S., Burgin, L., Sanders, C., & Carpenter, S. (2007). Will bluetongue come on the wind to the United Kingdom in 2007? *Vet Rec* **160**, 422-426.

Gould, A. R., & Pritchard, L. I. (1991). Phylogenetic analyses of the complete nucleotide sequence of the capsid protein (VP3) of Australian epizootic haemorrhagic disease of deer virus (serotype 2) and cognate genes from other orbiviruses. *Virus Res* **21**, 1-18.

Gumm, I. D., Taylor, W. P., Roach, C. J., Alexander, F. C., Greiner, E. C., & Gibbs, E. P. (1984). Serological survey of ruminants in some Caribbean and South American countries for type-specific antibody to bluetongue and epizootic haemorrhagic disease viruses. *Vet Rec* **114**, 635-638.

Hassan, S. S., & Roy, P. (1999). Expression and functional characterization of bluetongue virus VP2 protein: role in cell entry. *J Virology* **73**, 9832-9842.

Hewat, E. A., Booth, T. F., & Roy, P. (1994). Structure of correctly self-assembled bluetongue virus-like particles. *J Struct Biol* **112**, 183-191.

Hoar, B. R., Carpenter, T. E., Singer, R. S., & Gardner, I. A. (2004). Probability of introduction of exotic strains of bluetongue virus into the US and into California through importation of infected cattle. *Prev Vet Med* **66**, 79-91.

Homan, E. J., Taylor, W. P., De Ruiz, H. L., & Yuill, T. M. (1985). Bluetongue virus and epizootic haemorrhagic disease of deer virus serotypes in northern Colombian cattle. *J Hyg (Lond)* **95**, 165-172.

Iwata, H., Hirasawa, T., & Roy, P. (1991). Complete nucleotide sequence of segment 5 of epizootic haemorrhagic disease virus; the outer capsid protein VP5 is homologous to the VP5 protein of bluetongue virus. *Virus Res* **20**, 273-281.

Iwata, H., Chuma, T., & Roy, P. (1992). Characterization of the genes encoding two of the major capsid proteins of epizootic haemorrhagic disease virus indicates a close genetic relationship to bluetongue virus. *J Gen Virol* **73**, 915-924.

Killington, R. A., Stokes, A., & Hierholzer, J. C. (1996). Virus purification, p. 73-74 *In* Mahy, B. W. J., Kangro, H. O. (ed.), Virology Methods Manual. Academic Press, New York, New York.

Lee, V. H., Causey, O. R., & Moore, D. L. (1974). Bluetongue and related viruses in Ibadan, Nigeria: isolation and preliminary identification of viruses. *Am J Vet Res* **35**, 1105-1108.

Liao, Y. K., Lu, Y. S., Huang, S. T., & Lee, S. H. (1996). The etiological and epidemiological studies on the Ibaraki disease in Taiwan. *Exp Rep TAHRI* **32**, 49-57.

Maia, M. S., & Osburn, B. I. (1992). Differential serological responses to reassortant bluetongue viruses recovered from a bull, p. 776-780 *In* Walton, T. E., Osburn, B. I. (ed.), Bluetongue, African horse sickness, and related Orbiviruses: Proceedings of the Second International Symposium. CRC Press, Boca Raton, Florida.

MacLachlan, N. J., & Osburn, B. I. (2004). Epizootic haemorrhagic disease of deer, p. 1227-1230 *In* Coetzer, J. A. W., Tustin, R. C. (ed.), Infectious diseases of livestock, vol. II, second ed. Oxford University Press Southern Africa, Cape Town, South Africa.

Mellor, P. S., Boorman, J., & Baylis, M. (2000). *Culicoides* biting midges: their role as arbovirus vectors. *Annu Rev Entomol* **45**, 307-340.

Mertens, P. P., Pedley, S., Cowley, J., Burroughs, J. N., Corteyn, A. H., Jeggo, M. H., Jennings, D. M., & Gorman, B. M. (1989). Analysis of the roles of bluetongue virus outer capsid proteins VP2 and VP5 in determination of virus serotype. *Virology* **170**, 561-565.

Mertens, P. P., Burroughs, J. N., Walton, A., Wellby, M. P., Fu, H., O'Hara, R. S., Brookes, S. M., & Mellor, P. S. (1996). Enhanced infectivity of modified bluetongue virus particles for two insect cell lines and for two *Culicoides* vector species. *Virology* **217**, 582-593.

Mertens, P. P. C. (1999). Orbiviruses and Coltiviruses (*Reoviridae*): General Features, p. 1043-1061 *In* Granoff, A., Webster, R. G. (ed.), Encyclopedia of Virology, Second ed. Elsevier Academic Press, San Diego, California.

Mertens, P. P. C., Maan, S., Samuel, A., & Attoui, H. (2005). Genus *Orbivirus*, Family *Reoviridae*, p 466-483 *In* Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., Ball, L. A. (ed.), Virus Taxonomy: Eighth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, California.

Mohammed, M. E., & Mellor, P. S. (1990). Further studies on bluetongue and bluetongue-related orbiviruses in the Sudan. *Epidemiol Infect* **105**, 619-632.

Murphy, M. D., Howerth, E.W., MacLachlan, N.J., & Stallknecht, D.E. (2005). Genetic variation among epizootic hemorrhagic disease viruses in the southeastern United States: 1978-2001. *Infect Genet Evol* **5**, 157-165.

Nason, E. L., Rothagel, R., Mukherjee, S. K., Kar, A. K., Forzan, M., Prasad, B.V., & Roy, P. (2004). Interactions between the inner and outer capsids of bluetongue virus. J *Virology* **78**, 8059-8067.

Ohashi, S, Yoshida, K., Watanabe, Y., & Tsuda, T. (1999). Identification and PCR-restriction fragment length polymorphism analysis of a variant of the Ibaraki virus from naturally infected cattle and aborted fetuses in Japan. *J Clin Microbiol* **37**, 3800-3803.

Omori, T., Inaba, Y., Morimoto, T., Tanaka, Y., Ishitani, R., Kurogi, H., Munakata, K., Matsuda, K., & Matumoto M. (1969). Ibaraki virus, an agent of epizootic disease of cattle resembling bluetongue. I. Epidemiologic, clinical and pathologic observations and experimental transmission to calves. *Jpn J Microbiol* **13**, 139-157.

Parsonson, I. M., & Snowdon, W. A. (1985). Bluetongue, epizootic haemorrhagic disease of deer and related viruses: current situation in Australia. *Prog Clin Biol Res* **178**, 27-35.

Rabenau, H., Ohlinger, V., Anderson, J., Selb, B., Cinatl, J., Wolf, W., Frost, J., Mellor, P., & Doerr, H. W. (1993). Contamination of genetically engineered CHO-cells by epizootic haemorrhagic disease virus (EHDV). *Biologicals* **21**, 207-214.

Samal, S. K., Livingston, C. W., Jr, Mcconnell, S., & Ramig, R. F. (1987a). Analysis of mixed infection of sheep with bluetongue virus serotypes 10 and 17: evidence for genetic reassortment in the vertebrate host. *J Virology* **61**, 1086-1091.

Samal, S. K., El-Hussein, A., Holbrook, F. R., Beaty, B. J., & Ramig, R. F. (1987b). Mixed infection of *Culicoides variipennis* with bluetongue virus serotypes 10 and 17: evidence for high frequency reassortment in the vector. *J Gen Virol* **68**, 2319-2329.

Sellers, R. F., & Maarouf, A. R. (1989). Trajectory analysis and bluetongue virus serotype 2 in Florida 1982. *Can J Vet Res* **53**, 100-102.

Shope, R. E., MacNamara, L. G., & Mangold, R. (1955). Deer mortality – epizootic hemorrhagic disease of deer. *N J Outdoors* **6**, 17-21.

Stallknecht, D. E., Luttrell, M. P., Smith, K. E., & Nettles V. F. (1996). Hemorrhagic disease in white-tailed deer in Texas: a case for enzootic stability. *J Wildl Dis* **32**, 695-700.

Stott, J. L., Oberst, R. D., Channell, M. B., & Osburn, B. I. (1987). Genome segment reassortment between two serotypes of bluetongue virus in a natural host. *J Virology* **61**, 2670-2674.

St. George, T. D., Cybinski, D. H., Standfast, H. A., Gard, G. P., & Della-Porta, A. J. (1983). The isolation of five different viruses of the epizootic haemorrhagic disease of deer serogroup. *Aust Vet J* **60**, 216-217.

Sugiyama, K., Bishop, D. H. L., & Roy, P. (1982). Analyses of the genomes of bluetongue viruses recovered from different states of the United States and at different times. *Am J Epidemiol* **115**, 332-347.

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

Temizel, E. M., Yesilbag, K., Batten, C., Senturk, S., Maan, N. S., Clement-Mertens, P. P, & Batmaz, H. (2009). Epizootic hemorrhagic disease in cattle, Western Turkey. *Emerg Infect Dis* **15**, 317-319.

Uchinuno, Y., Ito, T., Goto, Y., Miura, Y., Ishibashi, K., Itou, T., & Sakai, T. (2003). Differences in Ibaraki virus RNA segment three sequences from 3 epidemics. *J Vet Med Sci* **65**, 1257-1263.

Uren, M. F. (1986). Clinical and pathological responses of sheep and cattle to experimental infection with five different viruses of the epizootic haemorrhagic disease of deer serogroup. *Aust Vet J* **63**, 199-201.

Weir, R. P., Harmsen, M. B., Hunt, N. T., Blacksell, S. D., Lunt, R. A., Pritchard, L. I., Newberry, K. M., Hyatt, A. D., Gould, A. R., & Melville, L. F. (1997). EHDV-1, a new Australian serotype of epizootic haemorrhagic disease virus isolated from sentinel cattle in the Northern Territory. *Vet Microbiol* **58**, 135-143.

Wilson, W. C., O'Hearn, E. S., Tellgren-Roth, C., Stallknecht, D. E., Mead, D. G., & Mecham, J.O. (2009). Detection of all eight serotypes of epizootic hemorrhagic disease virus by real-time reverse transcription polymerase chain reaction. *J Vet Diagn Invest* **21**, 220-225.

Xu, G., Wilson, W., Mecham, J., Murphy, K., Zhou, E. M., & Tabachnick, W. (1997). VP7: an attachment protein of bluetongue virus for cellular receptors in *Culicoides variipennis*. *J Gen Virol* **78**, 1617-1623.

Yadin, H., Brenner, J., Bumbrov, V., Oved, Z., Stram, Y., Klement, E., Perl, S., Anthony, S., Maan, S., Batten, C., & Mertens, P. P. (2008). Epizootic haemorrhagic disease virus type 7 infection in cattle in Israel. *Vet Rec* **162**, 53-56.

CHAPTER 5

OVERPRINTING:

CHARACTERIZATION OF DURHAM VIRUS, A NOVEL RHABDOVIRUS THAT ENCODES A C PROTEIN FROM AN OVERLAPPING OPEN READING FRAME (ORF) AND A SMALL HYDROPHOBIC PROTEIN FROM A NOVEL ORF

[1]Allison, A. B., G. Palacios, A. Travassos da Rosa, V. L. Popov, L. Lu, S. Y. Xiao, K. DeToy, T. Briese, W. Ian Lipkin, M. K. Keel, D. E. Stallknecht, G. R. Bishop, and R. B. Tesh. To be submitted to *Journal of General Virology*.

Abstract

The family *Rhabdoviridae* is a diverse group of non-segmented, negative-sense RNA viruses that are distributed worldwide and infect a wide range of hosts including vertebrates, invertebrates, and plants. Of the 115 currently recognized vertebrate rhabdoviruses, relatively few have been well characterized at both the antigenic and genetic level; hence, the phylogenetic relationships between many of the vertebrate rhabdoviruses remain unknown. The present report describes a novel rhabdovirus, isolated from the brain of a moribund American coot (*Fulica americana*) that exhibited neurological signs when found in Durham County, North Carolina, in 2005. Antigenic characterization of the virus revealed that it was serologically unrelated to 68 other known vertebrate rhabdoviruses. Genomic sequencing of the virus indicated that it shared the highest identity to Tupaia rhabdovirus (TUPV). Additionally, as only previously observed in TUPV, the genome encoded a putative C protein in an overlapping open reading frame (ORF) of the phosphoprotein gene and a small hydrophobic protein located in a novel ORF between the matrix and glycoprotein genes. Phylogenetic analysis of partial amino acid sequences of the nucleoprotein and polymerase proteins indicated that, in addition to TUPV, the virus was most closely related to avian and small mammal rhabdoviruses from Africa and North America. In this report, we present the morphological, pathological, antigenic, and genetic characterization of the new virus, tentatively named Durham virus (DURV), and discuss its potential evolutionary relationship to other vertebrate rhabdoviruses.

Introduction

The *Rhabdoviridae* is one of the four families classified within the Order *Mononegavirales* (Pringle and Easton, 1997). The rhabdovirus genome, like those of paramyxoviruses, filoviruses, and *Borna disease virus* (BDV), is a non-segmented, single strand of negative-sense RNA, although the genome of Orchid fleck virus (OFV) has been proposed to be bipartite (Kondo *et al*., 2006). Six genera are currently recognized within the *Rhabdoviridae*, four of which, *Vesiculovirus*, *Lyssavirus*, *Ephemerovirus*, and *Novirhabdovirus*, infect vertebrates, while members of *Cytorhabdovirus* and *Nucleorhabdovirus* infect plants (Tordo *et al*., 2005). Many of the vertebrate and plant rhabdoviruses replicate in, and are transmitted by, arthropod vectors, although some viruses (e.g., lyssaviruses) are spread by direct contact without any apparent arthropod component.

Of the presently recognized 115 vertebrate rhabdoviruses, 52 are classified as a member (or tentative member) of one of the four existing genera (Tordo *et al*., 2005). Of the 63 remaining unclassified vertebrate rhabdoviruses, 20 are placed within six serogroups based on antigenic cross-reactivity, while 43 remain unassigned to any existing serogroup. Although the antigenic relationships for many vertebrate rhabdoviruses have been determined through serological studies (Tesh *et al*., 1983; Calisher *et al*., 1989), the phylogenetic relationships between many of these viruses remain unknown due to the lack of available sequence data. Additionally, as many of the unclassified vertebrate rhabdoviruses are represented by only a single or few isolates, the normal host associations, transmission cycles, and geographical distributions of these viruses remain obscure.

The prototypical rhabdovirus genome consists of five genes [nucleoprotein (N), phosphoprotein (P), matrix (M), glycoprotein (G), polymerase (L)] that are sequentially transcribed in decreasing molar abundance according to their placement in the genome (i.e., 3'-N>P>M>G>L-5') (Stillman and Whitt, 1998). The genomic 3' and 5' ends contain the untranslated leader and trailer sequences, respectively, which exhibit some degree of terminal complementarity (Wertz *et al*., 1994; Rose and Whitt, 2001). The 3' leader of the genome contains the promoter region needed to initiate transcription of the mRNAs and replication of the positive-sense antigenome, whereas the 3' trailer of the antigenome (the complement of the 5' genomic trailer) acts as the promoter for the replication of genomic RNA (Barr *et al*., 2002). Each of the five individual genes are flanked by transcription initiation and termination/polyadenylation signals, which may be conserved among members of the same genus (Fu, 2005). Between each transcription unit (gene and associated flanking signals) is a nontranscribed intergenic region that usually contains a single or dinucleotide sequence [e.g., G or GG in Tupaia rhabdovirus (TUPV)] (Springfeld *et al*., 2005).

In addition to the five structural proteins (N-P-M-G-L) normally encoded in the genome, a number of additional proteins, either encoded in the same or overlapping open reading frame (ORF) within an existing gene, or in a novel ORF, have been identified in numerous rhabdoviruses. For instance, Wongabel virus (WONV), a putative member of the Hart Park serogroup [which includes the North American avian rhabdoviruses, Flanders virus (FLAV) and Hart Park virus (HPV)] has recently been demonstrated to encode three novel genes (U1, U2, U3) from three consecutive ORFs located between the P and M genes, in addition to two genes (U4, U5) in overlapping ORFs of the N and G

genes, respectively (Bourhy *et al*., 2005; Gubala *et al*., 2008). Other gene products in addition to N-P-M-G-L have also been described for vesiculoviruses (Spiropoulou and Nichol, 1993; Springfeld *et al*., 2005), ephemeroviruses (Walker *et al*., 1992; Wang *et al*., 1994), lyssaviruses (Chenik *et al*., 1995), novirhabdoviruses (Kurath *et al*., 1985; Alonso *et al*., 2004), nucleorhabdoviruses (Scholthof *et al*., 1994; Huang *et al*., 2003), and cytorhabdoviruses (Tanno *et al*., 2000; Dietzgen *et al*., 2006). As the genomic sequences of a number of both vertebrate and plant rhabdoviruses have become available in recent years, it appears that many members of the family do not conform to the prototypical N-P-M-G-L genomic organization. Currently, the function of most of these novel genes remains unknown (Fu, 2005).

In November 2005, a novel rhabdovirus was isolated from the brain of a moribund American coot (*Fulica americana*) found in Durham County, North Carolina. The virus was provisionally named Durham virus (DURV), after the county in North Carolina where the bird was originally recovered. Comparative analysis of the full-length genome of DURV to other rhabdoviruses indicated that it was most closely related to TUPV. Moreover, as previously only documented with TUPV, the genome of DURV encodes a putative C protein in a second overlapping ORF in the P gene, and also a unique small hydrophobic (SH) protein between the M and G genes (Springfeld *et al*., 2005). In this report, we present the morphological, pathological, antigenic, and genetic characterization of DURV and discuss the potential evolutionary relationships of DURV to other vertebrate rhabdoviruses.

Methods

Case history

On 08 November 2005, a moribund American coot (*Fulica americana*) was found by a private wildlife rehabilitator in Bahama, Durham County, North Carolina (NC), and was brought to the Piedmont Wildlife Center, Chapel Hill, NC, on the following day for supportive care. At the time of admittance, the bird was ataxic and was unable to stand and died before a physical exam could be completed. The bird was then shipped to the Southeastern Cooperative Wildlife Disease Study at the University of Georgia for diagnostic evaluation (i.e., necropsy and pathological, virological, bacteriological, and toxicological testing).

Virus isolation and identification

For virus isolation, samples of brain and liver (~0.5cm^3) obtained at necropsy were mechanically homogenized in 650µl of medium [1X minimum essential medium (MEM), 2.2g/l $NaHCO_3$, 20% fetal bovine serum (FBS), and 4X antibiotic/antimycotic solution (400 units/ml penicillin, 400µg/ml streptomycin, 1µg/ml amphotericin B) (Sigma, St. Louis, MO)]. Homogenized tissues were centrifuged (6700 x *g* for 10 min) to pellet debris, and an aliquot (100µl) of clarified supernatant was used to inoculate 2-day-old cultures of Vero cells in a 12-well plate format. Cells were maintained in MEM supplemented with 5% FBS and 2X antibiotic/antimycotic solution and incubated at 37 °C in a humidified atmosphere containing 5% CO_2. For wells exhibiting cytopathology, cell culture supernatant was harvested for (1) generation of stock virus and (2) RNA extraction using a QIAamp® Viral RNA Mini kit (Qiagen, Valencia, CA) according to

the manufacturer's instructions. For initial identification, stock virus was inoculated into three 75cm^2 flasks of Vero cells at an m.o.i. of 0.1 and virus was precipitated on day 4 post-inoculation with polyethylene glycol (PEG) as described by Killington *et al*. (1996). cDNA synthesis from extracted RNA was carried out using random decamers, ImProm-II reverse transcriptase (Promega, Madison, WI), and GoTaq Flexi DNA polymerase (Promega), according to the manufacturer's instructions. Amplicons were electrophoresed on a 2% agarose ethidium bromide-stained TAE gel, excised and purified from agarose using a QIAquick Gel Extraction Kit (Qiagen), cloned using a PCR Cloning^{Plus} Kit (Qiagen), and subsequently purified using a QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. Sequencing of clones was performed using an Applied Biosystems Inc. 3100 Genetic Analyzer (Foster City, California, USA).

Transmission electron microscopy

Infected monolayers of Vero cells were fixed in a mixture of 2.5% formaldehyde and 0.1% glutaraldehyde, containing 0.03% trinitrophenol and 0.03% CaCl2 in 0.05 M cacodylate buffer. Cells were scraped off the plastic, pelleted in buffer, postfixed in 1% OsO4 in 0.1 M cacodylate buffer, en bloc stained with 2% aqueous uranyl acetate, dehydrated in ethanol, and embedded in Poly/Bed 812 (Polysciences, Warrington, PA). Ultrathin sections were cut on a Reichert-Leica Ultracut S ultramicrotome, stained with 0.4% lead citrate, and examined in a Philips 201 electron microscope at 60 kV.

Experimental infection of newborn mice

Two-day old newborn mice (ICR strain, Harlan Sprague-Dawley, Indianapolis, IN) were inoculated intracerebrally with approximately 10^5 PFUs of a stock of DURV prepared from infected Vero cells. When the mice appeared severely ill, they were euthanized with CO_2 gas and a necropsy was performed. Samples of lung, liver, spleen, kidney and brain were removed and placed in 10% buffered formalin for fixation.

Histologic and immunohistochemical examinations

Tissues from the coot and from experimentally-infected mice (lung, liver, spleen, kidney, and brain) were fixed in 10% formalin for 24 hours and then transferred to 70% ethanol for storage and subsequent embedding in paraffin and sectioning. Mouse tissue sections (4-5 µm) and tissue sections from the coot (3 µm) were prepared and stained with hematoxylin and eosin (H&E) or immunohistochemically. Immunohistochemistry (IHC) for DURV antigen was performed as described previously (Xiao *et al*., 2001a; Xiao *et al*., 2001b). DURV mouse hyperimmune ascitic fluid was used as the primary antibody at a dilution of 1:100 and incubated overnight at 4 ^{o}C. An ISO-IHC immunostain kit (Inno-Genex, San Ramon, CA) was used to detect bound primary antibody and to prevent nonspecific binding between species (Xiao *et al*., 2001a; Xiao *et al*., 2001b).

Serologic testing

Complement-fixation (CF) tests were done according to a microtechnique described previously (Tesh *et al*., 1983), using 2 full units of guinea-pig complement.

Titers were recorded as the highest dilutions giving 3+ or 4+ fixation of complement on a scale of 0 to 4+.

Antigens and immune reagents

Antigens used in CF tests and for immunizing animals were infected newborn mouse brains, prepared by the sucrose-acetone extraction method (Clarke and Casals, 1958). Specific hyperimmune mouse ascitic fluids were prepared against each of the 68 vertebrate rhabdoviruses listed in Supplementary Table 1. The immunization schedule consisted of four intraperitoneal injections given at weekly intervals. Immunogens consisted of 10% suspensions of homogenized infected mouse brain mixed with equal volumes of Freund's adjuvant just prior to inoculation. Sarcoma 180 cells were also given intraperitoneally with the final immunization in order to induce ascites formation. Five adult hamsters were also inoculated with DURV antigen prepared from a frozen harvest of infected baby hamster kidney (BHK) cells grown in MEM with 10% hamster serum. The immunization schedule was the same, except that sarcoma 180 cells were not used. All animal work was carried out under an animal protocol approved by the University of Texas Medical Branch.

In vitro host range and growth kinetics

The host range and replicative capacity of DURV was assessed in the following mammalian, avian, reptilian, fish, and mosquito cell lines: (1) Vero [African green monkey (*Cercopithecus aethiops*) kidney epithelium], (2) CPAE [cattle (*Bos taurus*) pulmonary artery endothelium], (3) QNR/K2 [Japanese quail (*Coturnix coturnix*

japonica) Müller cells], (4) PDE [Pekin duck (*Anas platyrhynchus domesticus*) embryo fibroblasts], (5) VH-2 [Russell's viper (*Vipera russelli*) heart fibroblasts], (6) TH-1 [Eastern box turtle (*Terrapene carolina*) heart fibroblasts], (7) FHM [Fathead minnow (*Pimephales promelas*) connective tissue and muscle epithelium], and (8) C6/36 [Asian tiger mosquito (*Aedes albopictus*) larvae]. Mammalian and avian cell lines were maintained at 37 °C, while reptilian, fish, and mosquito cells were maintained at 28 °C. Cells were infected with approximately 10^3 PFUs of DURV in a 12-well (2ml) plate format and wells were harvested daily for 8 days. All titrations were performed on Vero cells overlaid with 1% gum tragacanth/1X MEM supplemented with 3% FBS and 2X antibiotic/antimycotic solution. Cultures were inactivated on day 3-7 post-adsorption (depending upon the virus; see below) with 10% buffered formalin and stained with 0.25% crystal violet for plaque visualization. Dilutions in which 20-100 plaques could be counted (when applicable) were used in determining viral titers ($\log_{10}$ PFU/mL). Four additional rhabdoviruses, 1) Farmington virus (FARV) [CT AN 114 Clone B, isolated from an unidentified bird species in Connecticut in 1969], 2) FLAV [WV 382-02, isolated from an House sparrow (*Passer domesticus*) from West Virginia in 2002], 3) Klamath virus (KLAV) [M-1056, isolated from a montane vole (*Microtus montanus*) in Oregon in 1962], and 4) *Vesicular stomatitis Indiana virus* (VSINV) [97-25323, isolated from a horse (*Equus caballus*) in New Mexico in 1997] were also analyzed for comparison.

Genomic sequencing

Preparation of genomic material for pyrosequencing was performed as described previously (Palacios *et al*., 2008). DURV RNA was extracted from infected Vero cell supernatant using TRIzol LS (Invitrogen, Carlsbad, CA). Total RNA extracts were treated with DNase I (DNA-*free* kit, Ambion, Austin, TX) and cDNA was subjected to a modified degenerate oligonucleotide-primed PCR (DOP-PCR) procedure (Palacios *et al*., 2007). Products >70 base pairs (bp) were selected by column purification (MinElute, Qiagen) and ligated to specific linkers for sequencing on the 454 Genome Sequencer FLX (454 Life Sciences, Branford, CT) without fragmentation of the cDNA (Palacios *et al*., 2008). Removal of primer sequences, redundancy filtering, and sequence assembly were performed with software programs accessible through the analysis applications at the GreenePortal website (http://tako.cpmc.columbia.edu/Tools/).

The libraries yielded 111,372 sequence reads, of which, 6,727 were retrospectively identified as belonging to DURV. Alignment of unique singleton and assembled contiguous sequences to the National Center for Biotechnology Information (NCBI) GenBank database using BLASTn and BLASTx analysis (Altschul *et al*., 1990) indicated coverage of approximately 8.1 kilobases (kB) of sequence distributed along rhabdovirus genome scaffolds. Sequence gaps between the aligned fragments were filled in by specific RT-PCR amplification with primers designed on the pyrosequence data. Conventional RT-PCRs were performed using a HotStar HiFidelity polymerase kit (Qiagen) according to the manufacturer's instructions. Specific primer sequences are available upon request. Amplification products were electrophoresed on 1% agarose gels, purified from agarose (MinElute, Qiagen), and sequenced using ABI PRISM Big

Dye Terminator v1.1 Cycle Sequencing kits on a ABI PRISM 3700 DNA Analyzer (Applied Biosystems Inc.). Terminal sequences were obtained by RACE (SMART RACE cDNA Amplification Kit, Clontech, Mountain View, CA). Overlapping primer sets based on the draft genome were designed to facilitate sequence validation by classical dideoxy sequencing. The accumulated data revealed the complete DURV genome (GenBank accession number FJ952155).

Protein analysis

All molecular visualizations of protein structures were carried out using the Visual Molecular Dynamics (VMD) program (http://www.ks.uiuc.edu/Research/vmd) running in an OpenGL 32-bit Windows XP Professional format (Humphrey *et al*., 1996). The crystal structure of the low pH form of the VSINV G protein (2CMZ) was obtained from the NCBI Protein Data Bank (PDB). DURV C and SH protein identity and homology to known sequences was assessed using BLASTp against both the non-redundant protein database and the PDB structure database. DURV protein pairwise identities to representative rhabdoviruses were determined by CLUSTALW alignment using the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) server (http://www.ebi.ac.uk/Tools/clustalw2/index.html) (Chenna *et al*., 2003). The predicted molecular weights, isoelectric points, and grand average hydrophobicity scores of the DURV proteins were determined using the ProtParam tool on the ExPASy server (http://www.expasy.ch/tools/protparam.html) (Kyte and Doolittle, 1982; Gasteiger *et al*., 2003). Phosphorylation sites of the P protein were predicted using the NetPhos 2.0 server (http://www.cbs.dtu.dk/services/NetPhos/) (Blom *et al*., 1999).

The putative transmembrane topology of the SH protein was determined using Phobius (http://phobius.sbc.su.se/), SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/cgi-bin/adv_sosui.cgi), and TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/) servers. Hydrophobicity plot analysis of the SH protein was performed using the ProtScale program on the ExPASy server (http://www.expasy.ch/tools/protscale.html). Glycosylation sites of the DURV G protein were predicted using the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/).

Phylogenetic analysis

Evolutionary relationships of DURV with representative rhabdoviruses were derived by construction of phylogenetic trees generated with CLUSTALW alignments of partial amino acid (aa) sequences of the N, G, and L proteins using the neighbor-joining method in the Molecular Evolutionary Genetics Analysis (MEGA) version 4 program (Tamura *et al*., 2007). Gaps in the alignments were analyzed by both pairwise and complete deletion. Bootstrap values were determined using 2000 replicates. The trees were calculated using Poisson correction and evolutionary distances were represented as the number of amino acid substitutions per site. N, G, and L sequences of 212, 334, and 156 aa in length, respectively, were used in generating phylogenies with the cognate regions of other rhabdovirus sequences available in GenBank (Bourhy *et al*., 2005; Kuzmin *et al*., 2006). For the L phylogeny, new sequences were obtained for six additional rhabdoviruses endemic to the United States: (1) KLAV M-1056, (2) Malpais Spring virus (MSPV) [85-488NM, isolated from *Aedes campestris* in New Mexico in 1985], (3) Bahia Grande virus (BGV) [TB4-1054, isolated from *Aedes sollicitans* in

Texas in 1974], (4) Muir Spring virus (MSV) [76V-23524, isolated from *Aedes* spp. in Colorado in 1976], (5) Bivens Arm virus (BAV) [BT82-55, isolated from *Culicoides insignis* in Florida in 1982], and (6) Kern Canyon virus (KCV) [M-206, isolated from a Yuma myotis (*Myotis yumanensis*) in California in 1956]. L sequences were amplified using primers developed to a conserved region in block IV as previously described (Bourhy *et al*., 2005).

Results

Virus isolation and preliminary genetic characterization

Cytopathic effects were observed in Vero cells inoculated with brain homogenate from the American coot and a virus (DURV) was subsequently isolated. Although diethyl ether and 5-bromo-2'-deoxyuridine treatment suggested the virus was an enveloped RNA virus, RT-PCR and plaque reduction neutralization testing against selected endemic bird-associated RNA viruses known to be cytopathic in mammalian cell culture yielded negative results. Subsequent RT-PCR, using random decamers, of RNA extracted from PEG-precipitated virus yielded a number of clones. One clone (640 bp), when translated, exhibited identity to a single sequence by BLASTp analysis: TUPV G protein (NCBI accession AAX47601). Primer walking from this clone resulted in a 2,850bp (950 aa) contig covering portions of the G and L genes that, similar to the initial clone, shared the highest amino acid identity (45%) to TUPV.

Pathological examination of bird tissues

A postmortem examination was performed on the carcass of the American coot. The bird was in fair nutritional condition and there was only mild autolysis. A small amount of hemorrhage was present in the muscle and subcutaneous tissues adjacent to the right caudolateral aspect of the sternum. No other significant gross lesions were identified. Microscopic examination of H&E-stained sections of the brain showed features of both meningitis and encephalitis. Specifically, there was vascular congestion involving both the meninges (Fig. 5.1A) and the brain parenchyma. In addition, inflammatory cellular infiltration was evident in the meninges, particularly in perivascular areas. Neuronal degeneration was seen in most areas of the parenchyma, with vacuolation and inflammation (Fig. 5.1B). IHC of brain tissue showed perinuclear staining with viral antigen in some neurons (Fig. 5.1C). The other tissues examined did not have any significant lesions.

Ultrastructural studies

In ultrathin sections of infected Vero cells, areas of massive virion formation could be observed in their cytoplasm (Fig. 5.2A). Rod-like virions ~30 nm in diameter and 140-160 nm long were budding from the limiting membranes into the expanded membrane-limited compartments formed by rough endoplasmic reticulum (Fig. 5.2B). Infected cells did not show significant cytopathy.

Antigenic relationships

In CF tests, the hyperimmune DURV mouse and hamster antisera were positive with the homologous (DURV) mouse brain antigen at dilutions of 512/≥2 and ≥2048/≥2, respectively. The DURV mouse brain antigen was negative in CF tests with antisera against each of the 68 rhabdoviruses listed in Supplementary Table 5.1, as well as with an antiserum against *Newcastle disease virus* (NDV) (*Avulavirus*: *Paramyxoviridae*).

Pathogenicity of DURV in mice and hamsters

Two-day old ICR mice inoculated intracerebrally, with approximately 10^4 tissue culture infectious $dose_{50}$ of DURV, became ill approximately the sixth day and most were dead by the ninth day after inoculation. Intraperitoneal inoculation of adult mice and hamsters with a crude 10% suspension of brains from the moribund and dead infant mice did not produce detectable illness, but both species had high levels of CF antibodies when tested approximately one month after infection.

Histopathology in newborn mice

Examination of stained sections of the brain showed that the overall pathologic process was that of multifocal necrosis of different stages. There was no specific preferential distribution, and the lesions were noted in the cerebral cortex, subcortical nuclei, hippocampus, and cerebellum. Earlier stage lesions showed degeneration or necrosis of neurons, followed by neutrophilic infiltration in later stage lesions, and eventually loss of neuropil, leading to the vacuoles, containing heavy mixed inflammatory cellular infiltration, but mainly neutrophils. In addition, there were

multiple foci of vascular reaction, characterized by prominence of endothelial cells, with intravascular and perivascular lymphocytic infiltration. Representative photomicrographs are shown in Fig. 5.1D-F. Examination of liver, spleen, and kidneys, did not reveal significant abnormalities.

In vitro growth characteristics

In vitro growth curve analysis demonstrated that quail glial cells of the neuroretina (QNR/K2) and monkey kidney (Vero), cattle (CPAE), and bat (Tb 1 Lu) cells all supported replication of DURV, but duck embryo (PDE), fish (FHM), mosquito (C6/36), and reptilian (VH2, TH-1) cell lines were either refractory or relatively non-permissive to infection (Table 5.1). DURV replicated to lower titers in permissive cell lines (Vero, CPAE, Tb 1 Lu, QNR/K2) than either VSINV, FARV, or KLAV; however, the extent to which the titer of the initial inoculum and passage history of each virus affects its ability to replicate on different cell lines is unknown, and therefore, direct comparisons between viruses of different passage histories may be biased. In contrast, DURV replicated much more efficiently than FLAV [same passage history (Vero P2) as DURV], with FLAV being severely restricted in both its host cell range and replicative capacity (Table 5.1), as previously noted by Whitney (1964).

Genomic and protein analysis

Genomic sequencing of DURV revealed seven ORFs, which based on standard rhabdovirus nomenclature, encode the following seven proteins: 1) nucleoprotein (N), 2) phosphoprotein (P), 3) C protein, 4) matrix (M) protein, 5) small hydrophobic (SH)

protein, 6) glycoprotein (G), and 7) polymerase (L) protein. The entire genome of DURV was 11,265 nt long, encoding 3,784 aa, and is schematically represented as 3-*l*-N-P/C-M-SH-G-L-*t*-5 (Fig. 5.3A). The length of each ORF and gene, and the deduced amino acid length, molecular weight, isoelectric point, and hydrophobicity index of each protein is listed in Table 5.2. The 3' leader was 84 nt long, while the 5' trailer sequence was significantly shorter, being 29 nt in length. The first three (UGC) and 10^{th} (U) nucleotides of the 3' leader sequence that are conserved among the vertebrate rhabdoviruses were also present in DURV (Fig. 5.3C) (Springfeld *et al*., 2005). The putative transcription start and stop/polyadenylation sequences were KUKY and NBACUUUUUUU ($NBACU_7$), respectively. The stop codon for each of the P, SH, and G genes was contained within the putative polyadenylation sequence (Fig. 5.3B). The deduced intergenic region between each transcription unit was GA, similar to *Vesicular stomatitis New Jersey virus* (VSNJV) (Stillman and Whitt, 1998). All transcription initiation, intergenic, and transcription termination/polyadenylation sequences, along with the 3' leader and 5' trailer sequences, are shown in Fig. 5.3B-C.

The DURV N gene was 1,290 nt long, coding for 430 aa, which was identical in size to the N protein of TUPV. Pairwise comparisons with the prototype species of each of the four vertebrate genera in *Rhabdoviridae* [i.e., RABV, VSINV, *bovine ephemeral fever virus* (BEFV), *infectious hematopoietic necrosis virus* (IHNV)], along with other selected rhabdoviruses, indicated that the DURV N protein shared the highest aa identity to TUPV (57%), followed by FLAV and VSINV (32%) (Table 5.3). Kolongo virus (KOLV), an African bird-associated rhabdovirus, shared the second closest amino acid identity (34%) (not shown). The region of the N protein reported to be the RNA binding

motif conserved among the rhabdoviruses (Kouznetzoff *et al*., 1998), was ^{285}-GISKNSPYSS-^{294} in DURV, with N289 being unique.

The DURV P gene was 1,011 nt long, coding for 336 aa, which was longer than the P protein of the prototype viruses of the four vertebrate genera and most similar in size to the TUPV P protein (337 aa). Twenty-five phosphorylation sites (16 serine; 8 threonine; 1 tyrosine) were predicted. The DURV P protein shared the lowest identity of all of the five major proteins when compared against other rhabdoviruses, with a maximum identity of 16% to TUPV (Table 5.3). Although the DURV and TUPV P proteins shared the highest identity and were the closest in size among the rhabdovirus P proteins analyzed, only three stretches of consecutive identities of at least three aa (^{15}-YDL-^{17}, ^{67}-FSR-^{69}, and ^{326}-LGL-^{328}) were identified between the two viruses. A putative dynein light chain binding motif [(K/R)XTQT] (Jacob *et al*., 2000; Lo *et al*., 2001) could not be identified in the DURV P gene.

Similar to that first described in VSNJV (Spiropoulou and Nichol, 1993), the DURV P gene encodes an additional protein, designated as the C protein, from a second overlapping ORF located near the 5' end of the P gene. The putative start codon for the C protein lies 40 nt downstream of the initiator AUG codon for the P gene. Position -3 from the start codon of the P gene is a C, consistent with Kozak's rules for leaky scanning through an upstream AUG codon (Kozak, 1989). The predicted DURV C protein was 136 aa in length, 85 aa shorter than the TUPV homolog (221 aa), but significantly larger than the C' proteins reported for VSINV (67 aa) and VSVNJV (65 aa) or the putative C proteins reported for *Cocal virus* (COCV) ($\geq$93aa), *Chandipura virus* (CHPV) (80 aa), *Siniperca chuatsi* rhabdovirus (SCRV) (72 aa), *Isfahan virus* (ISFV) (47

aa), *viral hemorrhagic septicemia virus* (VHSV) (46 aa), IHNV (42 aa), and *Piry virus* (PIRYV) (32aa) (Spiropoulou and Nichol, 1993; Schütze *et al*., 1999; Marriott, 2005; Pauszek *et al*., 2008; Tao *et al*., 2008). BLASTp analysis of the C protein of DURV revealed that it shared homologies, covering very similar regions of the DURV C protein (encompassing aa positions 15-75) (Fig. 5.3E), to two sequences of interest. Using the non-redundant sequence database, DURV C residues 16-75 shared a 22-23% identity and a 43-45% homology to the dynein heavy chain proteins of two mosquito species found in the southeastern United States, *Culex quinquefasciatus* (NCBI accession CPIJ002912) and *Aedes aegypti* (NCBI accession AAEL014313). Additionally, BLASTp analysis using the PDB demonstrated that the DURV C protein (aa positions 15-71) scored the highest (i.e., 26% identity and 36% homology) to the fusion domain of the G protein of VSINV (aa positions 55-103; PDB ID 2CMZ) (Fig. 5.4). No other viral sequences aligned with DURV C in the PDB BLAST. To determine an approximate number of viral structures in PDB databank, a query of "virus" revealed 3,011 structural hits (http://www.rcsb.org/pdb.).

The M protein of DURV was 606 nt long, coding for 193 aa, which was comparatively shorter in relation to most other rhabdoviruses. Similar to that observed with the P protein, the DURV M protein exhibited little sequence homology to other rhabdoviruses, with a maximum aa identity of 20% to TUPV (Table 5.3). The DURV M protein did not contain the highly conserved PPXY motif involved in viral budding reported for other rhabdoviruses (Harty *et al*., 1999).

As recognized previously only in TUPV, the DURV genome contains an additional ORF of 231 nt between the M and G genes that encodes a putative protein of

77 aa (Springfeld *et al*., 2005). Based on the nomenclature of Springfeld *et al*. (2005), we have also denoted this gene product as the small, hydrophobic (SH) protein. In comparison to TUPV, the DURV SH protein exhibited a 25% aa identity and was truncated in length, being 16 aa shorter. The DURV SH protein was extremely leucine rich, primarily in the N-terminal region, with leucine residues accounting for 20.8% (16/77) of the aa composition of the entire protein. Leucine residues 1, 5, 7, 8 and 9 (numbering based on DURV) were conserved between DURV and TUPV (Fig. 5.3D). Overall, the N-terminal portion of the protein exhibited a higher degree of conservation between DURV and TUPV, with the first 28 aa exhibiting a 50% identity. Hydrophobicity plot analysis, similar to that of TUPV (Springfeld *et al*., 2005), indicated that the N-terminal two-thirds of the protein contained regions of high hydrophobicity, while the C-terminal region (aa positions 49-77) was hydrophilic (not shown). The putative transmembrane topology of DURV SH was assessed using the Phobius program as was previously done for TUPV (Springfeld *et al*., 2005). The DURV SH protein was predicted to contain two transmembrane helices spanning residues 5-26 and 32-48, separated by a short extracellular domain (aa positions 27-31). Residues 1-4 and 49-77 were predicted to be cytoplasmic, with no putative signal sequence associated with the N-terminus. Determination of potential transmembrane domains using additional programs (SOSUI, TMHMM) yielded similar predictions regarding the location of the two transmembrane domains and the lack of a predicted signal peptide. Thus, the DURV SH protein appears to be a multipass transmembrane protein without a clear N-terminal signal sequence, potentially suggesting that an internal start-transfer sequence initiates translocation. The C-terminal region of the DURV SH protein stretching from residues

50-75 shows homology to the α1 and α2 regions of the sterile alpha motif (SAM) domains of the Eph family of receptor tyrosine kinases by BLASTp analysis of the PDB (Stapleton *et al*., 1999). Specifically, residues 57-63 (RLTGDWL) exhibited a high identity to a portion of the α1 motif in the SAM domain of EphA3 (RTTGDWL) (PDB ID 3FYD2).

The DURV G protein was 1,518 nt long, encoding 506 aa, being slightly shorter than most rhabdoviruses. Topological analysis using the Phobius server predicted an N-terminal signal peptide (^{1}-MWIILLHVSFVASQVII-^{17}), followed by an ectodomain (aa 18-475), transmembrane domain (^{476}-ILLASIITLIALITSTLLLCCVC-^{498}), and a short cytoplasmic tail (^{499}-KKRQHRSV-^{506}), consistent with the overall topology of other rhabdovirus G proteins (Coll, 1995; Walker and Kongsuwan, 1999). To deduce the cysteine configuration of the G protein, the G aa sequence of DURV was aligned against selected rhabdoviruses: TUPV, FLAV, VSINV, and Spring viremia of carp virus (SVCV). Both FLAV and SVCV exhibited the 12 cysteine residues previously reported for VSINV and other vesiculoviruses (Walker and Kongsuwan, 1999; Hoffman *et al*., 2002). DURV, as reported for TUPV (Springfeld *et al*., 2005), contained only 10 cysteine residues in the G protein and was missing residues C_{VIII} and $C_{X.}$ No cysteines were detected between C_{VII}-C_{IX} or C_{IX}-C_{XI}. Additionally, DURV G had two potential glycosylation sites (NXS/T) at aa positions ^{317}-NST-^{320} and ^{404}-NNT-^{407}. Similar to VSINV and other vesiculoviruses, the DURV G protein contains the aa residues ^{82}-WY-^{83} and ^{126}-YA-^{127} in the proposed two non-contiguous fusion loops that are responsible for G-mediated fusion (Roche *et al*., 2006; Roche *et al*., 2007; Sun *et al*., 2008). Unlike

DURV, TUPV contained the putative bipartite fusion loop sequences ^{90}-WL-^{91} and ^{134}-WS-^{135}.

The DURV L protein was 6,318 nt long, encoding 2,105 aa. The N-terminal LNSPL motif found in all animal rhabdoviruses, other than TUPV and *snakehead rhabdovirus* (SHRV) (Kuzmin *et al*., 2008), is also different in DURV, with an aspartic acid substitution for asparagine (^{41}-L<u>D</u>SPL-^{45}). Block I in DURV, as in all mammalian rhabdoviruses, contains the tripeptide ^{353}-GHP-^{355}, and, as in the lyssaviruses, contains a putative leucine zipper motif (^{219}-LX_6LX_6LX_6L-^{240} or ^{219}-LX_6LX_6LX_5L-^{240}) (Kuzmin *et al*., 2008). Other conserved polymerase motifs, including the block III ^{672}-GGLEG-^{676} and ^{706}-QGDNQ-^{710} pentapeptides, block IV ^{879}-GG-^{880} and ^{896}-DP-^{897} dipeptides, and block VI ^{1665}-GXGXGG-^{1670} 2'-O-ribose-methyltransferase motif, are also present in DURV (Poch *et al*., 1990; Kuzmin *et al*., 2008). Although the start of the polyadenylation sequence directly follows (N, M) or precedes (P, SH, G) the stop codon for other DURV genes, the first stop codon in the L gene was 68 nt upstream of the start of the putative polyadenylation site (NBACU_7). However, a second in-frame stop codon (AUC) starts two nt upstream of the start of the polyadenylation signal (Fig. 5.3B), which may suggest that the first stop codon encountered by the polymerase complex could possibly be read-through, culminating in an L protein containing an additional 22 aa.

Phylogenetic analysis

Neighbor-joining phylogenies of the N, G, and L proteins of DURV were constructed against the cognate regions of selected representative rhabdoviruses. Phylogenetic analysis of the partial N aa sequence of DURV indicated that it was most

closely related to TUPV and, as previously noted for TUPV by Kuzmin *et al*. (2006), both viruses clustered with KOLV and Sandjimba (SJAV) viruses, two rhabdoviruses isolated from birds from the Central African Republic (Fig. 5.5A). The G protein phylogeny demonstrated that DURV and TUPV formed a monophyletic clade completely separate from all other vertebrate rhabdoviruses (not shown). Phylogenetic analysis of a portion of the L protein indicated that, similar to the N and G phylogenies, DURV was most closely related to TUPV. Additionally, both DURV and TUPV clustered with KLAV to form a distinct group (89% bootstrap support), which was separate from other rhabdoviruses (Fig. 5.5B). The partial L sequence (136 aa) of KLAV exhibited a 76 and 80% identity to DURV and TUPV, respectively (not shown). China fish rhabdovirus (CFRV), an unclassified virus which shows the highest L aa identity to the fish vesiculoviruses (tentative classification), was joined to DURV in the L tree, albeit with minimal support (Fig. 5.5B).

In order to potentially elaborate on the evolutionary relationships of vertebrate rhabdoviruses isolated in the United States, partial L sequences were phylogenetically analyzed for an additional five rhabdoviruses: BAV, MSPV, MSV, BGV, and KCV (Fig. 5.5B). The L phylogeny demonstrated that BAV clustered tightly with Tibrogargan virus (TIBV) with 100% bootstrap support. Alignment of the partial L nt and aa sequences between TIBV and BAV demonstrated a 93 and 98% identity, respectively (not shown), confirming the close serological relationships previously noted between the two viruses (Gibbs *et al*., 1989; Calisher *et al*., 1989) and potentially suggesting BAV may be a strain of TIBV. MPSV grouped with CHPV, ISFV, and PERV in the L tree, verifying its previous antigenic designation as a tentative member of the *Vesiculovirus* genus (Clark *et*

al., 1988). KCV grouped with high support to LeDantec virus (LDV), the latter originally isolated from a human patient in Senegal in 1965 (Cropp *et al*., 1985). KCV and LDV further clustered with Fukouka virus (FUKV), affirming previous antigenic relationships recognized between KCV and FUKV (Calisher *et al*., 1989) and suggesting that, as alluded to by Kuzmin *et al.* (2009), members of the Kern Canyon and Le Dantec serogroups are phylogenetically related. BGV and MSV shared an 81 and 92% nt and aa identity, respectively, in their partial L sequences and formed a monophyletic clade separate from other rhabdoviruses, reiterating their classification as members of a distinct (Bahia Grande) serogroup (Kerschner *et al*., 1986; Calisher *et al*., 1989).

Discussion

Genomic sequencing of DURV revealed it shared the highest amino acid identity to, and an identical genomic organization with, TUPV, a virus originally isolated from a Northern tree shrew (*Tupaia belangeri*) imported from Thailand (Kurz *et al*., 1986). As tree shrews are geographically confined to Southeast Asia (Nowak, 1999), the recognition of a rhabdovirus isolated from a bird in the southeastern United States that was most closely related to TUPV was an intriguing finding and raised questions concerning the evolutionary history of these two viruses. Although DURV and TUPV are most closely related to one another based on currently available rhabdovirus sequences, they only share a 42% amino acid identity over their entire genomes, suggesting they are either distantly related and/or have undergone considerable genetic divergence due to contrasting transmission cycles and geographical ranges. Additionally, unlike that reported for TUPV (Kurz *et al*., 1986), DURV was not host-restricted either *in*

vitro (Table 5.1) or *in vivo* (Fig. 5.1). Nevertheless, DURV and TUPV share a number of common and unique genetic features (as outlined below) which may facilitate inferring evolutionary histories of the *Rhabdoviridae* as a whole, when additional sequence of other vertebrate rhabdoviruses (e.g., bird or small mammal-associated viruses) becomes available.

Like TUPV and most members of the genus *Vesiculovirus,* the DURV genome encodes a putative C protein from a second overlapping reading frame in the P gene (Spiropoulou and Nichol, 1993; Kretzschmar *et al*., 1996; Springfeld *et al*., 2005; Pauszek *et al*., 2008). The DURV C protein was not conserved among related viruses, as it exhibited an 8 and 14% amino acid identity to the C proteins of TUPV and VSINV, respectively (Table 5.3). Currently, the precise function of the C protein of the vesiculoviruses remains unknown, although it has been speculated to enhance transcriptional activity or be involved in host pathogenicity or insect transmission (Kretzschmar *et al*., 1996; Peluso *et al*., 1996). For VSNJV and VSINV, the C protein has been shown to be a small, highly basic, nonstructural protein that is translated in major (C) and minor (C') forms (Spiropoulou and Nichol, 1991; Kretzschmar *et al*., 1996). Previously, Kretzschmar *et al*. (1996) demonstrated that VSNJV mutants, incapable of expressing C proteins, produced infectious progeny and synthesized RNA and protein to levels indistinguishable from wild-type virus. Although this result suggested that the C proteins of VSNJV were dispensable for viral replication in the system employed, the authors did additionally speculate that the C proteins may possibly be important for replication in the natural host or vector. Similarly, Peluso *et al*. (1996) could not demonstrate an effect of purified or recombinant C proteins of VSINV on

transcription when using infected cell extracts or reconstituted polymerase complexes, respectively; however, it was additionally shown that, when added to purified virions, recombinant C' protein resulted in an increase in RNA synthesis in a serotype-specific manner. It was further proposed that one mechanism by which the VSINV C' protein may stimulate transcription is through interacting with and stabilizing the polymerase complex as it is bound to the RNA (Peluso *et al*., 1996).

Protein BLAST analysis of the C protein of DURV, using the NCBI non-redundant sequence database, indicated that it shared identity and homology to the ciliary dynein heavy chain proteins of *Culex* and *Aedes* species of mosquitoes. Whether the DURV C protein additionally shows homology to dynein proteins of other vector species known to be involved in rhabdovirus transmission (e.g., biting midges, sand flies) is unknown, as these sequences are currently not available. It has been demonstrated that the P protein of RABV has a dynein binding motif and that the interaction of the P protein with dynein, rather than promoting retrograde axonal transport as previously suggested (Raux *et al*., 2000), enhances viral transcription (Tan *et al*., 2007). By analogy, this may suggest that the C protein, at least in DURV, could potentially act as a competitive inhibitor or functional analog of dynein by binding to and interacting with the P protein during viral infection (e.g., in the vector, presuming DURV is arthropod-borne), thereby modulating transcriptional activity of the polymerase complex.

Additionally, BLAST analysis of the DURV C protein against the NCBI PDB database revealed it had the highest alignment score with the G protein of VSINV (PBD ID 2CMZ), with the area of homology mapping to the fusion domain in the G protein (Fig. 5.4). That the DURV C protein shared the greatest identity/homology to the

VSINV G protein of all the structural sequences in the PDB (~56,000 entries) was intriguing and possibly suggested that this homology represented a functional commonality, such that the DURV C protein may be fusogenic and, therefore, may play a role in mediating cell-cell fusion or virus release *in vivo*. Alternatively, another hypothesis is that the DURV C protein could potentially represent a more ancestral progenitor sequence from which the G protein evolved. Previously, nonstructural glycoprotein (G_{NS}) genes, located immediately downstream of the G gene, have been identified in *Adelaide River virus* (ARV) and BEFV (Walker *et al*., 1992; Wang and Walker, 1993). Although the G_{NS} proteins of these ephemeroviruses may have been acquired via non-homologous recombination with another rhabdovirus, it is generally perceived that they originated by homologous gene duplication of the G gene by a copy-choice mechanism (Walker *et al*., 1992; Wang and Walker, 1993). G and G_{NS} proteins of both ARV and BEFV show low levels of amino acid identity (e.g., 8 and 14% pairwise identity between NCBI accessions numbers AAA02763/AA02764 and AAG10413 /AAG10414), suggesting that, if the G_{NS} gene was derived by duplication of the G gene, it has undergone substantial genetic divergence, potentially adapting and evolving to a new biological role in the viral life cycle independent of the function of the ancestral G protein. As demonstrated for BEFV, the lack of G_{NS} incorporation into virions, its antigenic unrelatedness to the G protein, and its inability to induce neutralizing antibodies in the host implies that this may be the case (Hertig *et al*., 1995). Gene duplication followed by functional divergence has also been also hypothesized for White spot syndrome virus (WSSV), in which the nucleocapsid and envelope proteins (i.e., proteins carrying out substantially different biological roles) have been proposed to have arisen by

duplication of an ancestral gene (van Hulten *et al*., 2000). Although the evidence is limited, this could suggest that the ancestral G protein initially arose through an inter- or intra-molecular copy-choice mechanism of the C gene, with the duplicate C gene (possibly with fusogenic activity) subsequently adapting and specializing into a novel viral attachment/fusion protein (G protein). If the ancestral rhabdovirus C gene was bi-functional (e.g., transcription modulation, cell-cell fusion), similar to the multi-functional C proteins of paramyxoviruses (Lamb and Parks, 2007), this may additionally support the subfunctionalization of the duplicate gene to carry out a more specialized function (Lynch and Force, 1999; Hughes, 2002).

C protein sequences in vesiculoviruses have been demonstrated to be less conserved than their P counterparts and that negative selection is acting to conserve the region of overlap between the P and C genes, but not the C gene sequence (Jordan *et al*., 2000). As the C gene in TUPV and DURV encode putative polypeptides of 221 and 136 amino acids, respectively, this may suggest that the C' proteins in the vesiculoviruses are evolving to become non-essential, as the C' proteins are substantially shorter in VSINV (67 amino acids) and VSNJV (65 amino acids), or completely absent as in the case with *Vesicular stomatitis Alagoas virus* (VSAV) (Pauszek *et al*., 2008). However, the evolution of the C gene and its biological role during viral infection are only speculative until additional DURV isolates, or related rhabdoviruses containing bicistronic P genes, are sequenced, in addition to performing functional studies to investigate the biological properties of the DURV C protein.

Similar to only TUPV, the DURV genome encodes a small hydrophobic (SH) protein in an ORF located between the M and G genes (Springfeld *et al*., 2005).

Although TUPV and DURV are currently the only two rhabdoviruses known to encode an SH protein, SH proteins have been described, and the functional roles investigated, in other related negative-sense RNA viruses, such as pneumoviruses and rubulaviruses within the family *Paramyxoviridae*. In *parainfluenza virus 5* (PIV5) [*simian virus 5*], expression of the SH protein has been demonstrated to inhibit tumor necrosis factor (TNF)-mediated apoptosis, thereby suggesting the protein may have a role in abrogating viral clearance (Lin *et al*., 2003; Wilson *et al*., 2006). Although the SH proteins of paramyxoviruses such as PIV5, *respiratory syncytial virus* (RSV), and *mumps virus* (MuV), share little sequence homology and are predicted to adopt different topologies (i.e., both type I and type II transmembrane proteins), it has been suggested that they all may function similarly to inhibit programmed cell death (He *et al*., 2001; Lin *et al*., 2003; Fuentes *et al*., 2007). Hence, although the DURV SH protein is not predicted to adopt a type I transmembrane topology as TUPV (i.e., it lacks a predicted signal peptide), the 1) two transmembrane domains flanking a short extracellular domain, coupled with a long cytoplasmic C terminus, 2) high level of amino acid conservation in the N-terminal half of the protein, along with the conserved leucine residues (Fig. 5.3D), and 3) identical genomic placement (Fig. 5.3A), strongly suggests that the SH protein plays a similar, albeit unknown, biological function in both viruses.

The C terminal region of the DURV SH protein shares homology to the SAM domain present in the EphA3 tyrosine kinase receptor (TKR) (Fig. 5.3D), which is also located distal to the transmembrane domain near the C terminus in EphA TKRs (Stapleton *et al*., 1999). SAM domains are found in many membrane-bound receptors and cytoplasmic signaling proteins and have the ability to dimerize, promoting the

oligomerization of SAM domain-containing proteins, which may subsequently facilitate their functional properties (e.g., modulating signal transduction) (Stapleton *et al*., 1999). Although unknown, this homology could potentially suggest that the interaction of the intracellular C-terminal SAM-like domains between two SH proteins, or the heterodimerization of the SAM-like domain of the SH protein with an analogous domain in a cellular protein (e.g., signaling receptor or its ligand), may influence the biological function of the DURV SH protein (e.g., inhibit signal transduction). It is also of interest that disregulation of EphA3 TKR signaling (i.e., mutations in the SAM domain) has been implicated in development of hepatocellular carcinomas (Bae *et al*., 2009), the tumor type from which the related TUPV was isolated (Kurz *et al*., 1986).

Comparison of the G protein of DURV with other rhabdoviruses revealed that both DURV and TUPV share the same cysteine configuration, in that they are missing the 8^{th} and 10^{th} cysteine residues as determined by Walker and Kongsuwan (1999). Although the cysteine configuration in the G protein for rhabdoviruses is well conserved, some viruses deviate from the 12 cysteine residue configuration observed in the vesiculoviruses and ephemeroviruses, and those that do, are phylogenetically related. For example, viruses within the genus *Lyssavirus* were demonstrated to be missing the 3^{rd} and 5^{th} residues, while viruses within the genus *Novirhabdovirus* were missing the 2^{nd} and 4^{th} residues (Walker and Kongsuwan, 1999), possibly implying that other rhabdoviruses related to DURV and TUPV may also share the same cysteine configuration. Secondary structure predictions for each of four genera analyzed (*Lyssavirus*, *Vesiculovirus*, *Empherovirus*, and *Novirhabdovirus*) indicated disulfide bridges between the 8^{th}-11^{th} and 9^{th}-10^{th} cysteine residues (Walker and Kongsuwan, 1999), suggesting that the disulfide

bridge connections in the G protein of DURV and TUPV, and the subsequent structural motif of the G protein, may be unique.

Unlike other viral glycoproteins involved in cell attachment and fusion, the rhabdovirus G protein is not proteolytically cleaved for activation and exists in a dynamic equilibrium between pre- and post fusogenic states (i.e., low-pH induced activation is reversible); these unique features, along with structural differences in relation to class I or class II fusion proteins, led to the designation of the rhabdovirus (VSINV) G protein as the prototype of a novel class of fusion proteins, termed class III (Roche *et al*., 2008; Sun *et al*., 2008). With VSINV, crystallographic studies have determined the structural organization of the pre- and post-fusion forms of the G protein and mapped the location of the two non-contiguous fusion loops that are involved in mediating fusion to the host cell membrane (^{72}-WY-^{73} and ^{116}-YA-^{117}) (Roche *et al*., 2006; Roche *et al*., 2007). Among rhabdoviruses for which G protein sequences are available, the fusion loop amino acid configuration of VSINV is shared only by other vesiculoviruses such as VSAV, COCV, CHPV, ISFV, and PIRYV. Although the bipartite fusion loops contain hydrophobic residues that are conserved among vertebrate and plant rhabdoviruses (Roche *et al*., 2006; Sun *et al*., 2008), DURV appears to be the only other rhabdovirus, other than the aforementioned vesiculoviruses, to contain the same amino acid configuration (^{82}-WY-^{83} and ^{126}-YA-^{127}) in the exposed ends of the fusion loops, suggesting that the biological conditions under which fusogenic activity of the DURV G protein occurs is very similar to that of the vesiculoviruses.

Although DURV did not exhibit cross-reactivity to members of *Vesiculovirus* genus by CF (Supplementary Table 5.1), the fusion loop motif identity, in concert with

the fact that DURV encodes of putative C protein (which additionally shares homology to the fusion domain of the G protein in VSINV) suggests that DURV, as demonstrated previously for TUPV by immunofluorescence assay (Calisher *et al*., 1989), is related to the vesiculoviruses. However, based on the phylogenetic analysis presented here (Fig. 5.5), a direct relationship between DURV and the vesiculoviruses could not be demonstrated, implying that until additional unclassified (or currently unknown) vertebrate rhabdoviruses are sequenced, the evolutionary pathway of these relationships will remain obscure. It is also of interest that DURV, like TUPV, in addition to having a bicistronic P gene, also encodes a small hydrophobic transmembrane protein analogous to the SH proteins of paramyxoviruses, which could potentially suggest that they represent ancestral rhabdoviruses that emerged early in the evolution of the two families. Additionally, as the 3' and 5' untranslated regions of the mRNAs of DURV are extremely short in sequence relative to other rhabdoviruses [as also seen in TUPV (Springfeld *et al*., 2005)], with stop codons lying within the putative polyadenylation sequence for a number of genes (Fig. 5.3B), this may also suggest an ancestral origin (mechanism) of these viruses.

Currently, of the vertebrate rhabdoviruses previously reported in the United States, only three have been isolated from birds: FLAV, HPV, and FARV. The normal association of FLAV and HPV with avian hosts is well-established, as both of these viruses are known to circulate in passerine bird-mosquito cycles in the eastern and western United States, respectively (Whitney, 1964; Hall *et al*., 1969, Crane *et al*., 1970). On the other hand, FARV, like DURV, constitutes a single isolate, and, therefore, its relationship with birds as a normal amplifying host in its transmission is only speculative

(Travassos da Rosa *et al*., 2002). Outside of the United States, most avian rhabdovirus isolates have been recovered through arbovirus surveillance conducted in Africa and, to a lesser extent, South America (Karabatsos, 1985). The vast majority of these isolates have come from the Central African Republic and include, in addition to KOLV and SJAV, Bangoran virus (BGNV), Bimbo virus (BBOV), Garba virus (GARV), Landjia virus (LJAV), Mossuril virus (MOSV), Nasoule virus (NASV), and Ouango virus (OUAV) (Kokernot *et al*., 1962; Karabatsos, 1985; Calisher *et al*., 1989). Some of these viruses have been isolated from birds on multiple occasions (e.g., KOLV and MOSV) (Karabatsos, 1985), suggesting that these are not incidental findings and that birds may represent the normal hosts for these viruses. Previous antigenic analysis of MOSV has revealed a serological relationship to FLAV and HPV (Calisher *et al*., 1989), potentially suggesting that North American and African avian rhabdoviruses may be genetically related. This relationship may not be unlikely, as the oceanic transportation of these viruses by migratory birds may be mechanism by which these viruses spread transcontinentally.

Kuzmin *et al*. (2006) recently reported on the phylogenetic relationships of seven unclassified rhabdoviruses, including KOLV and SJAV. Neighbor-joining analysis of the N gene revealed that KOLV and SJAV formed a monophyletic clade separate from other genera, and it was suggested that these two African avian rhabdoviruses should be tentatively considered a new genus. Interestingly, KOLV and SJAV were demonstrated to consistently group with TUPV, possibly suggesting a phylogenetic relationship between African avian rhabdoviruses and TUPV. The fact that DURV is closely related genetically to TUPV, and that both viruses share the same unique genomic organization,

reiterates that notion that TUPV may be linked phylogenetically to avian rhabdoviruses. Phylogenetic analysis of the DURV N protein demonstrated a monophyletic origin with TUPV, with both viruses additionally clustering with KOLV and SJAV to from a distinct clade (Fig. 5.5A). As suggested by Kuzmin *et al*. (2006), the phylogenetic relationship between TUPV, KOLV, SJAV, and now DURV, suggest that members of this group may be widely distributed on multiple continents (i.e., Africa, Asia, North America) in numerous host species (i.e., small mammals and birds).

As aforementioned, TUPV has been demonstrated to be antigenic related to known vesiculoviruses [i.e., VSAV and *Maraba virus* (MARAV)] (Calisher *et al*., 1989), leading to its current classification as a tentative member of the *Vesiculovirus* genus (Tordo *et al*., 2005; Lyles and Rupprecht, 2007). Additionally, TUPV has been shown to be antigenically related to KLAV (Calisher *et al*., 1989), another tentative vesiculovirus originally isolated from a montane vole (*Microtus montanus*) in Oregon and subsequently from Northern red-backed voles (*Clethrionomys rutilus*) and tundra voles (*Microtus oeconomus*) in Alaska (Johnson, 1965; De and Banjerjee, 1999). However, sequence confirmation of the relationship between TUPV and KLAV has been lacking. Phylogenetic analysis of a portion of the L protein of KLAV indicated that it formed a monophyletic clade with TUPV, with both viruses additionally grouping with DURV (Fig. 5.5B). This grouping of DURV, TUPV, and KLAV was well supported (89% bootstrap value), suggesting that both DURV and KLAV may constitute North American representatives of this proposed genus of rhabdoviruses. From an evolutionary standpoint, it would be of interest to see if other avian (e.g., KOLV and SJAV) or small mammal (e.g., KLAV) rhabdoviruses also share some of the unique genetic features (i.e.,

C and SH proteins, G protein cysteine configuration, G fusion loop motif) observed in DURV. Further genetic characterization of these and other morphologically or serologically-confirmed, yet unclassified, rhabdoviruses will likely provide additional insight into the phylogenetic relationships among the many vertebrate rhabdoviruses that are, at present, arbitrarily grouped as 'dimarhabdoviruses' (Bourhy *et al*., 2005; Kuzmin *et al*., 2009).

Currently, the specific host, transmission cycle, and geographic distribution of DURV are unknown. Serological testing of American coots and other bird species may disclose the prevalence, if any, of the virus in wild bird populations in the southeastern United States. Whether the virus is normally neurotropic or neuropathogenic in birds is speculative. As the bird from which the virus was isolated additionally had bacterial meningitis, it is unknown if the bacterial infection preceded or was secondary to the CNS infection of DURV. As this is the only isolate of DURV yet recovered, it appears unlikely that (if the virus is maintained normally in a transmission cycle involving birds) the virus has a propensity to be pathogenic in its normal host. This may be somewhat analogous to FLAV in the southeastern United States, which, although commonly isolated from mosquitoes during arbovirus surveillance, is only rarely obtained from dead wild birds (Mead *et al*., 2009). Additionally, the zoonotic potential of DURV is currently unknown. The wildlife rehabilitator who initially found and handled the bird extensively, although having a history of leptomeningeal inflammation of unknown etiology, did not have neutralizing antibodies to DURV based on plaque-reduction neutralization testing. Although inoculation of newborn mice did confirm that DURV could replicate and cause fatal disease in mammalian hosts *in vivo*, adult mice and hamsters, although susceptible

to infection, were refractory to clinical disease, suggesting that additional studies are needed in order to determine the pathogenicity of DURV.

Table 5.1. In vitro host range and replicative capacity of DURV and other selected rhabdoviruses.

	DURV	FARV	FLAV	KLAV	VSINV
Vero	6.68 (3)[a]	9.15 (2)	5.73 (2)	7.63 (4)	9.20 (2)
CPAE	5.67 (2)	7.60 (3)	nd	6.67 (6)	8.06 (2)
TB I Lu	5.24 (3)	7.62 (3)	nd	4.26 (4)	7.27 (5)
PDE	nd[b]	7.63 (2)	nd	4.20 (3)	6.42 (5)[c]
QNR/K2	6.72 (4)	8.33 (3)	nd	nd	7.55 (2)
C6/36	nd	nd	3.30 (4)	nd	nd
FHM	nd	6.94 (3)	nd	nd	nd
TH-1	nd	6.89 (3)	nd	nd	8.63 (3)
VH-2	nd	nd	nd	nd	7.58 (5)

[a]Maximum titers (log_{10} PFU/mL) reached over the eight-day growth period are given, followed by the day at which the maximum titer was reached in parentheses.
[b]nd = not detected above input level; i.e., maximum titer recovered never exceeded the initial inoculum/mL.
[c]PDE cells were essentially non-permissive to VSINV infection, as previously reported by Levinson *et al*. (1978); titer given at day five represents the single well of eight in which a titer above the initial inoculum was recorded. Nucleotide sequencing of the G protein from the initial inoculum versus virus harvested on day 5 did not disclose any nucleotide substitutions, suggesting adaptation of VSINV to PDE cells was not at the level of receptor binding and cell entry.

Table 5.2. Length of the DURV ORFs and associated untranslated regions and predicted length, molecular weight, isoelectric point, and hydrophobic index of the putative proteins.

DURV gene	5' UTR (nt)	ORF (nt)	3' UTR (nt)	Gene (nt)	Protein (aa)	MW (kDa)	pI (pH)	Hydrop. index
N	4	1,293	11	1,308	430	48.4	8.48	-0.231
P	12	1,014	6	1,032	337	36.6	4.70	-0.521
C	-	411	-	411	136	15.7	4.99	-0.476
M	13	582	11	606	193	21.9	6.73	-0.302
SH	13	234	6	253	77	9.2	4.77	0.790
G	19	1,521	6	1,546	506	57.1	6.38	-0.303
L	4	6,318	79	6,401	2,105	239.9	8.74	-0.145

Table 5.3. Pairwise amino acid identity of DURV proteins to other selected rhabdoviruses.

DURV	TUPV	FLAV	SVCV	VSINV	BEFV	RABV	IHNV	LNYV	SYNV
N	57	32	32	32	31	17	5	2	5
P	16	4	9	8	7	7	7	6	6
C	8	-	-	14	-	-	-	-	-
M	20	15	8	11	15	5	4	12	2
SH	25	-	-	-	-	-	-	-	-
G	25	19	18	18	14	12	15	8	2
L	51	38	39	41	37	29	12	14	14

[DURV, Durham virus; TUPV, Tupaia rhabdovirus (tentative vesiculovirus); FLAV, Flanders virus (unclassified, Hart Park serogroup); SVCV, Spring viremia of carp virus (tentative vesiculovirus); VSINV, *Vesicular stomatitis Indiana virus* (Genus *Vesiculovirus*); BEFV, *Bovine ephemeral fever virus* (Genus *Ephemerovirus*); RABV, *Rabies virus* (Genus *Lyssavirus*); IHNV, *Infectious hematopoietic necrosis virus* (Genus *Novirhabdovirus*); LNYV, *Lettuce necrosis yellows virus* (Genus *Cytorhabdovirus*); SYNV, *Sonchus yellow net virus* (Genus *Nucleorhabdovirus*)].

Supplementary Table 5.1. Rhabdovirus antisera used in complement fixation tests with DURV.

Vesiculovirus genus:	**Unassigned grouped rhabdoviruses**
Carajas	**Bahia Grande group**
Chandipura	Bahia Grande
Cocal	**Hart Park group**
Isfahan	Flanders
Maraba	Hart Park
Piry	Kamese
Vesicular Stomatitis Alagoas	Mosqueiro
Vesicular Stomatitis Indiana	Mossuril
Vesicular Stomatitis New Jersey	**Kern Canyon group**
	Barur
Tentative Species in the genus:	Kern Canyon
Boteke	Nkolbisson
Calchaqui	**Le Dantec group**
Gray Lodge	Le Dantec
Jurona	Keuraliba
Klamath	**Sawgrass group**
Kwatta	Connecticut
La Joya	New Minto
Malpais Spring	Sawgrass
Perinet	**Timbo group**
Porton	Chaco
Radi	Timbo
Yug Bogdanovac	**Other Unassigned animal rhabdoviruses**
	Almpiwar
Lyssavirus genus	Aruac
Duvenhage	Bangoran
Mokola	Bimbo
Rabies	Charleville
Ephemerovirus genus	Coastal Plain
Tentative Species in the genus:	Curionopolis
Kimberley	Farmington
Malakal	Garba
Puchong	Gossas
	Inhangapi
	Iriri
	Joinjakaka
	Kannamangalam
	Kolongo
	Kotonkon
	Manitoba
	Marco
	Navarro
	Obodhiang
	Parry Creek
	Sandjimba
	Sripur
	Tibrogargan
	Yata

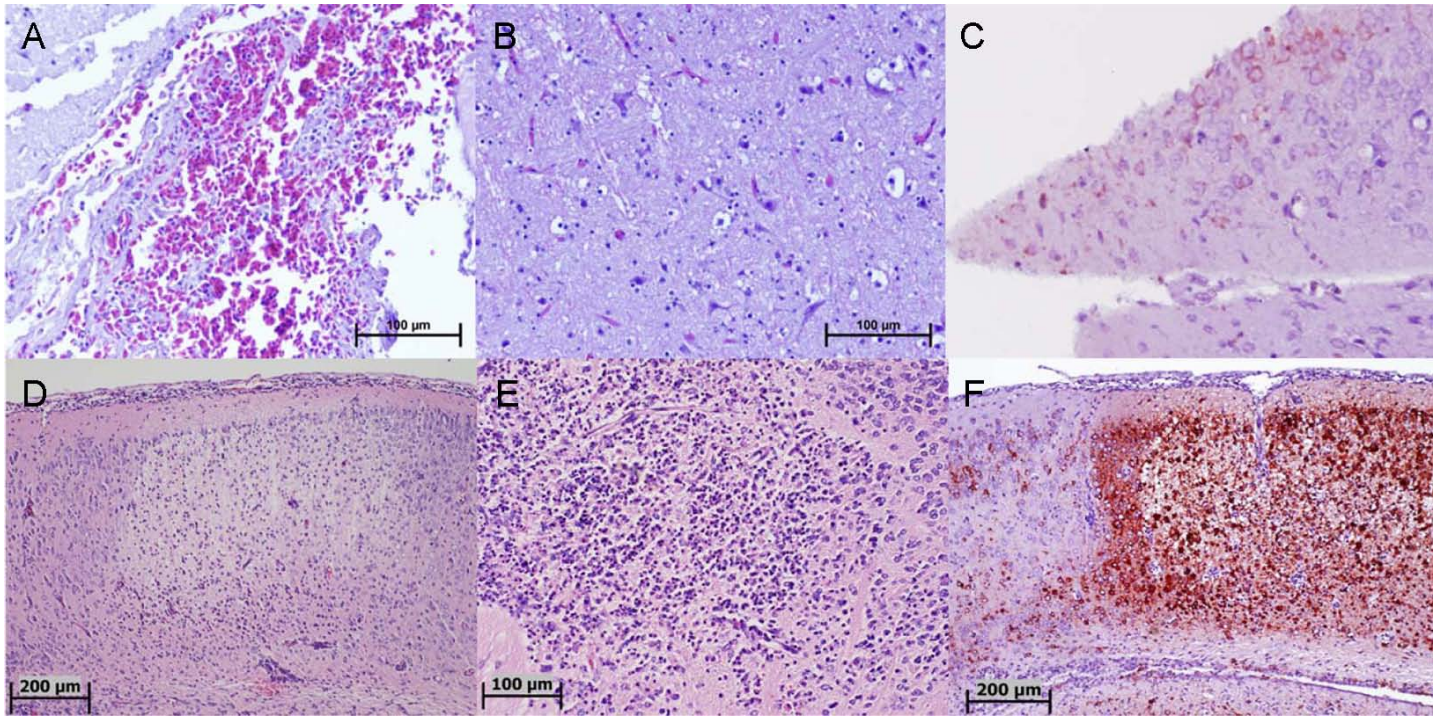


Figure 5.1. Pathology of DURV in avian and mammalian brain sections. (A-C): Sections of coot brain, naturally-infected with DURV. (A) Severe congestion in the meninges [H&E]; (B) Diffuse neuronal degeneration (accompanied by prominent vacuolation), mononuclear inflammatory cellular infiltration, and prominent capillary vessels [H&E]; (C) Immunohistochemical staining demonstrating some neurons with perinuclear DURV antigen. (D-F): Sections of mouse brain experimentally-infected with DURV. (D) Neuropil and neurons of the cerebral cortex; inflammatory cellular infiltrate has started from the bottom edge of the lesions. A prominent reactive blood vessel is visible (bottom) and marked meningitis is also present (top) [H&E]; (E) A necrotic focus in the deeper subcortical nuclei showing loss of neurons and infiltration with many neutrophils and macrophages [H&E]; (F) Immunohistochemical staining of same area of brain as shown in (D), demonstrating necrosis of the central cortex, with strong staining for DURV antigen. Scattered neurons in the adjacent area (far left) are also positive for viral antigen.

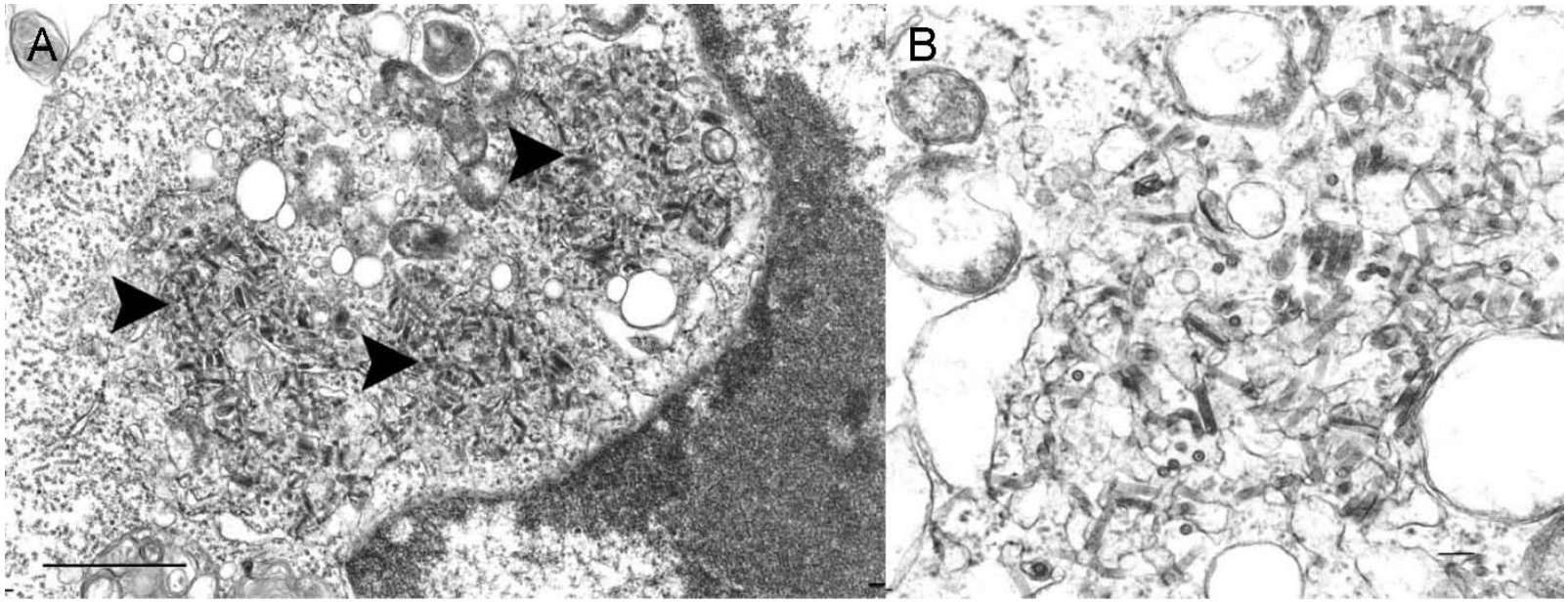
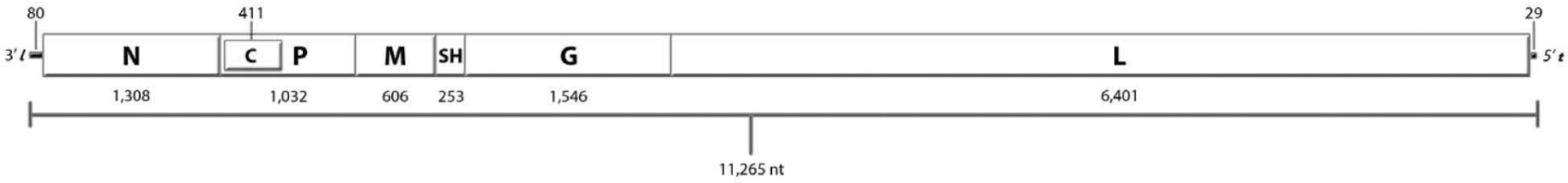


Figure 5.2. Ultrastructure of DURV in ultrathin sections of Vero cells. (A) Portion of a cell cytoplasm with three areas of virus formation (arrowheads). Bar = 1 µm. (B) Detail of a virus formation area, demonstrating virions budding into enlarged endoplasmic reticulum cisterns. Bar = 100µm.

Figure 5.3. DURV genome and regions of interest. (A) Schematic organization of the DURV genome. Including the 3' leader and 5' trailer sequences, the DURV genome was 11,265 nucleotides (nt) in length. The length of each gene, in nt, encoding the N, P, M, SH, G, and L proteins, is listed below the schematic. The length of the C gene (encoded in an overlapping reading frame of the P gene), along with the leader and trailer sequences, are noted above the genome schematic. (B) Transcription initiation, intergenic, and transcription termination/polyadenylation sequences. The start and stop codon for each gene is underlined and in bold. The intergenic sequence (i.s.) between each transcription unit, where applicable, is in bold. The consensus sequence for the transcription termination, intergenic (in bold), and transcription initiation regions is boxed. (C) Complementarity of the 3' leader and 5' trailer sequences, showing a 65% (19/29nt) inverse identity. Nucleotides that are conserved among the vertebrate rhabdoviruses (positions 1-3; 10) are in bold. (D) Amino acid alignment of the SH proteins of DURV and TUPV. Asterisks denote identity; colons and periods represent conserved and semi-conserved amino acid substitutions, respectively. Conserved leucine residues are underlined. The C-terminal region of the SH protein of DURV that shares homology to the α1 and α2 regions of the SAM domain of the EphA3 protein tyrosine kinase receptor (PDB 3FYD2), is shown in bold. (E) Amino acid sequence of the DURV C protein. The region of the C protein that shows homology to the fusion domain of VSINV (PDB 2CMZ), as well as to the dynein heavy chains of *Culex quinquefasciatus* (NCBI CPIJ002912) and *Aedes aegypti* (NCBI AAEL014313), is boxed.

A



B

Gene junction	stop	i.s.	start
leader-N		…GUUU**UAC**	
N-P	**ATC**UGACUUUUUUU**GA**UUGUGGGGAGGC**UAC**		
P-M	ACCUG**ACU**UUUUUU**GA**UUGUGUAGAGCAC**UAC**		
M-SH	**ACU**CGACUUUUUUU**GA**UUGUGGCAAGUCC**UAC**		
SH-G	GUAGG**ACU**UUUUUU**GA**UUGUGGAGAGCCACAUUUC**UAC**		
G-L	GCCAC**ACU**UUUUUU**GA**GUUC**UAC**		
L-trailer	**AUU**…65nt…AAUCUACUUUUUUU…		
Consensus	NBACUUUUUUU**GA**KUKY		

C

```
Leader    3′-UGCUUAUUUUGUUUUUAAGUCUAUUAAGU-
                |||||   |||||||    ||| | |||
Trailer   5′-ACGAAAUCAACAAAACAAAAGAGAUUUCG-
```

D

```
DURV   MYVELLVLGLFFLCGQRIFLFGLVYLLGRFNLLWFVFDAIMFCLWVF 47
TUPV   MITTLIIIGAAFLVGPRTFKFVLAYLLGYYNAFGPPLQIVQFMVWLI 47
       * . *:::*  ** * * * * *.**** :* :   :: : * :*::

DURV   CKNLPSEVLRLTGDWLKPSWESFQEEMYSS---------------- 77
TUPV   IIYFPKKFFSLGWYFCHDAFSSYFGDPNGGQLPVSTKFHSLTDMID 93
          :*.:.: *   : : ::.*:  :  ..
```

E

```
1-MTFQNSLTYVRSLKEGWMRTLEDWAEICQERWERDWARFLIDPGQVTTLLDCFHAMKEQCQSCVRRLYQMTTTPAWVEGTRSQIQSTLGLTNSTLSQTNPAVRPGEQDPIATELVALSRSLDESMEQACQAIYRPA-136
```

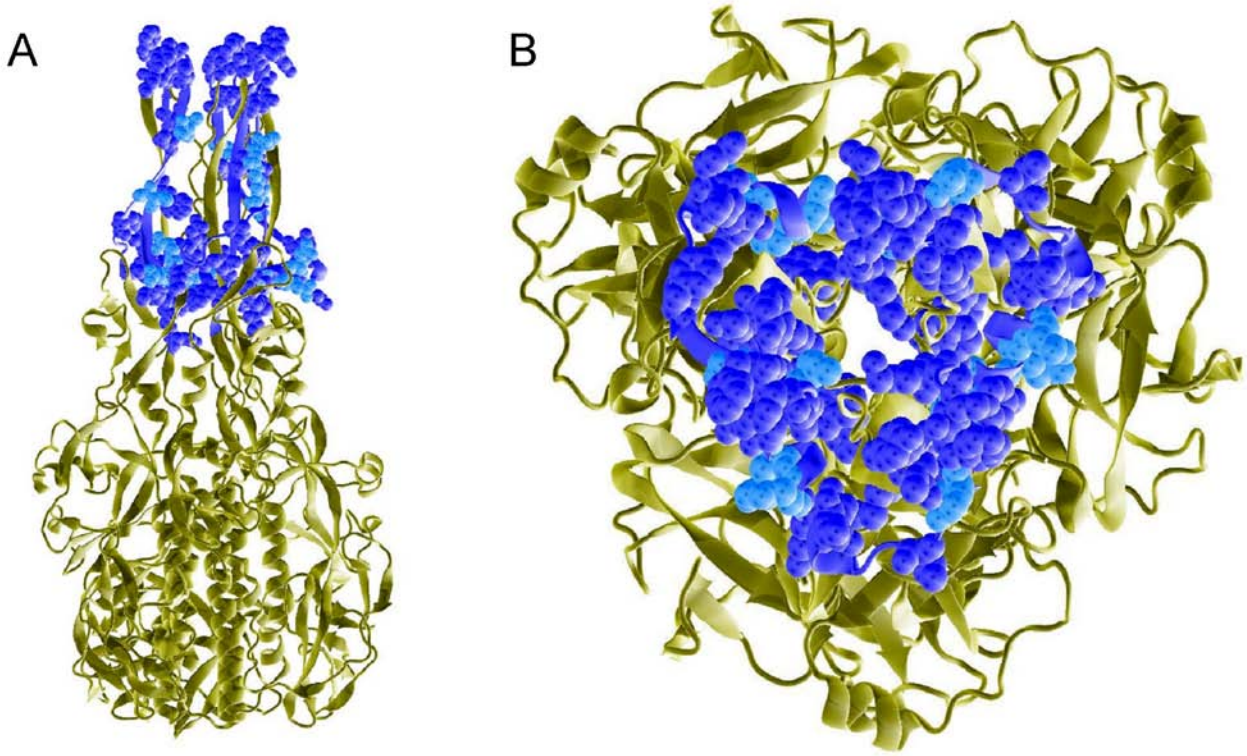
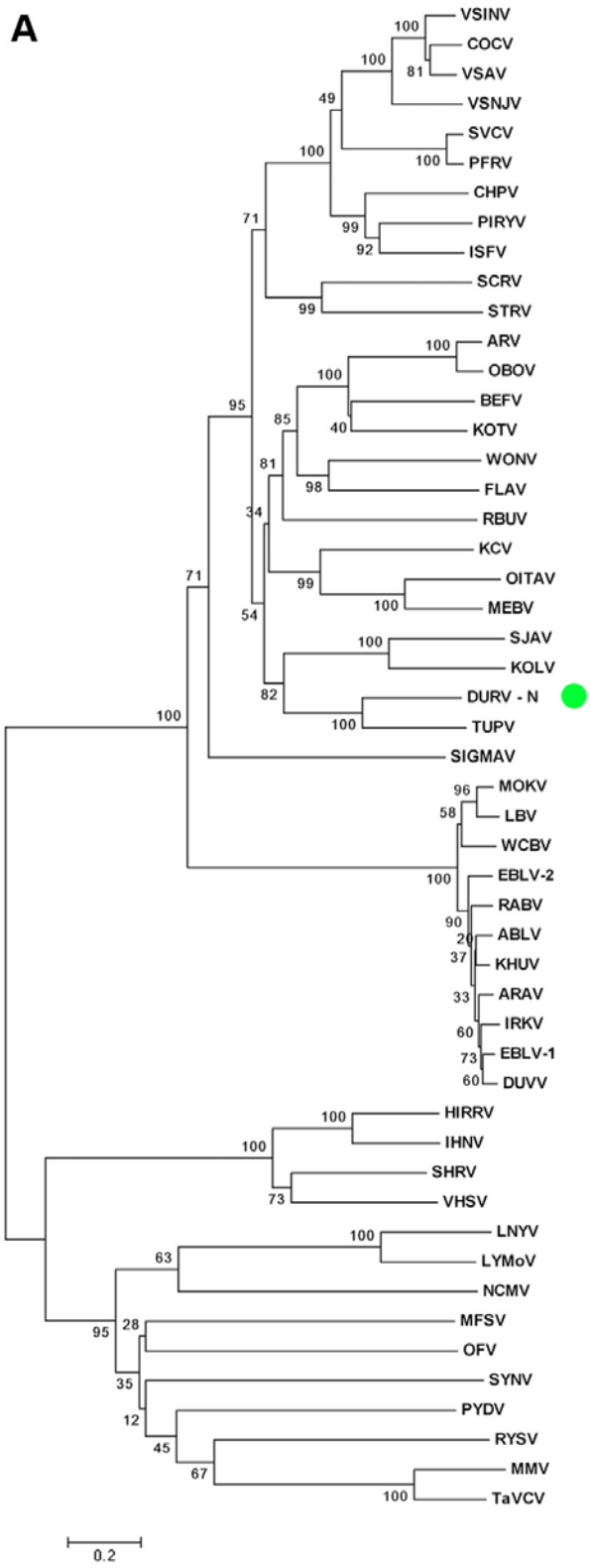
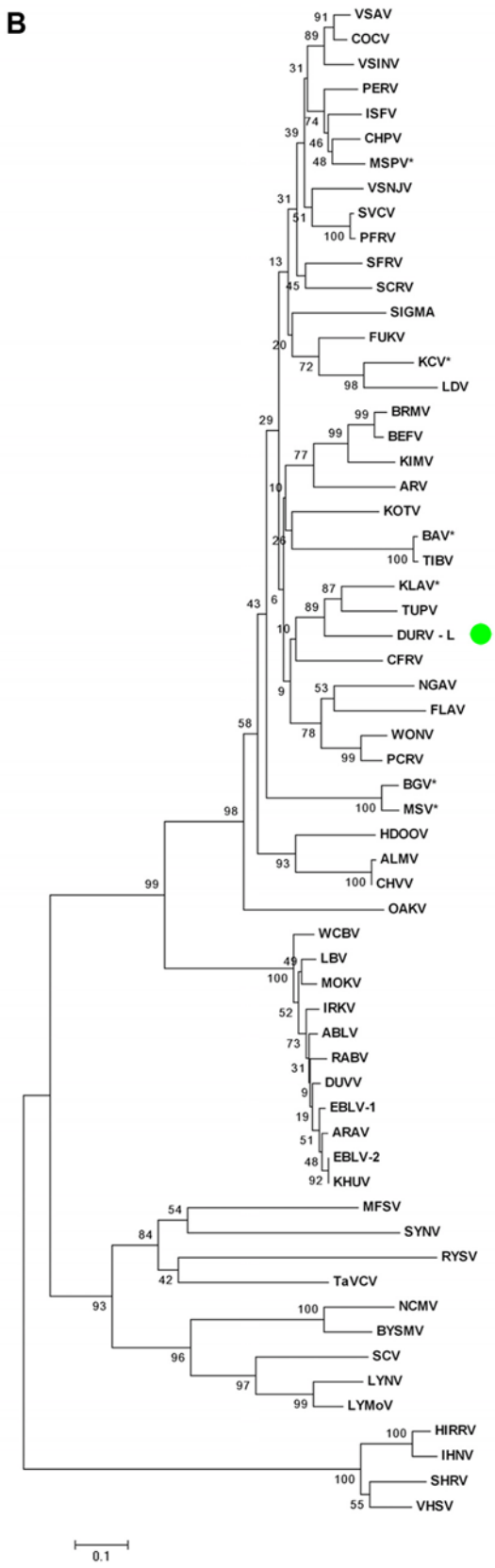


Figure 5.4. DURV C protein identity and homology with the fusion domain of the VSINV G protein. The ribbon schematic of the post-fusion homotrimeric form of the G protein of VSINV is shown in tan (PDB 2CMZ). The region of the DURV C protein (amino acids 15-71) which aligns with the VSINV G protein corresponds to domain IV (amino acids 55-103), which contains the fusion loops. Amino acids that exhibit identity or homology are rendered as van der Waals space-filling. The 15 amino acid residues of the DURV C protein which are identical to the VSINV G protein are shown in dark blue, while the five amino acids substitutions that are conserved are shown in light blue. Panel A shows a side view of the VSINV G protein with the distal ends of the bipartite fusion loops pointing upward, while panel B is an enlarged 90^{o} downward rotation of the fusion loops.

Figure 5.5. Evolutionary relationships of DURV with representative rhabdoviruses generated by neighbor-joining phylogenies of the N (A) and L (B) proteins. DURV sequences are indicated with a green circle. Asterisks beside ICTV abbreviations in the L tree denote new sequences. Phylogenetic trees were generated with a CLUSTALW alignment of the of partial N and L amino acid sequences of representative rhabdoviruses. Bootstrap values were determined using 2000 replicates and are listed at each node. Branch lengths are drawn to scale. The trees were calculated using Poisson correction and evolutionary distances are represented as the number of amino acid substitutions per site. Gaps in the alignments were analyzed by complete deletion. Amino acid sequences used to construct the N and L trees were: ABLV, Australian bat Lyssavirus (N: AAD01267; L: NP_478343); ALMV, Almpiwar virus (L: AAZ43273); ARAV, Aravan virus (N: Q6X1D8; L: ABV03822); ARV, Adelaide River virus (N: Q65111; L: AAG10421); BAV, Bivens Arm virus (L: GU085726); BEFV, Bovine ephemeral virus (N: NP_065398; L: NP_065409); BGV, Bahia Grande virus (L: tbs); BRMV, Berrimah virus (L: AAZ43265); BYSMV, Barley yellow striate mosaic virus (L: ACT21686); CFRV, China fish rhabdovirus (L: AAX86686); COCV, Cocal virus (N: ACB47434; L: ACB47438); CHPV, Chandipura virus (N: P11211; L: P13179); CHVV, Charleville virus (L: AAZ43300); DURV, Durham virus (FJ952155); DUVV, Duvenhage virus (N: Q66453; L: ABZ81216); EBLV-1, European bat lyssavirus (N: AAX62875; L: ABZ81181); EBLV-2, European bat lyssavirus 2 (N: YP_001285393; L: ABZ81191); FLAV, Flanders virus (N: AAN73283; L: AAN73288); FUKV, Fukuoka virus (L: AAZ43279); HDOOV, Humpty Doo virus (L: AAZ43271); HIRRV, Hirame rhabdovirus (N: ACO87995; L: NP_919035); IHNV, Infectious hematopoietic necrosis virus (N: Q08449; L: CAA52076); IRKV, Irkut virus (N: Q5VKP6; L: ABV03823); ISFV, Isfahan virus (N: Q5K2K7; L: Q5K2K3); KCV, Kern Canyon virus (N: ABE69215; L: tbs); KHUV, Khujand virus (N: Q6X1D4; L: ABV03824); KIMV, Kimberley virus (L: AAZ43266); KLAV, Klamath virus (L: GU085725); KOLV, Kolongo virus (N: ABE69214); KOTV, Kotonkon virus (N: ABE69213; L: AAZ43267); LBV, Lagos bat virus (N: ABF56214; L: ABZ81171); LDV, Le Dantec virus (L: AAZ43278); LNYV, Lettuce necrotic yellows virus (N: YP_425087; L: YP_425092); LYMoV, Lettuce yellow mottle virus (N: YP_002308371; L: YP_002308376); MEBV, Mount Elgon bat virus (N: ABE69217); MFSV, Maize fine streak virus (N: YP_052843; L: YP_052849); MMV, Maize mosaic virus (N: YP_052850; L: YP_052855); MOKV, Mokola virus (N: YP_142350; L: ABZ81211); MSPV, Malpais Springs virus (L: GU085727); MSV, Muir Springs virus (L: tbs); NCMV, Northern cereal mosaic virus (N: NP_057954; L: NP_597914); NGAV, Ngaingan virus (L: AAZ43277); OBOV, Obodhiang virus (N: ABE69212); OFV, Orchid fleck virus (N: BAH97109; L: YP_001294929); OITAV, Oita virus (N: BAD13431); OVRV: Oak Vale virus (L: AAZ43298); PCRV, Parry Creek virus (L: AAZ43275); PERV, Perinet virus (L: AAZ43280); PFRV, Pike fry rhabdovirus (N: ACP27998; L: ACP28002); PIRYV, Piry virus (N: P26037); PYDV, Potato yellow dwarf virus (N: ABW35154); RABV, Rabies virus (N: ACN51666; L: Q66T60); RBUV, Rochambeau virus (N: ABE69218); RYSV, Rice yellow stunt virus (N: NP_620496; L: NP_620502); SCRV, Siniperca chuatsi rhabdovirus (N: YP_802937; L: YP_802942); SCV, Strawberry crinkle virus (L: AAP03645); SFRV, Starry flounder rhabdovirus (L: AAS02285); SHRV, Snakehead rhabdovirus (N: NP_050580; L: NP_050585); SIGMAV, Sigma virus (N: ACV67011; L: ACU65438); SJAV, Sandjimba virus (N: ABE69216);

STRV, Sea trout rhabdovirus (N: AAL35756); SVCV, Spring viremia of carp virus (N: ABW24033; L: Q91DR9); SYNV, Sonchus yellow net virus (N: P10550; L: NP_042286); TaVCV, Taro vein chlorosis virus (N: YP_224078; L: YP_224083); TIBV, Tibrogargan virus (L: AAZ43274); TUPV, Tupaia rhabdovirus (N: YP_238528; L: YP_238534); VHSV, Viral hemorrhagic septicemia virus (N: P24378; L: CAB40833); VSINV, Vesicular stomatitis Indiana virus (N: P11212; L: NP_041716); VSAV, Vesicular stomatitis Alagoas virus (N: ACB47439; L: ACB47443); VSNJV, Vesicular stomatitis New Jersey virus (N: P04881; L: P16379); WCBV, West Caucasian bat virus (N: Q5VKP2; L: ABV03821); WONV, Wongabel virus (N: YP_002333271; L: AAZ43276).


A
VSINV
COCV
VSAV
VSNJV
SVCV
PFRV
CHPV
PIRYV
ISFV
SCRV
STRV
ARV
OBOV
BEFV
KOTV
WONV
FLAV
RBUV
KCV
OITAV
MEBV
SJAV
KOLV
DURV - N
TUPV
SIGMAV
MOKV
LBV
WCBV
EBLV-2
RABV
ABLV
KHUV
ARAV
IRKV
EBLV-1
DUVV
HIRRV
IHNV
SHRV
VHSV
LNYV
LYMoV
NCMV
MFSV
OFV
SYNV
PYDV
RYSV
MMV
TaVCV
0.2


B
VSAV
COCV
VSINV
PERV
ISFV
CHPV
MSPV*
VSNJV
SVCV
PFRV
SFRV
SCRV
SIGMA
FUKV
KCV*
LDV
BRMV
BEFV
KIMV
ARV
KOTV
BAV*
TIBV
KLAV*
TUPV
DURV - L
CFRV
NGAV
FLAV
WONV
PCRV
BGV*
MSV*
HDOOV
ALMV
CHVV
OAKV
WCBV
LBV
MOKV
IRKV
ABLV
RABV
DUVV
EBLV-1
ARAV
EBLV-2
KHUV
MFSV
SYNV
RYSV
TaVCV
NCMV
BYSMV
SCV
LYNV
LYMoV
HIRRV
IHNV
SHRV
VHSV
0.1

Acknowledgements

A. B. A., M. K. K., and D. E. S. were supported through sponsorship from SCWDS member state fish and wildlife agencies; through the Federal Aid to Wildlife Restoration Act (50 Stat. 917) and Grant Agreement 06ERAG0005, Biological Resources Division, U.S. Geological Survey, U.S. Department of the Interior; and through Cooperative Agreements 0596130032CA and 0696130032CA, Veterinary Services, Animal and Plant Health Inspection Service, U.S. Department of Agriculture. R. B. T, A. T. R., V. L. P. and S. Y. X. were supported by NIH contract NO1-AI25489. G. P., K. D., T. B., and W. I. L. were supported by NIH awards AI051292 and AI57158 (Northeast Biodefense Center - W. I. L.), the U. S. Department of Defense, and Google.org.

References

Alonso, M., Kim, C. H., Johnson, M. C., Pressley, M., & Leong, J. (2004). The NV gene of snakehead rhabdovirus (SHRV) is not required for pathogenesis, and a heterologous glycoprotein can be incorporated into the SHRV envelope. *J Virology* **78**, 5875-5882.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Bae, H. J., Song, J. H., Noh, J. H., Kim, J. K., Jung, K. H., Eun, J. W., Xie, H. J., Ryu, J. C., Ahn, Y. M., Kim, S. Y., Lee, S. H., Yoo, N. J., Lee, J. Y., Park, W. S., & Nam, S. W. (2009). Low frequency mutation of the Ephrin receptor A3 gene in hepatocellular carcinoma. *Neoplasma* **56**, 331-334.

Blom, N., Gammeltoft, S., & Brunak, S. (1999). Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **294**, 1351-1362.

Barr, J. N., Whelan, S. P. J., & Wertz, G. W. (2002). Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Biochim Biophys Acta* **1577**, 337-353.

Bourhy, H., Cowley, J. A., Larrous, F., Holmes, E. C., & Walker, P. G. (2005). Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene. *J Gen Virol* **86**, 2849–2858.

Calisher, C. H., Karabatsos, N., Zeller, H., Digoutte, J. P., Tesh, R. B., Shope, R. E., Travassos da Rosa, A. P., & St. George, T. D. (1989). Antigenic relationships among rhabdoviruses from vertebrates and hematophagous arthropods. *Intervirology* **30**, 241–257.

Chenik, M., Chebli, K., & Blondel, D. (1995). Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism. *J Virology* **69**, 707-712.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., & Thompson, J. D. (2003). Multiple sequence alignment with the CLUSTAL series of programs. *Nucleic Acids Res* **31**, 3497-3500.

Clarke, D. H., & Casals, J. (1958). Techniques for hemagglutination and hemagglutination-inhibition with arthropod-borne viruses. *Am J Trop Med Hyg* **7**, 561-573.

Clarke, G. G., Calisher, C. H., Crabbs, C. L., Canestorp, K. M., Tesh, R. B., Bowen, R. A., & Taylor, D. E. (1988). Malpais Spring virus: A new vesiculovirus from

mosquitoes collected in New Mexico and evidence of infected indigenous and exotic ungulates. *Am J Trop Med Hyg* **39**, 586-592.

Coll, J. M. (1995). The glycoprotein G of rhabdoviruses. *Arch Virol* **140**, 827-851.

Crane, G. T., Elbel, R. E., Klimstra, D. E., & Smart, K. L. (1970). Arbovirus isolations from mosquitoes collected in central Utah in 1967. *Am J Trop Med Hyg* **19**, 540-543.

Cropp, C. B., Prange, W. C., & Monath, T. P. (1985). LeDantec virus: Identification as a rhabdovirus associated with human infection and formation of a new serogroup. *J Gen Virol* **66**, 2749-2754.

De, B. P., & Banjeree, A. K. (1999). Ungrouped mammalian, bird and fish rhabdoviruses, p. 1541-1544. *In* A. Granoff, and R. G. Webster (ed.). Encyclopedia of Virology, 2nd ed. Elsevier Academic Press, San Diego, California.

Dietzgen, R. G., Callaghan, B., Wetzel, T., & Dale, J. L. (2006). Completion of the genome sequence of *Lettuce necrotic yellows virus*, type species of the genus *Cytorhabdovirus*. *Virus Res* **118**, 16-22.

Fu, Z. F. (2005). Genetic comparison of the rhabdoviruses from animals and plants. *Curr Top Microbiol Immunol* **292**, 1-24.

Fuentes, S., Tran, K. C., Luthra, P., Teng, M. N., & He, B. (2007). Function of the respiratory syncytial virus small hydrophobic protein. *J Virology* **81**, 8361-8366.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**, 3784-3788.

Gibbs, E. P. J., Calisher, C. H., Tesh, R. B., Lazuick, J. S., Bowen, R., & Greiner, E. C. (1989). Bivens Arm virus: A new rhabdovirus isolated from *Culicoides insignis* in Florida and related to Tibrogargan virus of Australia. *Vet Microbiol* **19**, 141-150.

Gubala, A. J., Proll, D. F., Barnard, R. T., Cowled, C. J., Crameri, S. G., Hyatt, A. D., & Boyle, D. B. (2008). Genomic characterisation of Wongabel virus reveals novel genes within the Rhabdoviridae. *Virology* **376**, 13-23.

Hall, R. R., McKiel, J. A., McLintock, J., & Burton, A. N. (1969). Arboviruses from Saskatchewan mosquitoes - isolation of a member of the Flanders-Hart Park group and of a strain as yet unidentified. *Can J Public Health* **60**, 486-488.

Harty, R. N., Paragas, J., Sudol, M., & Palese, P. (1998). A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with the WW domains of cellular proteins: Implications for viral budding. *J Virology* **73**, 2921-2929.

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

Hertig, C., Pye, A. D., Hyatt, A. D., Davis, S. S., McWilliam, S. M., Heine, H. G., Walker, P. J., & Boyle, D. B. (1995). Vaccinia virus-expressed bovine ephemeral fever virus G but not G_{NS} glycoprotein induces neutralizing antibodies and protects against experimental infection. *J Gen Virol* **77**, 631-640.

Hoffman, B., Schütze, H., & Mettenleiter, T. C. (2002). Determination of the complete genomic sequence and analysis of the gene products of the virus of spring viremia of carp, a fish rhabdovirus. *Virus Res* **84**, 89-100.

Huang, Y., Zhao, H., Luo, Z., Chen, X., & Fang, R.X. (2003). Novel structure of the genome of rice yellow stunt virus: identification of the gene 6-encoded virion protein. *J Gen Virol* **84**, 2259-2264.

Hughes, A. L. (2002). Adaptive evolution after gene duplication. *Trends Genet* **18**, 433-434.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD - Visual Molecular Dynamics. *J Mol Graphics* **14**, 33-38.

Jacob, Y., Badrane, H., Ceccaldi, P. E., & Tordo, N. (2000). Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein. *J Virology* **74**, 10217-10222.

Johnson, H. N. (1965). Diseases derived from wildlife. *Proc Vert Pest Cont Conf* **2**, 138-142.

Jordan, I. K., Sutter IV, B. A., & McClure, M. A. (2000). Molecular evolution of the Paramyxoviridae and Rhabdoviridae multiple-protein-encoding P gene. *Mol Evol* **17**, 75-86.

Karabatsos, N. (ed.) (1985). International catalogue of arboviruses, including certain other viruses of vertebrates, 3rd ed. American Journal of Tropical Medicine and Hygiene Press, San Antonio, Texas.

Kerschner, J. H., Calisher, C. H., Vorndam, A. V., & Francy, D. B. (1986). Identification and characterization of Bahia Grande, Reed Ranch, and Muir Springs viruses, related members of the family Rhabdoviridae with widespread distribution in the United States. *J Gen Virol* **67**, 1081-1089.

Killington, R. A., Stokes, A., & Hierholzer, J. C. (1996). Virus purification, p. 73-74 *In* Mahy, B. W. J., Kangro, H. O. (ed.), Virology Methods Manual, Academic Press, New York, New York.

Kondo, H., Maeda, T., Shirako, Y., & Tamada, T. (2006). Orchid fleck virus is a rhabdovirus with an unusual bipartite genome. *J Gen Virol* **87**, 2413-2421.

Kozak, M. (1989). The scanning model for translation: an update. *J Cell Biol* **108**, 229-241.

Kokernot, R. H., McIntosh, B. M., Worth, C. B., & De Sousa, J. (1962). Isolation of viruses from mosquitoes collected at Lumbo, Mozambique. II. Mossuril virus, a new virus isolated from the Culex (culex) sitiens Wiedemann group. *Am J Trop Med Hyg* **11**, 683-684.

Kouznetzoff, A., Buckle, M., & Tordo, N. (1998). Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. *J Gen Virol* **79**, 1005-1113.

Kretzschmar, E., Peluso, R., Schnell, M. J., Whitt, M. A., & Rose, J. K. (1996). Normal replication of vesicular stomatitis virus without C proteins. *Virology* **216**, 309-316.

Kurath, G., & Leong, J. C. (1985). Characterization of infectious hematopoietic necrosis virus mRNA species reveals a nonvirion rhabdovirus protein. *J Virology* **53**:462-468.

Kurz, W., Gelderblom, H., Flügel, R. M., & Darai, G. (1986). Isolation and characterization of a tupaia rhabdovirus. *Intervirology* **25**, 88-96.

Kuzmin, I. V., Hughes, G. J., & Rupprecht, C. E. (2006). Phylogenetic relationships of seven previously unclassified viruses within the family Rhabdoviridae using partial nucleoprotein gene sequences. *J Gen Virol* **87**, 2323–2331.

Kuzmin, I.V., Wu, X., Tordo, N., & Rupprecht, C.E. (2008). Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions. *Virus Res* **136**, 81-90.

Kuzmin, I. V., Novella, I. S., Dietzgen, R. G., Padhi, A., & Rupprecht, C. E. (2009). The rhabdoviruses: biodiversity, phylogenetics, and evolution. *Infect Genet Evol* **9**, 541-553.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-132.

Lamb, R. A. & Parks, G. D. (2007). Paramyxoviridae: The viruses and their replication, p. 1449-1496 *In* Knipe, D. M., Howley, P. M., Griffin, D. E., Martin, M. A., Lamb, R. A., Roizman, B., Straus, S. E. (ed.), Fields Virology, Vol. 1, Fifth ed. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.

Levinson, W., Oppermann, H., Rubenstein, P., & Jackson, J. (1978). Host range restriction of vesicular stomatitis virus on duck embryo cells. *Virology* **85**, 612-616.

Lin, Y., Bright, A. C., Rothermel, T. A., & He, B. (2003). Induction of apoptosis by paramyxovirus simian virus 5 lacking a small hydrophobic gene. *J Virology* **77**, 3371-3383.

Lo, K. W.-H., Naisbitt, S., Fan, J. S., Sheng, M., & Zhang, M. (2001). The 8-kDa dynein light chain binds to its targets via a conserved (K/R)XTQT motif. *J Biol Chem* **276**, 14059-14066.

Lyles, D. S., & Rupprecht, C.E. (2007). Rhabdoviridae, p. 1363-1408. *In* Knipe, D. M., Howley, P. M., Griffin, D. E., Martin, M. A., Lamb, R. A., Roizman, B., Straus, S. E. (ed.), Fields Virology, Vol. 1, Fifth ed. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.

Lynch, M., & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459-473.

Marriott, A. C. (2005). Complete genome sequences of Chandipura and Isfahan vesiculoviruses. *Arch Virol* **150**, 671-680.

Mead, D. G., Kelly, R., Allison, A. B., Lewandowki, H. B., Moulis, R. A., Blackmore, M., Gray, E. W., & Stallknecht, D. E. (2009). Mosquito surveillance for arboviruses in Georgia, USA: 2001-2007. *J Am Mosq Cont* (submitted).

Nowak, R. M. (ed.) (1999). Scandentia: Tupaiidae: Tree Shrews, p. 245-252 *In* Walker's Mammals of the World, Vol. 1, Sixth ed. John Hopkins University Press, Baltimore, Maryland.

Palacios, G., Quan, P. L., Jabado, O. J., Conlan, S., Hirschberg, D. L., Liu, Y., Zhai, J., Renwick, N., Hui, J., Hegyi, H., Grolla, A., Strong, J. E., Towner, J. S., Geisbert, T. W., Jahrling, P. B., Buchen-Osmond, C., Ellerbrok, H., Sanchez-Seco, M. P., Lussier, Y., Formenty, P., Nichol, M. S., Feldmann, H., Briese, T., & Lipkin, W. I. (2007). Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg Infect Dis* **13,** 73-81.

Palacios, G., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., Conlan, S., Quan, P. L., Hui, J., Marshall, J., Simons, J. F., Egholm, M., Paddock, C. D., Shieh, W. J., Goldsmith, C. S., Zaki, S. R., Catton, M., & Lipkin, W. I. (2008). A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med* **358,** 991-998.

Pauszek, S. J., Allende, R., & Rodriguez, L. L. (2008). Characterization of the full-length genomic sequences of vesicular stomatitis Cocal and Alagoas viruses. *Arch Virol* **153**, 1353-1357.

Peluso, R. W., Richardson, J. C., Talon, J., & Lock, M. (1996). Identification of a set of proteins (C' and C) encoded by the bicistronic P gene of the Indiana serotype of vesicular stomatitis virus and analysis of their effect on transcription by the viral RNA polymerase. *Virology* **218**, 335-342.

Poch, O., Blumberg, B. M., Bougueleret, L., & Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J Gen Virol* **71**, 1153-1162.

Pringle, C. R., & Easton, A. J. (1997). Monopartite negative strand RNA genomes. *Semin Virol* **8**, 49-57.

Raux, H., Flamand, A., & Blondel, D. (2000). Interaction of the rabies virus P protein with the LC8 dynein light chain. *J Virology* **21**, 10212-10216.

Roche, S., Bressanelli, S., Rey, F. A., & Gaudin, Y. (2006). Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science* **313**, 187-191.

Roche, S., Rey, F. A., Gaudin, Y., & Bressanelli, S. (2007). Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. *Science* **315**, 843-848

Roche, S., Albertini, A. A. V., Lepault, J., Bressanelli, S., & Gaudin, Y. (2008). Structures of vesicular stomatitis virus glycoprotein: membrane fusion revisited. *Cell Mol Life Sci* **65**, 1716-1728.

Rose, J. K., & Whitt, M. A. (2001). Rhabdoviridae: the viruses and their replication, p. 1221-1244 *In* Knipe, D. M., Howley, P. M., Griffin, D. E., Martin, M. A., Lamb, R. A., Roizman, B., Straus, S. E. (ed.), Fields Virology, Vol. 1, Fourth ed. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.

Scholthof, K. B., Hillman, B. I., Modrell, B., Heaton, L. A., & Jackson, A. O. (1994). Characterization and detection of sc4: a sixth gene encoded by sonchus yellow net virus. *Virology* **204**, 279-288.

Schütze, H., Mundt, E., & Mettenleiter, T. C. (1999). Complete genomic sequence of viral hemorrhagic septicemia virus, a fish rhabdovirus. *Virus Genes* **19**, 59-65.

Spiropoulou, C.F., & Nichol, S. T. (1993). A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus. *J Virology* **67**, 3103–3110.

Springfeld, C., Darai, G., & Cattaneo, R. (2005). Characterization of the tupaia rhabdovirus genome reveals a long open reading frame overlapping with P and a novel gene encoding a small hydrophobic protein. *J Virology* **79**, 6781-6790.

Stapleton, D., Balan, I., Pawson, T., & Sicher, F. (1999). The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization. *Nat Struct Biol* **6**:44-49

Stillman, E.A., & Whitt, M. A. (1998). The length and sequence composition of vesicular stomatitis virus intergenic regions affect mRNA levels and the site of transcript initiation. *J Virology* **72**, 5565-5572.

Sun, X., Belouzard, S., & Whittaker, G. R. (2008). Molecular architecture of the bipartite fusion loops of vesicular stomatitis virus glycoprotein G, a class III viral fusion protein. *J Biol Chem* **283**, 6418-6427.

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

Tan, G. S., Preuss, M. A., Williams, J. C., & Schnell, M. J. (2007). The dynein light chain 8 binding motif of rabies virus phosphoprotein promotes efficient viral transcription. *Proc Natl Acad Sci U S A* **104**, 7229-7234.

Tao, J. J., Zhou, G. Z., Gui, J. F., & Zhang, Q. Y. (2008). Genomic sequence of mandarin fish rhabdovirus with an unusual small non-transcriptional ORF. *Virus Res* **132**, 86-96.

Tanno, F., Nakatsu, A., Toriyama, S., & Kojima, M. (2000). Complete nucleotide sequence of Northern cereal mosaic virus and its genome organization. *Arch Virol* **145**: 1373-1384.

Tesh, R. B., Travassos da Rosa, A. P., & Travassos da Rosa, J. S. (1983). Antigenic relationship among rhabdoviruses infecting terrestrial vertebrates. *J Gen Virol* **64**, 169-76.

Tordo, N., Benmansour, A., Calisher, C., Dietzgen, R. G., Fang, R.-X., Jackson, A. O., Kurath, G., Nadin-Davis, S., Tesh, R. B., & Walker, P. J. (2005). Family Rhabdoviridae, p. 623-644 *In* Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., Ball, L. A. (ed.), Virus Taxonomy: Eighth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, California.

Travassos da Rosa, A. P., Mather, T. N., Takeda, T., Whitehouse, C. A., Shope, R. E., Popov, V. L., Guzman, H., Coffey, L., Araujo, T. P., & Tesh, R. B. (2002). Two new rhabdoviruses (Rhabdoviridae) isolated from birds during surveillance for arboviral encephalitis, northeastern United States. *Emerg Infect Dis* **8**, 614-618.

van Hulten, M. C. W., Goldbach, R. W., & Vlak, J. M. (2000). Three functionally diverged major structural proteins of white spot syndrome virus evolved by gene duplication. *J Gen Virol* **81**, 2525-2529.

Walker, P. J., Byrne, K. A., Riding, G. A., Cowley, J. A., Wang, Y., & McWilliam, S. (1992). The genome of bovine ephemeral fever rhabdovirus contains two related glycoprotein genes. *Virology* **191**, 49-61.

Walker, P. J., & Kongsuwan, K. (1999). Deduced structural model for animal rhabdovirus glycoproteins. *J Gen Virol* **80**, 1211-1220.

Wang, Y. & Walker, P. J. (1993). Adelaide River rhabdovirus expresses consecutive glycoprotein genes as polycistronic mRNA: new evidence of gene duplication as an evolutionary process. *Virology* **195**, 719-731.

Wang, Y., McWilliam, S. M., Cowley, J. A., & Walker, P. J. (1994). Complex genome organization in the GNS-L intergenic region of Adelaide River rhabdovirus. *Virology* **203**, 63-72.

Wertz, G. W., Whelan, S., LeGrone, A., & Ball, L. A. (1994). Extent of terminal complementarity modulates the balance between transcription and replication of vesicular stomatitis virus. *Proc Natl Acad Sci U S A* **91**, 8587-8591.

Whitney, E. (1964). Flanders strain, an arbovirus newly isolated from mosquitoes and birds of New York state. *Am J Trop Med Hyg* **13**, 123-131.

Wilson, R. L., Fuentes, S. M., Wang, P., Taddeo, E. C., Klatt, A., Henderson, A. J., & He, B. (2006). Function of small hydrophobic proteins of paramyxovirus. *J Virology* **80**, 1700-1709.

Xiao, S. Y., Guzman, H., Zhang, H., Travassos da Rosa, A. P. A., & Tesh, R. B. (2001a). West Nile virus infection in the golden hamster (*Mesocricetus auratus*): A model for West Nile encephalitis. *Emerg Infect Dis* **7**, 714-721.

Xiao, S. Y., Zhang, H., Guzman, H., & Tesh, R. B. (2001b). Experimental yellow fever virus infection in the golden hamster (*Mesocricetus auratus*). 2. Pathology. *J Infect Dis* **183**, 1437-1444.

CHAPTER 6

MUTATION:
THE MOLECULAR EVOLUTION OF A SPECIES JUMP BY CANINE PARVOVIRUS[1]

[1]Allison, A. B., J. D. Brown, M. G. Ruder, M. K. Keel, and D. E. Stallknecht. To be submitted to *Virology*.

Abstract

Canine parvovirus (CPV) is the etiological agent of acute gastroenteritis in dogs and regarded as one of the most important canine pathogens due to its severe clinical course in the absense of vaccination and prolonged stability in the environment. CPV presumably arose as a host range variant of *feline panleukopenia virus* (FPV), mediated by subtle amino acid changes in the VP2 capsid, the protein which controls the antigenicity and host range of all carnivore parvoviruses. Although the initial CPV variant that emerged in 1978 (CPV-2) was infectious only to dogs, the more recent antigenic strains of CPV that are currently circulating worldwide (CPV-2a, -2b, -2c) can also infect cats, demonstrating the ability of the virus to expand its host range after adaptation. Previously, raccoons have been shown not to be susceptible to CPV infection, although they are known to be the host of *raccoon parvovirus* (RPV), a virus very similar, if not identical, to FPV. However, during 2007-2010, outbreaks of severe gastroenteritis were documented in raccoons associated with rehabilitation facilities in the southeastern U. S., in which the virus isolated in all cases was not RPV, but rather a genetic variant of CPV-2a. The geographical distribution and high rate of nucleotide substitutions recognized among the isolates suggest that the virus is already widespread in the southeastern U. S. and may still be evolving to its new host. The genetic changes that facilitated the species jump and the epidemiological implications of this host range expansion are discussed.

Introduction

Aberrant interspecies transfer of viruses is a normal consequence of viral transmission and is most often recognized during periods of high virus activity (i.e., epidemic transmission), whereby bystander species not normally considered part of the transmission cycle of the virus become infected ('spillover'). For example, during periods of epidemic transmission of bird-associated mosquito-borne viruses, such as *West Nile virus* (WNV), spillover of the virus into humans or horses is often noted. In these cases, the jump of the virus (mediated by an arthropod vector) from its normal host (bird) into a new species (e.g., human, horse) results in a 'dead-end' of the transmission cycle, as the virus is not adapted to be transmitted efficiently between individuals of the new host species. As a consequence, the transmission cycle is not perpetuated in the new host, although it may be associated with mortality or morbidity.

The recognition of a virus jumping from one vertebrate species to another, followed by its adaptation and permanent maintenance in the new species, is a very rare event. However, such events, when they occur, can lead to the evolution of a new virus, which can have severe epidemiological consequences. The current AIDS pandemic was facilitated by such a species jump, whereby a chimpanzee virus [*simian immunodeficiency virus* (SIV) cpz] jumped into humans and subsequently adapted to become a human virus [*human immunodeficiency virus-1* (HIV-1)] (Gao *et al*., 1999). Although the possibilities for interspecies viral jumps seem endless, there are many barriers that exist to inhibit such jumps from being successful after they occur. Such barriers have been suggested to act on different fronts, including those at the level of cellular entry, intracellular replication, cellular exit, and intra and inter-host transmission

(Webby *et al*., 2004). However, in most cases, exactly what these barriers are and how they act to inhibit cross-species transfers and adaptation are essentially unknown. However, one noted exception was the jump of a feline parvovirus, *feline panleukopenia virus* (FPV), into dogs, which led to the emergence of a novel virus, *canine parvovirus* (CPV). In this case, the specific amino acid residues of the virus that determined how it was able to cross the species barrier and evolve into a new pathogen have been determined (Parker *et al*., 2001; Hueffer and Parrish, 2003; Hueffer *et al*., 2003; Palermo *et al*., 2006). Surprisingly, only very subtle genetic changes from its parental virus (i.e., two amino acid substitutions) were needed in order to initially change the host range of the virus.

It has been hypothesized that the ancestor of CPV arose by infection of a non-felid carnivore species (e.g., fox) with FPV or a variant of FPV, whereby this species may have served as an 'intermediate' host prior to the transmission and subsequent adaptation of the virus to dogs (Truyen, 2006). It has also been suggested that this ancestral virus may have first crossed-over into the dog population in Europe during the early 1970's, during which time it went undetected, presumably as it underwent further adaptation in its newly acquired canine host (Hueffer and Parrish, 2003; Shackelton *et al*., 2005). It wasn't until 1978 when CPV was first documented as a recognized pathogen of dogs, suggesting that it took several years for the virus to efficiently infect and be transmitted by its new host. Subsequently, after this period of 'silent' adaptation, CPV initiated a pandemic in the dog population, spreading worldwide in two years (Parrish *et al*., 1988b). The initial pandemic strain, CPV-2, although it was apparently a host range

variant of FPV, was demonstrated to have lost its feline host range (Truyen and Parrish, 1992; Truyen *et al*., 1996).

One of the major questions regarding the emergence of CPV was to identify the molecular determinants that allowed the virus to jump from cats into dogs. As FPV is not infectious to dogs, identifying mutations in CPV would likely disclose the amino acid substitutions responsible for the differences in host range between the two viruses. Subsequently, it was demonstrated that by changing only two amino acids in VP2 of FPV, residues 93 (K→N) and 323 (D→N), the virus could bind to and infect canine cells (Parrish, 1999). These two residues were not only demonstrated to be responsible for CPV binding to the canine transferrin receptor responsible for initiating infection in dog cells, but were also responsible for differences in the pH-dependent hemagglutination and antigenic (i.e., monoclonal antibody binding) properties observed between FPV and CPV (Chang *et al*., 1992), thereby affording biological characterization between the viruses prior to the advent of routine genetic analysis. In addition to changes at residues 93 and 323, three other amino acid substitutions, at positions 87 (M→Y), 300 (A→G), and 305 (D→Y), became fixed in the population as new CPV variants (CPV-2a/b) became widespread, quickly replacing the original CPV (CPV-2) (Parrish, 1999). One new biological consequence of the emergence of these new variants was their ability to infect cats, a property, as aforementioned, lost in the original CPV-2. Surprisingly, reacquisition of the feline host range by CPV-2a/b was not associated with reversion to the FPV sequence, but rather was likely the result of the additional changes occurring at residues 87, 300 and 305 (Truyen *et al*., 1995; Truyen *et al*., 1996). Two of these residues, 300 and 305, have now changed in the raccoon (*Procyon lotor*) CPV isolates

presented here, with, in some cases, position 305 reverting back to the original CPV-2 and FPV residue (Asp305; see Table 6.1).

Little is known regarding parvovirus infection in raccoons and the effects that parvoviruses may have on raccoon or other procyonid populations in the United States. Indeed, as there are no prevalence of infection or case fatality rates for wild raccoons, determining the impacts of parvoviruses at the population level is currently impossible (Barker and Parrish, 2001). Nevertheless, in rehabilitation settings (as witnessed here), infection and mortality may be high, especially in juvenile animals (Barker and Parrish, 2001). Although reports of raccoons exhibiting clinical symptoms suggestive of parvovirus infection (i.e., gastroenteritis, diarrhea) are fairly common, the subsequent isolation and genetic characterization of the viruses associated with those symptoms or lesions is very rare. Parvoviruses that have been isolated from raccoons have been designated as *raccoon parvovirus* (RPV), simply by host association (Parrish *et al*., 1987; Steinel *et al*., 2001); unfortunately, this designation has also been used to describe all parvovirus-like infections in raccoons, even in the absense of genetic characterization. Consequently, little information exists on which parvoviruses normally circulate in raccoon populations. Currently, only a single isolate of RPV is available in GenBank, which was obtained from a shipment of raccoons being translocated from Texas to West Virginia in 1978 (Nettles *et al*., 1980). Biological characterization (pH-specific hemagglutination and neutralization testing) of one of these isolates (79-4176) led the authors to suggest that the viruses might in fact be CPV, rather than FPV. However, retrospective sequence analysis of 79-4176 demonstrated that the virus was essentially identical to FPV (Parrish et al., 1988a). Raccoons have also been suggested to play a role

in the transmission of *Aleutian mink disease virus* (AMDV) [Genus *Amdovirus*], a parvovirus common to wild and farmed mink (*Mustela vision*) (Oie *et al*., 1996). Although raccoons were shown to be susceptible to experimental infection with AMDV, serial raccoon-to-raccoon transmission could not be demonstrated, nor were any pathological lesions observed (Oie *et al*., 1996), suggesting that natural infection of raccoons with AMDV infection may only be the result of spillover during epizootics among wild and farmed mink, and thus their role in maintaining the virus in nature is minimal or non-existent.

The comparative pathology and susceptibility of raccoons to CPV, FPV, and mink enteritis virus (MEV) was assessed by Barker *et al*. (1983). Although raccoons were shown to be highly susceptible to both FPV and MEV, resulting in severe clinical disease and death within six to 10 days of inoculation, raccoons experimentally infected with CPV-2 demonstrated no clinical signs, no lesions or tissue pathology, and shed only low levels of virus sporadically. These results, in addition to other experimental trials with raccoons and CPV (Appel and Parrish, 1992), suggested that while raccoons were clearly susceptible to FPV and MEV-induced disease, they were apparently refractory to CPV. However, despite these previous findings, the viruses isolated from raccoons from 2007-2010 in the southeastern United States are host range variants of CPV and are not FPV or RPV. Additionally, based on gross and histopathological examination of the raccoons, the CPV variants are capable of inducing severe pathological lesions in raccoons formerly only observed with RPV, FPV, or MEV (Nettles *et al*., 1980; Barker *et al*., 1983). Genetic analysis of the complete coding region of eight of the isolates from five different states indicated a number of amino acids residues in the VP2 protein that appear

to be under positive selection, suggesting these changes are involved in raccoon transferrin receptor binding, and hence, were responsible for mediating the species jump. The potential significance of these findings from wildlife (procyonid species) and domestic animal health (dog and cat) perspectives are discussed.

Material and Methods

Case histories

As of April 2010, a total of 21 CPV-positive cases (isolates or PCR positives) were collectively recovered from 11 raccoon accessions submitted to the Southeastern Cooperative Wildlife Disease Study (SCWDS) during 2008-2010. Tentative confirmation of parvovirus infection as the cause of death was determined by histopathology and/or immunohistochemistry. Differentiation of CPV from other carnivore parvoviruses as the etiological agent of the observed pathology was determined by sequencing of a portion of the VP2 gene from either virus isolated in culture or, in cases in which there was no available fresh of frozen tissue, by PCR directly from formalin-fixed tissues (see below). The eight isolates (six cases) in which full-length coding sequences were analyzed are listed below in chronological order from the date of death of the individual(s) (i.e., rather than date of submission or onset of outbreak) and are also shown in Fig. 6.1:

Case 1: CC 118-09-A - Clarke County, Virginia - September, 2007. Over a five-day period, mortality was documented in eight of eleven juvenile (4- to 5-months old) raccoons housed in outdoor enclosures at a rehabilitation facility. Four of the eight

raccoons exhibited clinical signs prior to death, including lethargy, anorexia, vomiting, and diarrhea; the other four died acutely without any observed clinical symptoms. This was the second consecutive year in which an outbreak in juvenile raccoons had been documented at the facility. All eight raccoons (CC 118-09, A-H) were submitted to the SCWDS for diagnostic evaluation and virus isolation.

Case 2: CC 287-08, 289-08, 349-08 - Glynn County, Georgia - October (*287-08, 289-08*) and November (*349-08*), 2008. Three juvenile (2- to 7-months-old) raccoons were submitted from a rehabilitation center on St. Simon's Island to SCWDS for diagnostic evaluation. These three individual animals, along with reportedly multiple other raccoons, were found dead either prior to, or after, release from the rehabilitation facility. In cases where clinical signs could be observed prior to death, symptoms consistent with parvovirus gastroenteritis, including severe diarrhea, lethargy, and dehydration, were noted.

Case 3: CC 351-09 - Knox County, Tennessee - May 2009. A wildlife rehabilitator acquired two litters of orphaned raccoons (newborns to 1.5-months of age) in April and May, 2009, respectively. Within days after receiving the second litter, members of both litters began displaying signs of parvovirus infection including anorexia, depression, vomiting and/or diarrhea. Although members from both litters died, some raccoons that were treated with subcutaneous fluids and parenteral or oral antibiotics displayed clinical disease for approximately 20 days and subsequently survived the infection. A fecal sample of one of the dead raccoons was submitted to the SCWDS for virus isolation.

Case 4: CC 381-09 - Lee County, Florida - July 2009. An infant raccoon at rehabilitation facility on Sanibel Island, after being hand-raised indoors, was moved to an outdoor enclosure with other juvenile raccoons. Within one month, this raccoon (along with its cage mates) began to display clinical signs including lethargy, ataxia and vomiting. Although symptoms initially improved, this individual was found ten days later exuding blood-tinged froth from it nares and in respiratory distress, and subsequently was euthanized and submitted to SCWDS.

Case 5: 278-A-09 - Fairfax County, Virginia - August 2009. Over the course of one month, four raccoons (<6 months of age) in two outdoor enclosures at a rehabilitation center in Lorton, Virginia died of apparent gastroenteritis. Animals began acting lethargic and displayed bloody and/or watery diarrhea prior to death. Previously, all four animals were vaccinated (twice; three weeks apart) with a Pfizer Felocell® FPV vaccine. Two animals were shipped to SCWDS for diagnostic evaluation.

Case 6: CC 358-B-09 - Scott County, Kentucky - November 2009. From July to November, 2009, a private raccoon rehabilitator had 23 raccoons die of suspected parvovirus enteritis. Twelve of these animals were sent to the Kentucky Livestock Diagnostic Disease Center for clinical evaluation and were diagnosed with severe diffuse necrotizing enteritis consistent with parvoviral enteritis. Animals had been vaccinated [e.g., Merial Purevax® Feline 4 (RCCP), Pfizer Vanguard® Plus, Schering Plough Galaxy®-DA2PPv and Eclipse 4®] prior to the recognition of clinical disease. Two animals (CC 358-09, A-B) were shipped to SCWDS for diagnostic evaluation.

Virus isolation

For virus isolation, samples of gastrointestinal tract, mesenteric lymph node, and/or brain (~0.5cm^3) obtained from the raccoons at necropsy were mechanically homogenized in 650µl of medium [1X minimum essential medium (MEM), 2.2g/l $NaHCO_3$, 20% fetal bovine serum (FBS), and 4X antibiotic/antimycotic solution (400 units/ml penicillin, 400µg/ml streptomycin, 1µg/ml amphotericin B) (Sigma)]. Homogenized tissues were centrifuged (6700 x *g* for 10 min) to pellet debris, and an aliquot (100µl) of clarified supernatant was used to inoculate freshly-seeded 12.5 cm^2 cultures of Crandell-Rees feline kidney (CRFK) cells and incubated at 37 °C in a humidified atmosphere containing 5% CO_2. CRFK cells were propagated in standard growth media [1X minimal essential media (MEM), 2.2g/l $NaHCO_3$, 5% fetal bovine serum (FBS), 100 units/ml penicillin, 100µg/ml streptomycin, 0.25µg/ml amphotericin B] (Sigma). No viral cytopathology was observed in any of the cultures. Once complete confluency was reached in the CRFK cell cultures, supernatant was harvested and DNA was extracted using a QIAamp® Viral DNA Mini kit (Qiagen, Valencia, CA) according to the manufacturer's instructions.

Pathology and immunohistochemistry

For pathological analysis of the raccoons, brain, small intestine, and mesenteric lymph node were fixed in 10% formalin, dehydrated in graded ethanol, cleared in xylene, and infiltrated and embedded in paraffin by routine methods. Deparaffinized and rehydrated tissue sections (5µm) were either used for immunohistochemistry (IHC) or stained with hematoxylin and eosin for microscopic evaluation. CPV IHC was

performed using a mouse anti-CPV/FPV monoclonal antibody (AbD Serotec) coupled with a biotinylated goat anti-mouse IgG and a streptavidin-horseradish peroxidase conjugate (Dako) according to the manufacturer's instructions. Antigen-antibody complexes were visualized using 3,3-diaminobenzidine (Dako) as the substrate chromogen. Tissue sections were counterstained with Gills II hematoxylin to demarcate nuclear structures, dehydrated in graded alcohol, cleared in xylene, and mounted using a xylene-based medium. Positive demonstration of viral antigen was based on chromogen deposition in the nucleus, with or without deposition in the cytoplasm.

Genetic and phylogenetic analysis

Virus that was isolated in culture was identified as CPV using primers covering a 417nt portion of the distal end of the VP2 gene that encodes the phylogenetically informative amino acid positions 564 and 568, thereby allowing differentiation of CPV from RPV, FPV, and MEV (see Table 6.1). Primers were then designed to cover the complete coding sequence of CPV (NS1, NS2, VP1, VP2), in addition to partial regions of the 3' and 5' ends of the genome. A list of primers may be obtained from the authors. Total sequence coverage of the overlapping primer sets was 4,627 nt or ~87% of the genome. PCR using extracted culture supernatant DNA was carried out with GoTaq Flexi DNA polymerase (Promega) according to the manufacturer's instructions. Cycling parameters were an initial denaturation step at 94 °C for 2 min to destabilize any secondary structure, followed by 40 cycles of denaturation at 94 °C for 1 min, annealing at 50 °C for 1 min, and extension at 72 °C for 1 min, with a final extension step at 72 °C for 10 min. Amplicons were electrophoresed on a 2% agarose ethidium bromide-stained

TAE gel, excised and purified from agarose using a QIAquick Gel Extraction Kit (Qiagen), cloned using a PCR Cloning^{Plus} Kit (Qiagen), and subsequently purified using a QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. Sequencing of clones was performed using an Applied Biosystems Inc. 3100 Genetic Analyzer (Foster City, California, USA). Contigs covering the full-length coding sequences were constructed using Sequencher 4.1.4. (Gene Codes Corporation).

Evolutionary relationships of the raccoon CPV variants with representative carnivore parvoviruses were derived by construction of phylogenetic trees generated with CLUSTALW alignments of the complete nucleotide sequence of the VP2 gene using the neighbor-joining method in the Molecular Evolutionary Genetics Analysis (MEGA) version 4 program (Tamura *et al*., 2007). Gaps in the alignments were analyzed by complete deletion. Bootstrap values were determined using 2000 replicates. The trees were calculated using the Kimura 2-parameter method and evolutionary distances were represented as the number of nucleotide substitutions per site. Nucleotide and protein identities among the isolates were determined by CLUSTALW alignment using the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) server (http://www.ebi.ac.uk/Tools/clustalw2/index.html) (Chenna *et al*., 2003).

Sequencing of the raccoon transferrin receptor

As the carnivore parvoviruses are known to utilize the transferrin receptor (TfR) in order to infect cells in their respective host, the mRNA of the raccoon TfR was amplified and sequenced. Pl 1Ut (NBL-9) cells, derived from uterine tissue of an adult female raccoon, were obtained from ATCC and were used as the source of mRNA. Pl 1

Ut cells were grown in standard growth media, except that it contained 10% iron-supplemented calf serum (Sigma) rather than FBS. At confluency, media was removed and exchanged with 3% FBS without iron supplementation. Twenty-four hours later, cells were trypsinized in a 1:10 dilution of 10X trypsin-EDTA solution (Sigma) in phosphate-buffered saline. Cells were then pelleted by low speed centrifugation (800 x *g* for 10 min), the trypsin solution was decanted, and mRNA was extracted from the cell pellet using an Oligotex Direct mRNA Mini kit (Qiagen) according to the manufacturer's instructions. Primers were designed based on the feline and canine TfR sequences; raccoon-specific primers were then designed from cloned sequences in order to fill in gaps and confirm ends. The full-length raccoon TfR mRNA was derived from a contig of overlapping sequences constructed using Sequencher 4.1.4.

CPV DNA amplification from formalin-fixed paraffin-embedded (FFPE) tissues

In order to retrospectively analyze tissues from previous suspect parvovirus cases for which only formalin-fixed paraffin-embedded (FFPE) tissues were available, primers flanking a portion of the VP2 gene containing the amino acid residues 297, 300, and 305 were developed. This primer set, CPV-FFPE-VP2-1-F and -1-R (5-ATGGCAAACAAATAGAGCA-3 and 5-GCACTATAACCAACCTCAGC-3), amplify a 186 bp portion of the VP2 gene covering VP2 nucleotide positions 834-1,019. Tissue sections from FFPE blocks that corresponded to areas that were positive by IHC were cut from each block using a sterile razor blade. Excess paraffin was removed and sections were placed in microcentrifuge tubes containing QuickExtract® FFPE DNA Extraction Solution (Epicentre). Samples were briefly centrifuged, incubated at 56 °C for 1 hr,

followed by incubation at 98 °C for 2 min. PCR and cloning using the CPV-FFPE-VP2-1-F/R primer set and 1uL of extracted FFPE DNA was performed as in the *Genetic and phylogenetic analysis* section.

Results

Pathological analysis and antigen distribution

Gross lesions in the raccoons were limited to the small intestines and lymph nodes. Depending upon the individual case, severe fibrinous exudate or brown-tinged fluid containing small flecks of fibrin were observed in the intestines. Mesenteric lymph nodes were enlarged in some animals. No other gross lesions were observed. Microscopically, the majority of the lesions were observed in the small intestines, including blunting and fusion of villi, ulceration of the mucosal surface, and dilation of the crypts of Lieberkuhn. When available for examination, similar lesions were observed in the large intestines. No neuropathology was observed.

Immunohistochemical analysis detected viral antigen most prominently in the lymph nodes and intestines (Fig. 6.2B). However, antigen was also seen in non-lymphoid tissues, including the brain, where positive staining for CPV was observed in the external germinal layer and in Purkinje cells (large neurons) of the cerebellum (Fig. 6.2A) in some raccoons.

Genetic and phylogenetic analysis

Contigs generated from overlapping cloned sequences of the eight raccoon CPV isolates listed in the *Case history* section were 4,646 nt long, and encompassed the entire region encoding the nonstructural and structural polyproteins (4,269nt). Nucleotide BLAST analysis of the earliest isolate from a 2007 outbreak in northern Virginia (CC 118-A-09), when trimming off the terminal non-coding regions (i.e., analyzing nonstructural and structural polyprotein gene sequences only), revealed the best alignment was to CPV-2 (CPV-6.us.80), with 13 nt changes between the two viruses (4256/4269 nt; 99%). Overall, the viruses indicated a high rate of evolution amongst themselves, such that an alignment of the complete sequences of all eight isolates revealed 75 nucleotide substitutions (Table 6.2). Whether the high rate of substitutions recognized amongst the eight different isolates is a result of the transfer of CPV into the raccoon population multiple times (i.e., these are different viruses) or is indicative of the rapid temporal and spatial evolution of the virus as it initially adapts to its new host (i.e., same virus), or both, is not clear and requires further analysis.

For the Georgia isolates (case 2), in addition to the three individual raccoons submitted during from the fall of 2008, the same variants of CPV have been retrospectively identified as the etiological agents of parvoviral enteritis in previous accessions from 2006 and early 2008, as well as from recent cases in 2009, suggesting that St. Simon's Island may be an endemic focus of ongoing virus activity. In the Kentucky case (case 6), two animals were submitted for diagnostic evaluation (CC 358-09, A-B). The Arg224/Asp300/Asp305 variant, also isolated from the Tennessee and Virginia (2007) cases, was recovered from the brain and gastrointestinal tract of one of

the raccoons (CC 358-B-09) (Fig. 6.3), which also had pathological lesions consistent with parvoviral enteritis. On the other hand, for the other animal submitted in tandem (CC 358-A-09), CPV could not be detected nor did the raccoon have lesions suggestive of parvovirus infection, but rather apparently died from a *Klebsiella pneumoniae* infection. However, FPV was isolated from the mesenteric lymph node of this raccoon (Fig. 6.3), possibly suggesting that it had been recently vaccinated, which was later confirmed, as it was disclosed that the raccoon had received a third and final dose of a Merial Purevax® Feline 4 live vaccine (containing FPV) 17 days prior to death. This additionally suggests that co-infection of raccoons with wild-type and vaccine strains of CPV and FPV may readily occur under these conditions and that the potential to generate new recombinants between the CPV variants and modified-live vaccine strains of FPV and CPV in these artificial rehabilitation settings may be great. However, in the isolates analyzed here, recombination was not detected.

Retrospective FFPE analysis of an accession from Virginia (CC 170-04) disclosed the presence of CPV in raccoons as far back as 2004, suggesting that, at least with this specific variant (His305), the 'jump' into raccoons occurred a minimum of six years prior (Fig. 6.3). Testing of a FFPE block from an accession originally submitted from Kentucky in 1993 (CC 121-C-93) identified *raccoon parvovirus* (RPV) as the etiological agent of the observed gastroenteritis, providing the first genetic evidence of RPV infection in raccoons during the 1990's in the southeastern U. S. Further comprehensive retrospective analysis of FFPE blocks from multiple southeastern and Midwestern states is currently underway, which will likely aid in attempting to determining an approximate

time frame of the jump (or jumps) of CPV into raccoons, as well as the prevalence of RPV in the raccoon population.

Phylogenetic analysis of the complete VP2 nucleotide sequence of the eight isolates with representative carnivore parvoviruses indicated that all the raccoon isolates clustered together within the major CPV clade, separate from FPV, MEV and RPV (Fig. 6.4). Within the raccoon CPV grouping, the eight sequences could be subdivided into two subclades: 1) a 'coastal' group containing Florida, Georgia, and Virginia (2009) viruses and an 2) 'Appalachian' group containing Kentucky, Tennessee, and Virginia (2007) isolates. This phylogenetic grouping into two subclades was likely based on the 224-305 residue configurations of the isolates (see below).

Comparative protein alignments of the isolates with each other and canine CPV isolates demonstrated that amino acid substitutions were observed throughout the entire genome, such that there was not a major preponderance for substitutions to occur more often in the structurally exposed proteins (e.g. VP2) relative to those that were nonstructural. In fact, when comparing all eight isolates, the number of amino acid substitutions occurring in the nonstructural polyprotein (NS1 and NS2; 17) was much greater than that collectively observed in the structural polyprotein (VP1 and VP2; 8) (Table 6.2). However, it should be noted that nine of the amino acid changes noted in NS2 were in the C terminus of the protein, which is encoded in an overlapping reading frame with NS1.

For the raccoon isolates, six amino acid substitutions were noted in VP2 at positions 190, 224, 269, 300, 305, and 541. Of these six changes, only positions 300, 305, and possibly 224, appeared to be under positive selection (Table 6.1 and Fig. 6.3).

Substitutions at residues 224 and 305 appeared to be potentially linked, as all southeastern isolates with an Asp at position 305 also had the Arg224 substitution (Table 6.1). Protein BLAST analysis revealed that the Arg224 was a very uncommon change, but was present in the Felocell® FPV vaccine strain. Nucleotide comparisons around the 224 region disclosed that the raccoon sequences were more similar to CPV than the Felocell® FPV vaccine (not shown), suggesting that the Arg224 seen in the Appalachian subclade may be a convergent substitution and did not arise by recombination. Isolates which had a His substitution (rather than the Asp substitution) at position 305, had a Gly residue at position 224, similar to CPV and other parvoviruses (Table 6.1). The His305 had previously never been documented, but similar to the Asp305 substitution, is likely involved in TfR binding. The amino acid substitution that is conserved among all the raccoon isolates is Asp300, which previously had only been documented in leopard cat isolates from Southeast Asia (Ikeda *et al*., 2000; Ikeda *et al*., 2002). Amino acid changes at positions 190, 269, and 541 were observed in only one isolate each, and do not appear to be informative substitutions at this time.

For substitutions unique to VP1 (i.e., not contained in VP2), only two were noted (residues 116 and 135), both which were confined to only one to two of the isolates and were highly conservative substitutions (i.e., K→R and V→I). NS1 (668aa) contained eight amino acid substitutions (at positions 350, 376, 470, 492, 544, 573, 583, and 602). For NS2, there were nine amino substitutions, all occurring in the C-terminal region that arises by alternative splicing and is contained within an overlapping reading frame of NS1. The very high nonsynonymous substitution rate within the C-terminal region of

NS2 suggests that negative selection is acting to conserve the function of NS1 more so than NS2.

Genetic analysis of the raccoon transferrin receptor (TfR)

The raccoon TfR was 2,310 nt long and encoded 770 amino acids, which was the same length as the canine TfR or one amino acid shorter than the feline TfR (Parker *et al*., 2001). In comparison to other mammalian TfRs that have been sequenced, the raccoon TfR shared the highest amino acid identity, in decreasing order, to the following species: 1) giant panda (*Ailuropoda melanoleuca*), 90%; 2) cat (*Felis catus*), 88%; 3) dog (*Canis lupus familiaris*), 88%; 4) horse (*Equus caballus*), 82%; and 5) human (*Homo sapiens*), 78%. The raccoon TfR, like other mammalian TfRs, can be divided into 5 domains: a cytoplasmic domain (residues 1 - 64), a transmembrane domain (residues 65 - 92), a stalk region (residues 93 - 131), a bipartite protease-like domain (residues 132 - 196 and 394 - 616), an apical domain (residues 197 - 393), and a helical domain (residues 617 - 770) (Lawrence *et al*., 1999). Analysis of the apical domain, which is believed to encompass the residues which interact with the CPV and FPV capsids (Palermo *et al*., 2003; Palermo *et al*., 2006), indicated that, in comparison to the canine and feline TfRs, the raccoon TfR was more similar to the feline TfR. Of the 197 amino acid residues in the apical domain, 155 were conserved between the raccoon, canine, and feline TfRs, while six residues were unique to all three (Fig. 6.5). Of the residues that were conserved only between two of the species, 17 were unique to only raccoon and feline, eight were unique between raccoon and canine, and 11 were unique to canine and feline.

Overall, the raccoon TfR had 50 amino acids (6.5%) that were unique in comparison to feline and canine sequences. The frequency of unique amino acid substitutions was highest in the stalk portion (0.21), while lowest in the transmembrane domain (0.00). The frequency of unique amino acid substitutions in the apical domain (0.056) was intermediate between the other extracellular domains [i.e., protease-like domain (0.087) and helical domain (0.026)]. When comparing the complete raccoon, canine, and feline TfR protein sequences, a total of 16 unique sites were identified, of which six occurred in the apical domain.

One of the differences previously noted between canine and feline TfR was a Asn substitution as position 384 in the canine TfR that introduced a potential glycosylation site ($^{384-}$NLT^{-386}), suggesting that additional glycosylation may be involved in mediating CPV attachment to the canine TfR (Palermo *et al*., 2003). Analysis of the raccoon TfR, however, revealed a Lys at position 384 (i.e., $^{384-}$KLT^{-386}), similar to the feline receptor, and thus it does not have the potential glycosylation site, suggesting that glycosylation at that position may not be essential for CPV binding to the raccoon TfR. The raccoon TfR protein contained five Asn-X-Ser/Thr motifs, three of which, $^{261-}$NGS^{-263}, $^{327-}$NHT^{-330} and $^{737-}$NET^{-740} were predicted to be N-glycosylated. Interestingly, the three potential sites that are predicted to be glycosylated were strictly conserved between raccoon, canine and feline TfR sequences, while the other two motifs (not predicted to be glycosylated) at $^{537-}$NWS^{-539} and $^{732-}$NNS^{-734} were not conserved between the three species.

Discussion

The close association of raccoons to developed areas and their known predilection to opportunistically forage around human dwellings (i.e., more so than most wild terrestrial mammal species), brings raccoons into close contact with domestic species such as cats and dogs on a regular basis. This close and recurrent contact undoubtedly facilitates the transfer of viruses between these domestic and wild species, which the recognition of CPV in raccoons presented here attests to. However, when or where the transfer event(s) of CPV from dogs and/or cats into raccoons occurred is unknown.

Based on the defining antigenic residue at position 426, all the raccoon CPV isolates appear to be host range variants of CPV-2a. Although the raccoon viruses are also very similar to the presumably extinct CPV-2 (although it is still used in vaccines), none of the isolates contain a Met at position 87 characteristic of CPV-2 (Table 6.1). In comparison to the VP2 protein sequences of other CPV isolates, six amino acid substitutions were noted in the raccoon variants at positions 190, 224, 269, 300, 305, and 541 (Table 6.2). Of these six changes, only positions 300 and 305 (and possibly 224) appeared to under positive selection, with 224 and 305 potentially being concomitant compensatory substitutions (Table 6.1). For example, for southeastern isolates in which an Asp305 substitution (from a Tyr) was observed, this amino acid change was always concomitant with an Arg224 substitution (from a Gly), such that an isolate containing only one of these two substitutions has not yet been recovered. The Wisconsin Asp305 isolate, although not containing the Arg224 substitution, did contain an additional substitution (Thr232) in the same location (loop 2 of antigenic site A; see Figure 6.6). Viruses which had the His305 change (rather than the Asp), which was a substitution that

had previously never been observed before in any CPV isolates, did not have a reciprocal change in loop 2, but rather maintained the Gly224 or Ile232 as observed in all other carnivore parvoviruses (Table 6.1). This result may suggest that the Asp305 substitution, when it occurred in CPV-2a, may have been associated with an additional compensatory change in loop 2, or vice versa. If these amino acids changes are structurally linked (305 and 224/232), it may be suggestive that both are involved in transferrin binding to some degree, although until functional studies are performed, this is speculative. It may also be that Arg224 or Thr232 in the Asp305 viruses [Kentucky, Tennessee, Wisconsin, or Virginia (2007)] are simply independent, non-related changes.

For CPV, two dominant neutralizing immunogenic sites have been defined on the VP2 capsid, designated as antigenic sites A and B (Fig. 6.6). Site A involves three separate loop structures (loops 1, 2, and 4) and includes the CPV-specific epitope, residue 93, which (along with residue 323) defines the canine host range (Strausshiem *et al*., 1994). Site B, involves a single loop (loop 3) which is located on the shoulder of the three-fold spike, and includes amino acids in close proximity to residue 300 (Strausshiem *et al*., 1994). Both antigenic sites contain residues that are known to be in contact with the host transferrin receptor. Residue 224 is located on loop 2, which along with loop 1 and loop 4 from either the same or a neighboring VP2 molecule, respectively, form the conformationally-dependent antigenic site A. Previous mutational analysis of the exact same change at position 224 as observed in the raccoon variants (i.e., G224R) disclosed that it induced antigenic changes in the virus such that site A/residue 93-specific monoclonal antibodies (mAbs) no longer recognized the virus, while binding by a site B mAb was unaffected (Hueffer *et al*., 2003; Parker and Parrish, 1997). Thus, the change

to Arg224 in the raccoon isolates likely alters the structure of antigenic site A of the virus, including residues known to contact the host cell receptor (e.g., residue 93), but likely does not alter site B (which contains position 305). As such, the Asp305 substitution may therefore compensate for this alteration, or vice versa. However, as aforementioned, how and if the Arg224/Thr232 and Asp305 substitutions are functionally linked remain to be determined.

In addition to the VP2 substitution at residue 305 (Asp/His), the second VP2 position that was changed in all variants was at residue 300, in which all viruses had an Asp substituted for a Gly (Table 6.1). As stated earlier, residue 300 is located on loop 3 (Fig. 6.6), which along with other residues around position 300 (i.e., residue 305), are contained within antigenic site B (Strausshiem *et al*., 1994). Prior to this study, the Asp300 substitution had previously only been documented during natural infections of leopard cats (*Prionailurus bengalensis*) from Vietnam (Ikeda *et al*., 2000; Ikeda *et al*., 2002). These isolates have been arbitrarily designated as CPV-2c(a) or 2c(b) in order to differentiate them from CPV-2a and 2b, although the parenthesized letter demonstrates whether they were derived from either the CPV-2a (Asn) or 2b (Asp). In other words, the CPV-2c(a) and CPV-2c(b) leopard cat isolates are host range variants of CPV-2a and CPV-2b, respectively. However, it should be noted that another antigenic variant of CPV, containing a characteristic Glu substitution at position 426, has also been designated as CPV-2c (Table 6.1) (Decaro *et al*., 2009). Nevertheless, the defining mutation of significance of these Vietnamese isolates is the Asp300.

The Asp300 substitution in the leopard cat isolates was demonstrated to induce remarkable changes in their antigenic profiles as demonstrated by hemagglutination-

inhibition tests using a number of mAbs (Ikeda *et al*., 2002), which raises the question if this antigenic change was significant enough to avoid vaccine-elicited immunity. Although the Asp300 leopard cat CPV isolates were shown to be more infectious and induced a higher frequency of disease in cats than CPV-2a (Nakamura *et al*., 2001), the ability of currently-used vaccines to neutralize the virus, or the ability of the virus to infect dogs *in vivo*, has never been assessed.

The fact that the Asp300 substitution has only been demonstrated in leopard cats and raccoons could imply that either 1) the Asp300 substitution is associated with infection of a non-canid species, such that raccoons and leopard cats may have similar TfR binding sites, or 2) the Asp300 substitution was selected by passage of CPV-2a from dogs into cats, and then into raccoons, such that cats served as an intermediate host in the emergence of the variants. Previously, the Asp300 substitution has been shown to occur upon serial passage of CPV isolates in CRFK cells (Parrish and Carmichael, 1986; Hoelzer *et al*., 2008), suggesting that long-term maintenance (i.e., *in vivo* passage) of CPV in cats may be associated with positive selection for Asp300 over Gly300. However, the Asp300 substitution may also be a convergent substitution due to similarities between the transferrin receptors of raccoons and cats, as the apical domain of the TfR was demonstrated to be more similar genetically between raccoons and cats than either between raccoons and dogs or between dogs and cats (Figure 6.5).

With CPV-2, a change from Ala300 to Asp300 in VP2 has been demonstrated experimentally to result in neutralization escape and a loss of the canine host range (Llamas-Suiz *et al*., 1996). Specifically, the Asp300 substitution, unlike Ala300, was demonstrated to form a salt bridge with Arg81. Not only did the formation of a salt

bridge induce a change in the structure and charge distribution of antigenic site B that allowed the virus to evade neutralization, but it was also suggested to be responsible for the reduction and/or elimination of the ability of the virus to replicate in canine cells due to an increase in stability (i.e., difficulty of uncoating of the virus) (Llamas-Suiz *et al*., 1996). Alternatively, the loss or reduction in the canine host range may have been related to a decrease in the binding efficiency of CPV-2 to the canine TfR due to the Asp300 salt bridge formation, rather than an increase in the stability of the virus. Nevertheless, as the Asp300 substitution is associated with changes in neutralization and receptor binding, it may have profound influences on the epidemiology of the virus, such that, in conjunction with changes at positions 305 and/or 224, these mutations may induce such a major antigenic change (i.e., occurring in both major immunodominant sites) that current vaccines are no longer protective, as has been suggested in some instances for other (less variable) antigenic variants of CPV (Decaro *et al*., 2008). However, at the same time, since Asp300 is associated with the loss of canine host range *in vitro*, these CPV variants may have lost their ability to infect dogs *in vivo*.

Although CPV was first recognized as a pathogen in 1978, it has been suggested that it may have emerged up to 10 years before it was initially described, possibly circulating in the dog population while acquiring beneficial mutations that increased its infectivity to dogs, and hence transmissibility, allowing it to spread globally (Shackelton *et al*., 2005). As such, the raccoon CPV variants may represent a similar potential epidemiological scenario. Comparative alignment of the raccoon CPV isolates disclosed that all viruses, other than the Wisconsin isolate (CC 37-10), contained an Ala297 rather than the Ser297 (Figure 6.3). Global analysis of CPV isolates demonstrated that Ala297

first appeared in Europe in 1993, and quickly spread in the dog population, suggesting that the replacement of Ser297 with Ala297 was advantageous to infection and/or transmission, with Ser297 ultimately becoming distributed worldwide during the 1990's (Parrish and Kawaoka, 2005; Truyen, 2006). However, analysis of U. S. isolates from 1997 did not indicate that all viruses had the Ala297 substitution (Truyen, 1999), suggesting that Ala297 may have not become dominant in the dog population in the United States until after 1997. Nevertheless, based on the fact that the Wisconsin isolate has a Ser at position 297 is suggestive that the jump, in this particular case, was not a recent event and may have occurred 10 or more years ago. Obviously, the discrepancy of a Ser or Ala residue at position 297 in different isolates also suggests that the jump was not a single event, but that CPV has been transferred into raccoons on multiple independent occasions. This is somewhat analogous to the emergence of HIV-1, as it has been suggested that the progenitor virus, SIV cpz, may have jumped into the human population on at least three different occasions (Gao *et al*., 1999).

Although the Ser297 of the Wisconsin isolate suggests that that particular virus may have been present in raccoon populations for quite some time, the fact that numerous reports of widespread morbidity and mortality of rehabilitated raccoons have been described only within the last few years in the southeastern U. S. is suggestive that this is likely a new phenomenon (as implied by the Ala297) and that the incidence or severity of disease associated with these CPV variants is on the rise, which may presumably be a result of the adaptation and/or recent emergence of these genetic variants. Similar to that hypothesized with CPV-2 in the dog population, these Ala297 variants may have been circulating at a low level for a number of years (~5-10 years) in the wild raccoon

population in the southeastern U. S., during which time the virus has been adapting to infect raccoons more efficiently. Consequently, the increased frequency of mortality events and the spread of the virus into rehabilitation settings may simply be an indication that the virus is now being transmitted efficiently among raccoons and is becoming endemic in wild populations. Whether these variants of CPV have (or will) lead to the displacement of RPV in raccoon populations, or whether both viruses will be sympatric in nature, is unknown. As very little is known about the current distribution or prevalence of RPV in raccoon populations, gauging the impacts these variants will have on RPV is impossible.

Based on the ability of the virus to persist for very long periods in the environment (Bouillant and Hanson, 1965), and as raccoons stretch across Canada and the entire continental United States south to Panama (Nowak, 1999), there is little reason to suspect that the virus should not reach all raccoon populations throughout its entire geographical range. Once CPV emerged as a recognized pathogen in 1978, the virus went pandemic in two years. It is estimated that between 1978 and 1979, >80% of the world's population of domestic dogs and wild canids were infected with CPV (Parrish and Kawaoka, 2005). Although this monumental spread was likely facilitated by the social and gregarious nature of humans and their pets, it does demonstrate the ease at which an environmentally-stable virus can spread in a naive population. Whether background immunity to RPV (if it exists) will limit the spread and impact of the new variants is speculative. One possible consequence of the emergence of these CPV variants is their spread to other related procyonid species. In contrast to North America, which has only three endemic procyonid species [raccoons, ringtails (*Bassariscus*

astutus), and white-nosed coatis (*Nasua narica*)], Central and South American has a rich diversity of procyonid species, from raccoons and coatis to kinkajous and olingos (Nowak, 1999). Based on the similarity of the raccoon TfR to both canine and feline TfR (Figure 6.5), it is likely that other procyonid species related to raccoons may be susceptible to infection with these variants, if the virus was to spread southward.

In susceptible hosts, carnivore parvoviruses are spread through the fecal-oral/nasal route and due to their requirement for host cell machinery in order to replicate, parvovirus infection is generally believed to be confined to cells that are in the S-phase of cell division (Barker and Parrish, 2001). About two to four days post-infection, the virus infects the crypts of the villi in the ileum and jejunum, resulting in the destruction and sloughing of the epithelial layer of the intestines leading to hemorrhagic diarrhea (Parrish and Kawaoka, 2005). For FPV, a clinical feature that appears to be distinct from CPV is that intrauterine or perinatal infection can lead to infection of cells in the central nervous system (CNS), leading to neurological symptoms and neuropathology (Url *et al*., 2003; Truyen *et al*., 2009). Therefore, the isolation of CPV from the brain of the raccoons, coupled with the detection of antigen within the brain, is an intriguing finding, especially since the infections did not occur *in utero* or perinatally. The relevance of the infection of cells in the CNS to clinical disease and its association or dependence upon the genetic changes observed in the virus is currently unknown. Although FPV has been associated with neuronal infection in neonatal kittens, CPV is normally not associated with CNS infection and antigen has previously never been detected in neurons of CPV-infected dogs (Url and Schmidt, 2005). However, a recent study analyzing the tissue distribution of CPV-2a, -2b, and -2c in dogs disclosed that CPV DNA was detected in all tissues

tested, including considerable amounts in the brain (Decaro *et al*., 2007). The discrepancies between these two studies may be related to the detection methods employed or, less likely, it may be possible that the degree of neurotropism of CPV is dependent upon the infecting strain. For raccoons, infection of the CNS, as demonstrated by immunohistochemistry, was observed in the external germinal layer and in Purkinje cells of the cerebellum (Fig. 6.2). Previously, with FPV, Purkinje cell infection in perinatal kittens has been associated with cerebellar hypoplasia and a cerebellar ataxia syndrome (Url *et al*., 2003). It may also be of interest that CNS vasculitis (e.g., encephalopathy, neurological impairment) and encephalitis in humans caused by parvovirus B19 has only recently been recognized as a pathological complication of infection, suggesting that CNS infection may, in fact, be a normal opportunistic tropism of many parvoviruses (Heegaard *et al*., 1995; Bilge *et al*., 2005). Since the observed neurotropism in the raccoons is limited in its breath (i.e., both the number and cell types involved are few) and that the antigen deposition was not associated with neuropathology, whether CNS infection by these CPV variants is pathological (e.g., induces behavioral changes, alters motor coordination or mental status) is uncertain. It is also unknown if the observed neurotropism is solely a unique property of the genetic changes in these CPV variants or common to other parvoviruses that infect raccoons (i.e., RPV). However, future *in vivo* experimental trials with these CPV variants and RPV may provide additional clues regarding the differential ability of these viruses to infect the CNS and whether these CPV variants are normally neurotropic and/or neuropathogenic.

Whether the antigenic changes that have occurred due to the amino acid substitutions are significant enough to suggest that, if the virus is still infectious to dogs and cats, current vaccines would not be protective, is speculative. However, it is interesting to note that in many of these instances (e.g., see cases 5 and 6 in *Case histories* section), the raccoons had received FPV or CPV vaccines well in advance of the recognition of clinical disease. Although there is the possibility of improper administration of the vaccine, or expired or improper storage of the vaccine leading to its failure, circumstantial evidence suggests that currently-used vaccines may not be completely protective, as morbidity and mortality has been documented in multiple cases even after proactive vaccine administration.

One difference between the original CPV (CPV-2) and the variants which later replaced it (CPV-2a/b) was the ability of the latter to additionally infect cats. The *in vivo* host range of the raccoon CPV isolates are currently unknown; even if the current strains are no longer infectious or replicate at very low levels in dogs, the previous demonstration of the virus to continually adapt and shift host ranges may suggest that the ability of the virus to regain the capacity to infect dogs (if it is lost), once it is significantly changed from an antigenic standpoint, should be considered a possibility. Future avenues of research will entail examining the *in vivo* host range (raccoons, cats, dogs) of the raccoon CPV variants and examining the ability of different viruses (CPV, FPV, raccoon variants) to bind to and infect cells expressing the raccoon TfR. Additionally, examining how well the raccoon variants, or mutants in which the observed amino acids are sequentially introduced into a CPV-2a infectious clone background, infect cells expressing the canine and feline TfR *in vitro* may also help correlate *in vivo*

and *in vitro* results and also clarify the potential epidemiological significance of these viruses to both wildlife and domestic animal populations.

Table 6.1. Comparison of the phylogenetically informative VP2 amino acid positions responsible for the differential host range and antigenicity observed among the carnivore parvoviruses.

	VP2 residue												
Virus	80	87	93	103	224	232	297	300	305	323	426	564	568
RPV	K	M	K	V	G	V	S	A	D	D	D	N	A
FPV	K	M	K	V	G	V	S	A	D	D	N	N	A
MEV	K	M	K	V	G	V	S	V	D	D	N	N	A
CPV-2	R	M	N	A	G	I	S	A	D	N	N	S	G
CPV-2a	R	L	N	A	G	I	S/A	G	Y	N	N	S	G
CPV-2b	R	L	N	A	G	I	S/A	G	Y	N	D	S	G
CPV-2c[a]	R	L	N	A	G	I	A	G	Y	N	E	S	G
CPV-2c(a)[b]	R	L	N	A	G	I	A	D	Y	N	N	S	G
CPV-2c(b)	R	L	N	A	G	I	A	D	Y	N	D	S	G
CC 287-08 (GA)	R	L	N	A	G	I	A	D	H	N	N	S	G
CC 289-08 (GA)	R	L	N	A	G	I	A	D	H	N	N	S	G
CC 349-08 (GA)	R	L	N	A	G	I	A	D	H	N	N	S	G
CC 118-A-09 (VA)	R	L	N	A	R	I	A	D	D	N	N	S	G
CC 278-09 (VA)	R	L	N	A	G	I	A	D	H	N	N	S	G
CC 351-09 (TN)	R	L	N	A	R	I	A	D	D	N	N	S	G
CC 358-B-09 (KY)	R	L	N	A	R	I	A	D	D	N	N	S	G
CC 381-09 (FL)	R	L	N	A	G	I	A	D	H	N	N	S	G
CC 37-10 (WI)	R	L	N	A	G	T	S	D	D	N	N	S	G

a = CPV-2c refers to antigenic variants isolated primarily from dogs containing a glutamic acid substitution at position 426

b = CPV-2c(a) and (b) refers to antigenic variants isolated from leopard cats containing an aspartic acid substitution at position 300.

[RPV: *raccoon parvovirus*; FPV: *feline panleukopenia virus*; MEV: *mink enteritis virus*; CPV: *canine parvovirus*; CC: SCWDS raccoon clinical case isolate].

Table 6.2. Comparison of the nucleotide substitutions occurring among the CPV raccoon isolates from 2007-2009, for which full-length coding sequences were obtained.

Contig position	Consensus codon	Residue position	Amino acid	FL 2009 381	TN 2009 351	KY 2009 358-B	VA 2007 118-A	VA 2009 278-A	GA 2008 287	GA 2008 289	GA 2008 349	Syn. subst.	Nonsyn. subst.	AA change
1	G	*	*			A						*	*	*
40	A	*	*						G	G	G	*	*	*
41	A	*	*						G	G	G	*	*	*
72	G	*	*						A	A	A	*	*	*
103	A	*	*						G	G	G	*	*	*
144	T	*	*		C	C	C					*	*	*
195	GTA	NS1/2-14	V						GTG			V		
228	GCA	25	A			GCG						A		
315	GAA	54	E		GAG							E		
351	GCA	66	A			GCG						A		
360	CAA	69	Q	CAG								Q		
549	GTT	NS1-132	V						GTC	GTC	GTC	V		
879	AAG	242	K		AAA	AAA	AAA					K		
1005	TCA	284	S			TCG						S		
1020	TTG	289	L			TTA						L		
1074	CAA	307	Q	CAG								Q		
1092	GAA	313	E						GAG	GAG	GAG	E		
1201	GAT	**350**	**D**						AAT	AAT	AAT		**N**	**D↔N**
1279	ATT	**376**	**I**	GTT									**V**	**I↔V**
1303	GCA	383	A				GCG					A		
1530	ATT	459	I					ATC	ATC	ATC	ATC	I		
1562	AAA	**470**	**K**								AGA		**R**	**K↔R**
1581	CAA	476	Q	CAG								Q		
1608	ACA	485	T						ACG	ACG	ACG	T		
1627	ATT	**492**	**I**				GTT						**V**	**I↔V**
1629	ATT	492	I						ATC	ATC	ATC	I		
1653	AGA	500	R						AGG	AGG	AGG	R		
1701	TTA	516	L				TTG					L		
1776	AAA	541	K						AAG	AAG	AAG	K		
1784	TAT	**544**	**Y**					TTT	TTT	TTT	TTT		**F**	**Y↔F**
1833	GAA	560	E			GAG						E		
1870	GGT	**573**	**G**	AGT									**S**	**G↔S**
1900	GAG	**583**	**E**					AAG					**K**	**E↔K**
1905	ACA	584	T					ACG				T		
1926	CAG	591	Q			CAA						Q		
1941	GTT	596	V						GTC	GTC	GTC	V		
1942	CTA	597	L					TTA				L		
1957	TCG	**602**	**S**					CCG	CCG	CCG	CCG		**P**	**S↔P**
1980	CTG	609	L	CTA								L		
2079	ACA	642	T					ACG	ACG	ACG	ACG	T		
2103	CTG	650	L						CTA	CTA	CTA			
1900	AGA	**NS2-92**	**R**					AAA					**K**	**R↔K**
1905	ACA	**94**	**T**					GCA					**A**	**T↔A**
1926	GAG	**101**	**E**			AAG							**K**	**E↔K**
1941	TCT	**106**	**S**						CCT	CCT	CCT		**P**	**S↔P**
1942	TCT	**106**	**S**					TTT					**F**	**S↔F**
1957	TTC	**111**	**F**					TCC	TCC	TCC	TCC		**S**	**F↔S**
1980	GGA	**119**	**G**	AGA									**R**	**G↔R**
2079	ATG	**152**	**M**					GTG	GTG	GTG	GTG		**V**	**M↔V**
2103	GAG	**160**	**E**						AAG	AAG	AAG		**K**	**E↔K**
2301	GGG	VP1-21	G					GGA	GGA	GGA	GGA	G		
2394	TCT	52	S		TCC							S		
2454	AAG	72	K	AAA								K		
2532	GAT	98	D					GAC				D		
2585	AAA	**116**	**K**				AGA						**R**	**K↔R**
2643	GTA	**136**	**V**					ATA					**I**	**V↔I**
2787	ATT	VP1/2-40*	I		ATC							I		
2814	ACG	49	T						ACA	ACA	ACA	T		
2847	GAA	60	E						GAG	GAG	GAG	E		
2970	ACC	101	T	ACC								T		
2980	ATT	105	I						ATC	ATC	ATC	I		
3075	TTG	136	L			TTA						L		
3084	GTT	139	V	GTC								V		
3192	GTT	175	V	GTC								V		
3238	ATG	**190**	**M**	ATA									**I**	**M↔I**
3252	TTG	195	L						TTA	TTA	TTA	L		
3337	GGA	**224**	**G**		AGA	AGA	AGA						**R**	**G↔R**
3354	CCA	229	P						CCG			P		
3360	AAT	231	N						AAC	AAC	AAC	N		
3472	GAT	**269**	**D**		AAT								**N**	**D↔N**
3568	GGT	**300**	**G**	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT		**D**	**G↔D**
3580	TAT	**305**	**Y**	CAT	GAT	GAT	GAT	CAT	CAT	CAT	CAT		**H/D**	**Y↔H/D**
3615	GTA	316	V			GTG						V		
3672	GAG	335	E				GAA					E		
3675	GTT	336	V					GTA				V		
3759	GCG	364	A	GCA								A		
3816	CAA	383	Q					CAG				Q		
3955	TTG	430	L	CTG								L		
3957	TTG	430	L						TTA	TTA	TTA	L		
3958	CTA	431	L					TTA				L		
4062	CCA	465	P	CCT								P		
4167	GTA	500	V					GTG				V		
4288	GCC	**541**	**A**		ACC								**T**	**A↔T**
4290	GCC	541	A						GCT	GCT	GCT	A		
4480	A	*	*	G				G	G	G		*	*	*
4576	G	*	*						A	A	A	*	*	*
4584	T	*	*			C						*	*	*
4594	A	*	*					G				*	*	*
4617	G	*	*					A				*	*	*

*Amino acid residue designation based on VP2

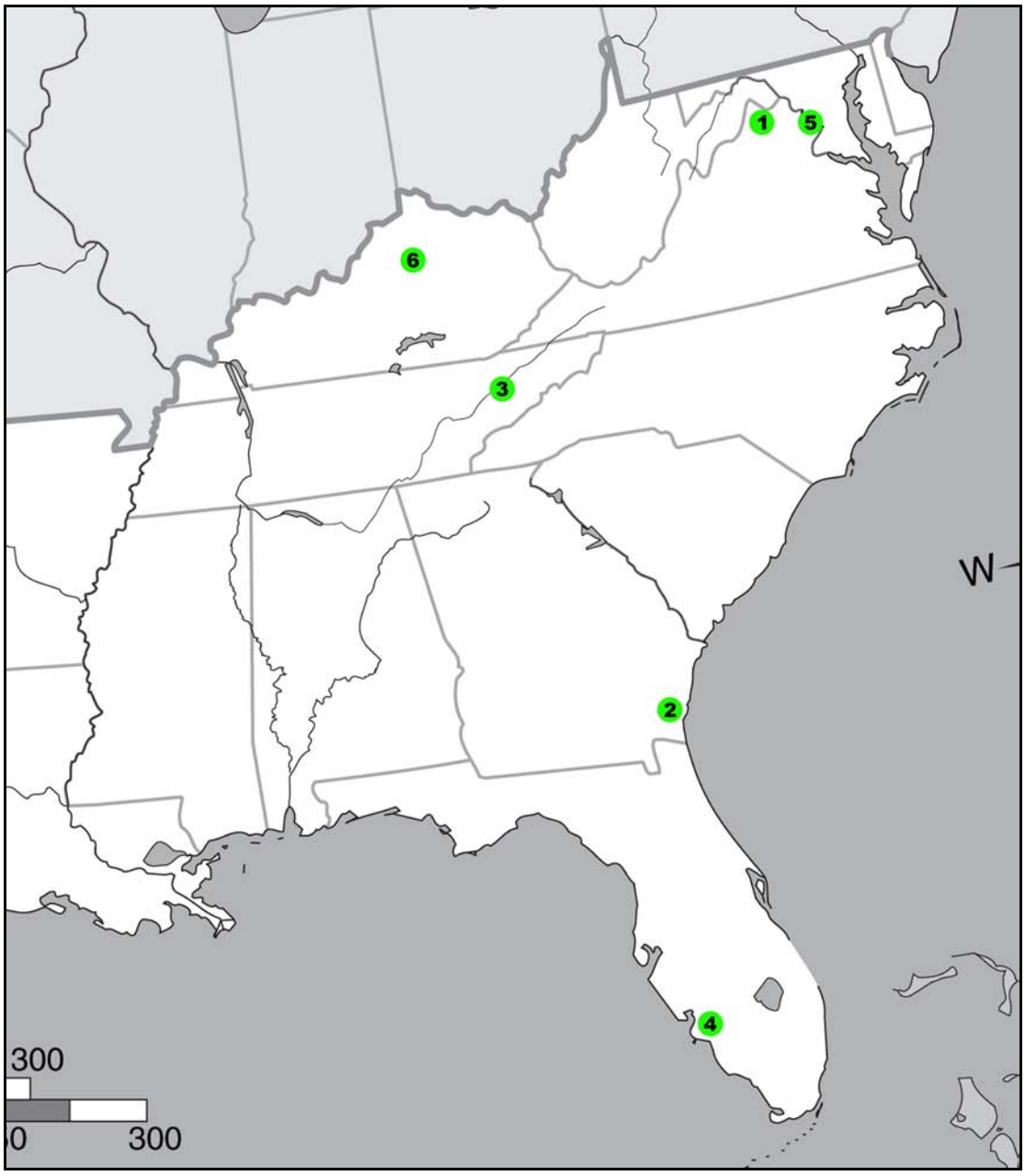


Figure 6.1. Geographical locations of outbreaks in raccoons caused by novel genetic variants of CPV associated with rehabilitation centers in the southeastern United States during 2007-2009. Numerical designations refer to the chronological order of the individual case histories used in comprehensive genetic analyses (see *Case history* section in *Material and methods* for details).

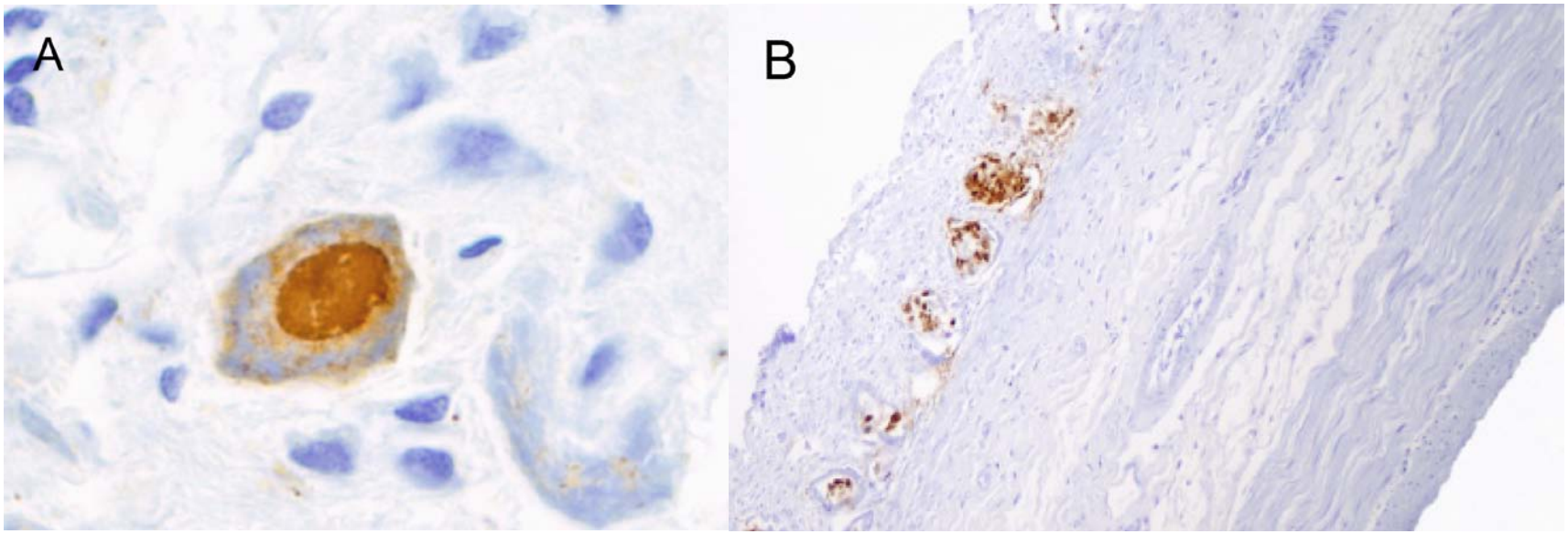


Figure 6.2. Immunohistochemistry of CPV infection in raccoons. Detection of viral antigen in the nucleus of a Purkinje cell in the cerebellum (A; 1000X) and in the intestinal crypts of the small intestines (B; 40X) of a raccoon. Viral antigen was detected with a mouse anti-CPV/FPV monoclonal antibody coupled to a biotinylated goat anti-mouse IgG and a streptavidin-horseradish peroxidase conjugate using 3,3'-diaminobenzidine as the chromogen substrate.

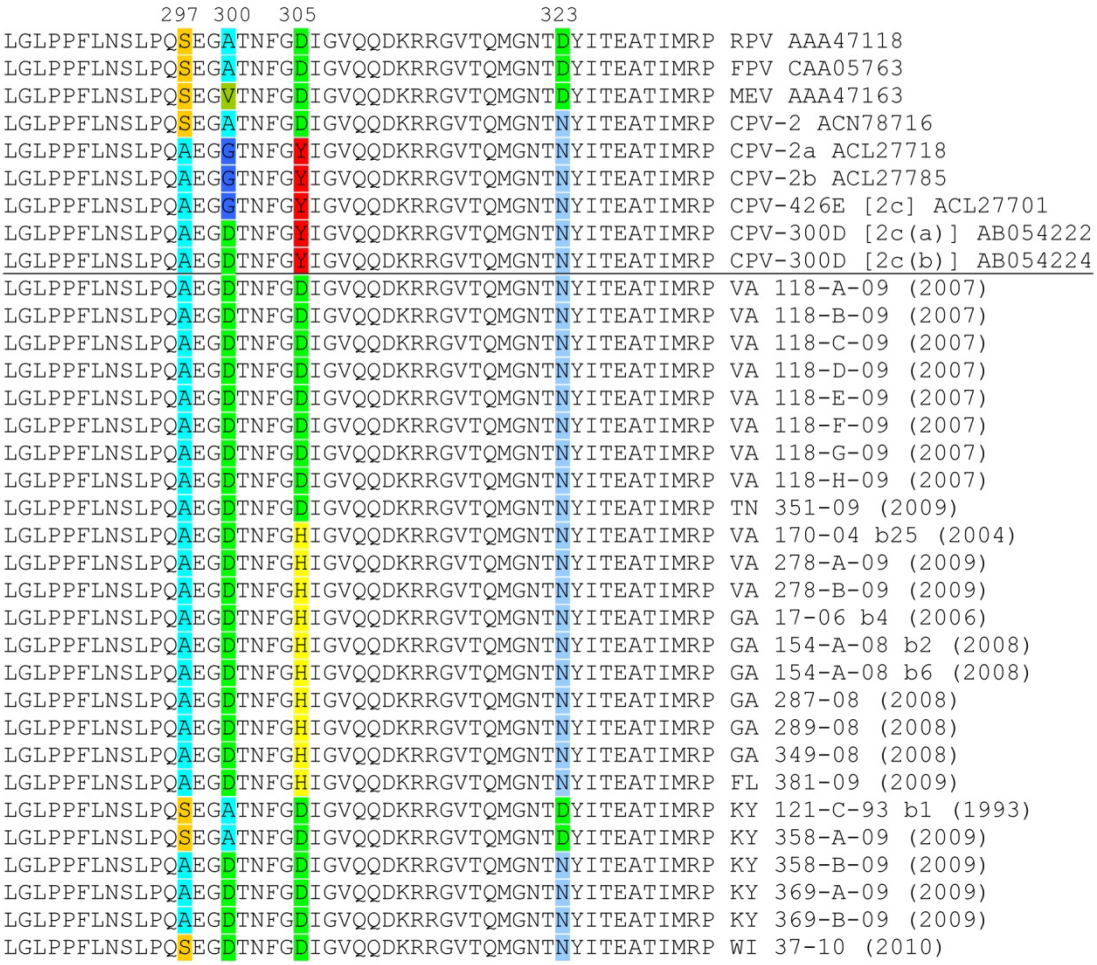


Figure 6.3. VP2 alignment of all CPV raccoon cases identified by either virus isolation or by PCR amplification from FFPE blocks. FFPE samples are indicated by their block number designation (b#). VP2 residues shown indicate sequence coverage of the primer set used in amplifying CPV DNA from FFPE tissues and include the phylogenetically informative amino acids at VP2 positions 297, 300, 305 and 323. Color coding of amino acids correspond to those shown in Table 1. For the carnivore parvoviruses (RPV, FPV, MEV, CPV), GenBank accession numbers are given. For the raccoon CPV isolates, the state of origin, followed by the accession number and year of isolation is shown.

Figure 6.4. Evolutionary relationships of CPV raccoon isolates to other carnivore parvoviruses. The neighbor-joining phylogeny was constructed using a CLUSTALW alignment the nucleotide sequence of VP2 using representative CPV, FPV, MEV, and RPV sequences. Bootstrap values were determined using 2000 replicates and are those that are ≥40% are listed at each node. Branch lengths are drawn to scale. The trees were calculated using the Kimura 2-parameter method and evolutionary distances are represented as the number of nucleotide substitutions per site. Raccoon CPV isolates are indicated with a green circle with each number corresponding to the case history as in Figure X. The single NCBI RPV sequence is highlighted with a red circle. Sequences used in construction of the VP2 tree were: CPV-410.us.00 (EU659119), CPV-411a.us.98 (EU659120), CPV-2b SAH (FJ222822), CPV-431 (AY742951), CPV-2b 42/05-49 (FJ005263), CPV-193 (AY742932), CPV-2c 56/00 (AY380577), CPV-2c G367/97 (FJ005202), CPV-447 (AY742934), CPV-395(AY742936), CPV-411b.us.98 (EU659121), CPV-ZD13 (EU483515), CPV KU4-08 (FJ869124), CPV KU3-08 (FJ869123), CPV-39 (AY787930), CPV-13.us.81 (EU659118), CPV-Y1 (D26079), CPV BR6-80 (DQ340404), CPV BR137-80 (DQ340406), CPV BR135-80 (DQ340405), CPV-2a 80/08 (FJ005258), CPV-435 (AY742953), CPV BR209-94 (DQ340428), CPV BR227-93 (DQ340425), CPV BR8155-00 (DQ340434), CPV-339 (AY742933), CPV-2a 54/08 (FJ005257), CPV-U6 (AY742935), CPV-2a 17/08 (FJ005256), CPV-2a 100/08 (FJ005259), CPV-04S30 (DQ025999), CPV-03C3 (DQ025959), LCPV-V139 (AB054222), LCPV-V140 (AB054223), CPV-5.us.79 (EU659116), CPV-b (M38245), CPV-6.us.80 (EU659117), CPV-128(U22186), MEV-Abashiri (D00765), MEV-B (FJ592174), FPV-kai.us.06 (EU659115), FPV-4.us.64 (EU659112), RPV-M24005 (same), FPV-193/70 (X55115), FPV-3.us.67 (EU659111), FPV-8a.us.89 (EU659113), and FPV-8b.us.89 (EU659114).

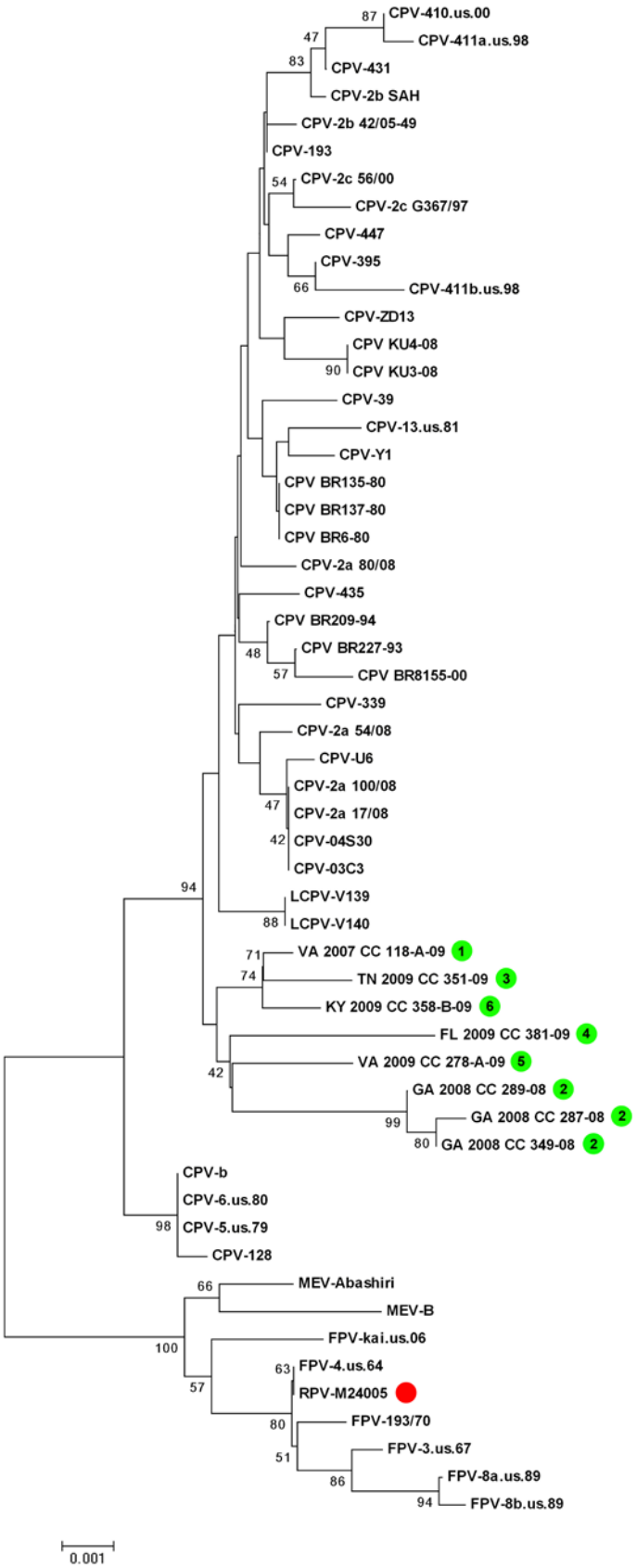
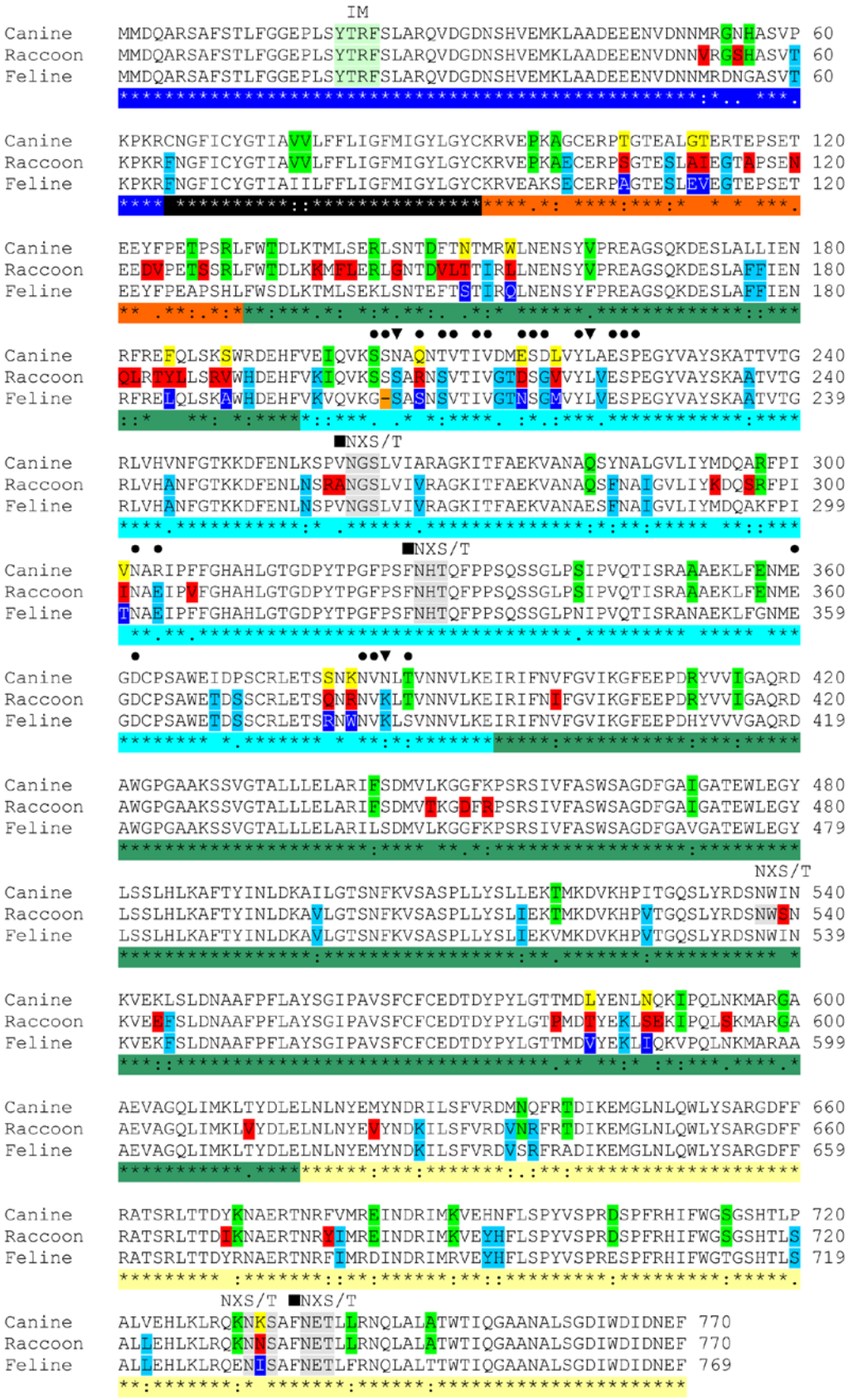

CPV-410.us.00
CPV-411a.us.98
CPV-431
CPV-2b SAH
CPV-2b 42/05-49
CPV-193
CPV-2c 56/00
CPV-2c G367/97
CPV-447
CPV-395
CPV-411b.us.98
CPV-ZD13
CPV KU4-08
CPV KU3-08
CPV-39
CPV-13.us.81
CPV-Y1
CPV BR135-80
CPV BR137-80
CPV BR6-80
CPV-2a 80/08
CPV-435
CPV BR209-94
CPV BR227-93
CPV BR8155-00
CPV-339
CPV-2a 54/08
CPV-U6
CPV-2a 100/08
CPV-2a 17/08
CPV-04S30
CPV-03C3
LCPV-V139
LCPV-V140
VA 2007 CC 118-A-09 1
TN 2009 CC 351-09 3
KY 2009 CC 358-B-09 6
FL 2009 CC 381-09 4
VA 2009 CC 278-A-09 5
GA 2008 CC 289-08 2
GA 2008 CC 287-08 2
GA 2008 CC 349-08 2
CPV-b
CPV-6.us.80
CPV-5.us.79
CPV-128
MEV-Abashiri
MEV-B
FPV-kai.us.06
FPV-4.us.64
RPV-M24005
FPV-193/70
FPV-3.us.67
FPV-8a.us.89
FPV-8b.us.89
0.001

Figure 6.5. Comparative amino acid alignment of the raccoon, canine and feline transferrin receptors (TfR). The TfR can be divided into five domains, which are colored coded underneath the alignment: a cytoplasmic domain (royal blue; residues 1-64), a transmembrane domain (black; residues 65-92), a stalk domain (orange; residues 93-131), a bipartite protease-like domain (sea green; residues 132-196 and 394-616), an apical domain (aqua; residues 197-393), and a helical domain (cream; residues 617-770). Underneath the alignment, asterisks denote identity, while semi-colons, periods, or a blank indicate conserved, semi-conserved, or non-conserved amino acid substitutions between the three TfRs. Within the alignment, residues that are unique to raccoon TfR (i.e., but conserved in both canine and feline TfR) are shown in red. Amino acids which are variable between raccoon, canine, and feline TfR are shown in red, bright yellow, and royal blue, respectively. Amino acids that are conserved between raccoon and canine TfR are shown in bright green, while those conserved in both raccoon and feline TfR are shown in pale blue. The five potential glycosylation sites (NXS/T) of raccoon TfR are shaded in grey, with the three asparagines predicted to be glycosylated indicated with a black square. TfR amino acids in the apical domain mutated in previous CPV and FPV binding studies are indicated above the alignment with a circle (no clear effect on virus binding) or downward arrowhead (effected binding) (see Palermo *et al*., 2003). The TfR internalization motif (IM) is highlighted in light green


IM
Canine MMDQARSAFSTLFGGEPLSYTRFSLARQVDGDNSHVEMKLAADEEENVDNNMRGNHASVP 60
Raccoon MMDQARSAFSTLFGGEPLSYTRFSLARQVDGDNSHVEMKLAADEEENVDNNVRGSHASVT 60
Feline MMDQARSAFSTLFGGEPLSYTRFSLARQVDGDNSHVEMKLAADEEENVDNNMRDNGASVT 60
Canine KPKRCNGFICYGTIAVVLFFLIGFMIGYLGYCKRVEPKAGCERPTGTEALGTERTEPSET 120
Raccoon KPKRFNGFICYGTIAVVLFFLIGFMIGYLGYCKRVEPKAECERPSGTESLAIEGTAPSEN 120
Feline KPKRFNGFICYGTIAIILFFLIGFMIGYLGYCKRVEAKSECERPAGTESLEVEGTEPSET 120
Canine EEYFPETPSRLFWTDLKTMLSERLSNTDFTNTMRWLNENSYVPREAGSQKDESLALLIEN 180
Raccoon EEDVPETSSRLFWTDLKKMFLERLGNTDVLTTIRLLNENSYVPREAGSQKDESLAFFIEN 180
Feline EEYFPEAPSHLFWSDLKTMLSEKLSNTEFTSTIRQLNENSYFPREAGSQKDESLAFFIEN 180
NXS/T
Canine RFREFQLSKSWRDEHFVEIQVKSSNAQNTVTIVDMESDLVYLAESPEGYVAYSKATTVTG 240
Raccoon QLRTYLLSRVWHDEHFVKIQVKSSSARNSVTIVGTDSGVVYLVESPEGYVAYSKAATVTG 240
Feline RFRELQLSKAWHDEHFVKVQVKG-SASNSVTIVGTNSGVVYLVESPEGYVAYSKAATVTG 239
NXS/T
Canine RLVHVNFGTKKDFENLKSPVNGSLVIARAGKITFAEKVANAQSYNALGVLIYMDQARFPI 300
Raccoon RLVHANFGTKKDFENLNSRANGSLVIVRAGKITFAEKVANAQSFNAIGVLIYKDQSRFPI 300
Feline RLVHANFGTKKDFENLNSPVNGSLVIVRAGKITFAEKVANAESFNAIGVLIYMDQAKFPI 299
NXS/T
Canine VNARIPFFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPSIPVQTISRAAAEKLFENME 360
Raccoon INAEIPVFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPSIPVQTISRAAAEKLFENME 360
Feline TNAEIPFFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPNIPVQTISRANAEKLFGNME 359
Canine GDCPSAWEIDPSCRLETSSNKNVNLTVNNVLKEIRIFNVFGVIKGFEEPDRYVVIGAQRD 420
Raccoon GDCPSAWETDSSCRLETSQNRNVKLTVNNVLKEIRIFNIFGVIKGFEEPDRYVVIGAQRD 420
Feline GDCPSAWETDSSCRLETSRNWNVKLSVNNVLKEIRIFNVFGVIKGFEEPDHYVVVGAQRD 419
Canine AWGPGAAKSSVGTALLLELARIFSDMVLKGGFKPSRSIVFASWSAGDFGAIGATEWLEGY 480
Raccoon AWGPGAAKSSVGTALLLELARIFSDMVTKGDFRPSRSIVFASWSAGDFGAIGATEWLEGY 480
Feline AWGPGAAKSSVGTALLLELARILSDMVLKGGFKPSRSIVFASWSAGDFGAVGATEWLEGY 479
NXS/T
Canine LSSLHLKAFTYINLDKAILGTSNFKVSASPLLYSLLEKTMKDVKHPITGQSLYRDSNWIN 540
Raccoon LSSLHLKAFTYINLDKAVLGTSNFKVSASPLLYSLIEKTMKDVKHPVTGQSLYRDSNWSN 540
Feline LSSLHLKAFTYINLDKAVLGTSNFKVSASPLLYSLIEKVMKDVKHPVTGQSLYRDSNWIN 539
Canine KVEKLSLDNAAFPFLAYSGIPAVSFCFCEDTDYPYLGTTMDLYENLNQKIPQLNKMARGA 600
Raccoon KVEEFSLDNAAFPFLAYSGIPAVSFCFCEDTDYPYLGTPMDTYEKLSEKIPQLSKMARGA 600
Feline KVEKFSLDNAAFPFLAYSGIPAVSFCFCEDTDYPYLGTTMDVYEKLIQKVPQLNKMARAA 599
Canine AEVAGQLIMKLTYDLELNLNYEMYNDRILSFVRDMNQFRTDIKEMGLNLQWLYSARGDFF 660
Raccoon AEVAGQLIMKLVYDLELNLNYEVYNDKILSFVRDVNRFRTDIKEMGLNLQWLYSARGDFF 660
Feline AEVAGQLIMKLTYDLELNLNYEMYNDKILSFVRDVSRFRADIKEMGLNLQWLYSARGDFF 659
Canine RATSRLTTDYKNAERTNRFVMREINDRIMKVEHNFLSPYVSPRDSPFRHIFWGSGSHTLP 720
Raccoon RATSRLTTDIKNAERTNRYIMREINDRIMKVEYHFLSPYVSPRDSPFRHIFWGSGSHTLS 720
Feline RATSRLTTDYRNAERTNRFIMRDINDRIMRVEYHFLSPYVSPRESPFRHIFWGTGSHTLS 719
NXS/T NXS/T
Canine ALVEHLKLRQKNKSAFNETLLRNQLALATWTIQGAANALSGDIWDIDNEF 770
Raccoon ALLEHLKLRQKNNSAFNETLLRNQLALATWTIQGAANALSGDIWDIDNEF 770
Feline ALLEHLKLRQENISAFNETLFRNQLALTTWTIQGAANALSGDIWDIDNEF 769

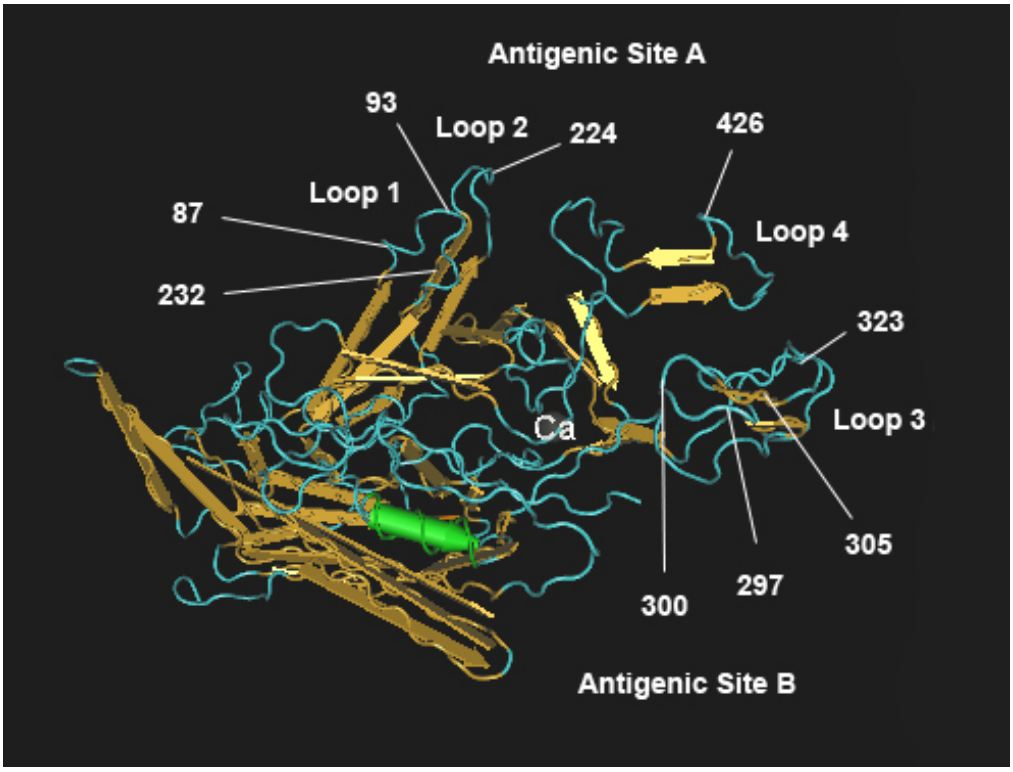


Figure 6.6. Ribbon diagram of the crystal structure of the VP2 protein of CPV showing amino acid substitutions of importance in defining the host range of the raccoon CPV isolates [PDB entry 1C8D]. The calcium-binding site (Ca) of CPV is also highlighted (see Simpson *et al*., 2000).

Acknowledgements

Funds were provided through sponsorship from the fish and wildlife agencies of Alabama, Arkansas, Florida, Georgia, Kansas, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Puerto Rico, South Carolina, Tennessee, Virginia, and West Virginia; through the Federal Aid to Wildlife Restoration Act (50 Stat. 917) and Grant Agreement 06ERAG0005, Biological Resources Division, U.S. Geological Survey, U.S. Department of the Interior; and through Cooperative Agreements 0596130032CA and 0696130032CA, Veterinary Services, Animal and Plant Health Inspection Service, U.S. Department of Agriculture.

References

Appel, M. J., Parrish, C. R., 1982. Raccoons are not susceptible to canine parvovirus. J. Am. Vet. Med. Assoc. 181, 489.

Barker, I. K., Povey, R. C., Voigt, D. R., 1983. Response of mink, skunk, red fox and raccoon to inoculation with mink virus enteritis, feline panleukopenia and canine parvovirus and prevalence of antibody to parvovirus in wild carnivores in Ontario. Can. J. Comp. Med. 47, 188-197.

Barker, I. K., Parrish, C. R., 2001. Parvovirus infections, p. 131-146 *In* Williams, E. S., Barker, I. K. (ed.), Infectious Diseases of Wild Mammals, Iowa State University Press, Ames, IA.

Bilge, I., Sadikoğlu, B., Emre, S., Sirin, A., Aydin, K., Tatli, B., 2005. Central nervous system vasculitis secondary to parvovirus B19 infection in a pediatric renal transplant patient. Pediatr. Nephrol. 20, 529-533.

Bouillant, A., Hanson, R. P., 1965. Epizootiology of mink enteritis: I. Stability of the virus in feces exposed to natural environmental factors. Can. J. Comp. Med. Vet. Sci. 29, 125-128.

Chang, S. F., Sgro, J. Y., Parrish, C. R., 1992. Multiple amino acids in the capsid structure of canine parvovirus coordinately determine the canine host range and specific antigenic and hemagglutination properties. J. Virol. 66, 6858-6867.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., Thompson, J.D., 2003. Multiple sequence alignment with the CLUSTAL series of programs. Nucleic Acids Res. 31, 3497-3500.

Decaro, N., Martella, V., Elia, G., Desario, C., Campolo, M., Lorusso, E., Colaianni, M. L., Lorusso, A., Buonavoglia, C., 2007. Tissue distribution of the antigenic variants of canine parvovirus type 2 in dogs. Vet. Microbiol. 121, 39-44.

Decaro, N., Desario, C., Elia, G., Martella, V., Mari, V., Lavazza, A., Nardi, M., Buonavoglia, C., 2008. Evidence for immunisation failure in vaccinated adult dogs infected with canine parvovirus type 2c. New Microbiol. 31, 125-130.

Decaro, N., Desario, C., Parisi, A., Martella, V., Lorusso, A., Miccolupo, A., Mari, V., Colaianni, M. L., Cavalli, A., Di Trani, L., Buonavoglia, C., 2009. Genetic analysis of canine parvovirus type 2c. Virology 385, 5-10.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O. Peeters, M., Shaw, G. M., Sharp, P. M., Hahn, B. H., 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 397, 436-441.

Heegaard, E. D., Peterslund, N. A., Hornsleth, A., 1995. Parvovirus B19 infection associated with encephalitis in a patient suffering from malignant lymphoma. Scand. J. Infect. Dis. 27, 631-633.

Hoelzer, K., Shackelton, L. A., Parrish, C. R., Holmes, E. C., 2008. Phylogenetic analysis reveals the emergence, evolution and dispersal of carnivore parvoviruses. J. Gen. Virol. 89, 2280-2289.

Hueffer, K., Parker, J. S., Weichert, W. S., Geisel, R. E., Sgro, J. Y., Parrish, C. R., 2003. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. J. Virol. 77, 1718-1726.

Hueffer, K., Parrish, C. R., 2003. Parvovirus host range, cell tropism and evolution. Curr. Opin. Microbiol. 6, 392-398.

Ikeda, Y., Mochizuki, M., Naito, R., Nakamura, K., Miyazawa, T., Mikami, T., Takahashi, E., 2000. Predominance of canine parvovirus (CPV) in unvaccinated cat populations and emergence of new antigenic types of CPVs in cats. Virology 278, 13-19.

Ikeda, Y., Nakamura, K., Miyazawa, T., Takahashi, E., Mochizuki, M., 2002. Feline host range of canine parvovirus: recent emergence of new antigenic types in cats. Emerg. Infect. Dis. 8, 341-346.

Lawrence, C. M., Ray, S., Babyonyshev, M., Galluser, R., Borhani, D. W., Harrison, S. C., 1999. Crystal structure of the ectodomain of human transferrin receptor. Science 286, 779-82.

Llamas-Saiz, A. L., Agbandje-McKenna, M., Parker, J. S., Wahid, A.T., Parrish, C. R., Rossmann, M. G., 1996. Structural analysis of a mutation in canine parvovirus which controls antigenicity and host range. Virology 225, 65-71.

Nakamura, K., Sakamoto, M., Ikeda, Y., Sato, E., Kawakami, K., Miyazawa, T., Tohya, Y., Takahashi, E., Mikami, T., Mochizuki, M., 2001. Pathogenic potential of canine parvovirus types 2a and 2c in domestic cats. Clin. Diagn. Lab. Immunol. 3, 663-668.

Nettles, V. F., Pearson, J. E., Gustafson, G. A., Blue, J. L., 1980. Parvovirus infection in translocated raccoons. J. Am. Vet. Med. Assoc. 177, 787-789.

Nowak, R. M. (ed.), 1999. Carnivora: Family Procyonidae, p. 694-704. *In* Walker's Mammals of the World, Volume 1, Sixth ed. John Hopkins University Press, Baltimore, Maryland.

Oie, K. L., Durrant, G., Wolfinbarger, J. B., Martin, D., Costello, F., Perryman, S., Hogan, D., Hadlow, W. J., Bloom, M. E., 1996. The relationship between capsid protein (VP2) sequence and pathogenicity of Aleutian mink disease parvovirus (ADV): a possible role for raccoons in the transmission of ADV infections. J. Virol. 70, 852-861.

Palermo, L. M., Hueffer, K., Parrish, C. R., 2003. Residues in the apical domain of the feline and canine transferrin receptors control host-specific binding and cell infection of canine and feline parvoviruses. J. Virol. 77, 8915-8923.

Palermo, L. M., Hafenstein, S. L., Parrish, C. R., 2006. Purified feline and canine transferrin receptors reveal complex interactions with the capsids of canine and feline parvoviruses that correspond to their host ranges. J. Virol. 80, 8482-8492.

Parker, J. S., Murphy, W. J., Wang, D., O'Brien, S. J., Parrish, C. R., 2001. Canine and feline parvoviruses can use human or feline transferrin receptors to bind, enter, and infect cells. J. Virol. 75, 3896-3902.

Parker, J. S., Parrish, C. R., 1997. Canine parvovirus host range is determined by the specific conformation of an additional region of the capsid. J. Virol. 71, 9214-9222.

Parrish, C. R., Carmichael, L. E., 1986. Characterization and recombination mapping of an antigenic and host range mutation of canine parvovirus. Virology 148, 121-132.

Parrish, C. R., Leathers, C. W., Pearson, R., Gorham, J. R., 1987. Comparisons of feline panleukopenia virus, canine parvovirus, raccoon parvovirus, and mink enteritis virus and their pathogenicity for mink and ferrets. Am. J. Vet. Res. 48, 1429-1435.

Parrish, C. R., Aquadro, C. F., Carmichael, L. E., 1988a. Canine host range and a specific epitope map along with variant sequences in the capsid protein gene of canine parvovirus and related feline, mink, and raccoon parvoviruses. Virology 166, 293-307.

Parrish, C. R., Have, P., Foreyt, W. J., Evermann, J. F., Senda, M., Carmichael, L. E., 1988b. The global spread and replacement of canine parvovirus strains. J. Gen. Virol. 69, 1111-1116.

Parrish, C. R., 1999. Host range relationships and the evolution of canine parvovirus. Vet. Microbiol. 69, 29-40.

Parrish, C. R., Kawaoka, Y., 2005. The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. Annu. Rev. Microbiol. 59, 553-586.

Simpson, A. A., Chandrasekar, V., Hébert, B., Sullivan, G. M., Rossmann, M. G., Parrish, C. R., 2000. Host range and variability of calcium binding by surface loops in the capsids of canine and feline parvoviruses J. Mol. Biol. 300, 597-610.

Shackelton, L. A., Parrish, C. R., Truyen, U., Holmes, E. C., 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc. Natl. Acad. Sci. U. S. A. 102, 379-384.

Steinel, A., Parrish, C. R., Bloom, M. E., Truyen, U., 2001. Parvovirus infections in wild carnivores. J. Wildl. Dis. 37, 594-607.

Strassheim, M. L., Gruenberg, A., Veijalainen, P., Sgro, J. Y., Parrish, C. R., 1994. Two dominant neutralizing antigenic determinants of canine parvovirus are found on the threefold spike of the virus capsid. Virology 198, 175-184.

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

Truyen, U., Parrish, C. R., 1992. Canine and feline host ranges of canine parvovirus and feline panleukopenia virus: distinct host cell tropisms of each virus in vitro and in vivo. J. Virol. 66, 5399-5408.

Truyen, U., Gruenberg, A., Chang, S. F., Obermaier, B., Veijalainen, P., Parrish, C. R., 1995. Evolution of the feline-subgroup parvoviruses and the control of canine host range in vivo. J. Virol. 69, 4702-4710.

Truyen, U., Evermann, J. F., Vieler, E., Parrish, C. R., 1996. Evolution of canine parvovirus involved loss and gain of feline host range. Virology 215, 186-189.

Truyen, U., 1999. Emergence and recent evolution of canine parvovirus. Vet. Microbiol. 69, 47-50.

Truyen, U., 2006. Evolution of canine parvovirus - a need for new vaccines? Vet. Microbiol. 117, 9-13.

Truyen, U., Addie, D., Belák, S., Boucraut-Baralon, C., Egberink, H., Frymus, T., Gruffydd-Jones, T., Hartmann, K., Hosie, M. J., Lloret, A., Lutz, H., Marsilio, F., Pennisi, M. G., Radford, A. D., Thiry, E., Horzinek, M. C., 2009. Feline panleukopenia. ABCD guidelines on prevention and management. J. Feline Med. Surg. 11, 538-446.

Url, A., Truyen, U., Rebel-Bauder, B., Weissenböck, H., Schmidt, P., 2003. Evidence of parvovirus replication in cerebral neurons of cats. J Clin Microbiol. 41, 3801-3805.

Url, A., Schmidt, P., 2005. Do canine parvoviruses affect canine neurons? An immunohistochemical study. Res. Vet. Sci. 79, 57-59.

Webby, R., Hoffmann, E., Webster, R., 2004. Molecular constraints to interspecies transmission of viral pathogens. Nat. Med.10, S77-81.

CHAPTER 7

PERSPECTIVES

Understanding the mechanisms by which viruses evolve in nature has been the subject of intense research for many years, particularly in response to the sudden emergence of variants of increased pathogenicity. A fundamental challenge arising during such episodes of 'viral emergence' is to determine if the recognized increase in the incidence or severity of disease is directly correlated with a genetic change in the virus, and if so, how does that genetic change alter biological characteristics such as transmissibility or virulence. In recent history, the single most devastating viral outbreak was the 1918-1919 H1N1 pandemic of (avian) *influenza A virus* (AIV). As an estimated 20-50 million people were killed by this virus worldwide, a great deal of research has gone into understanding why this particular strain was so lethal, from reconstruction of the virus by reverse genetics, to manipulating the reassortment configuration of the virus in aims of identifying the genetic determinants of pathogenicity and transmissibility (Taubenberger *et al*., 2005, Tumpey *et al*., 2005, Pappas *et al*., 2008). Although whether the 1918 H1N1 virus was or was not a reassortant between avian and human viruses is a matter of debate (Taubenberger *et al*., 2005; Gibbs and Gibbs, 2006), recent genetic studies have uncovered some other surprising findings. In 2001, even though the genetics of AIV had been studied extensively for decades, a novel 'hidden' protein was

discovered (PB1-F2) that was encoded within an overlapping reading frame of PB1 (Chen *et al*., 2001). Subsequently, PB1-F2 was shown to induce apoptosis in immune cells and was suggested to be an important virulence factor, such that it may have played a role in determining the highly pathogenic phenotype of the 1918 virus (Cononello *et al*., 2007). The overprinted nature of PB1-F2 is similar to the C gene of *Durham virus* (DURV) reported in this research, whose sequence is contained within the P gene, such that it appears 'hidden' from conventional view (Fig. 7.1). Although the functional role of the DURV C protein is currently unknown, it was shown to exhibit identity and homology to the G protein of *Vesicular stomatitis Indiana virus* (VSINV), which could possibly suggest that the rhabdovirus G protein, which is the protein by which all rhabdoviruses infect cells, could have arisen from the C protein (Fig. 7.1).

Although to suggest that the rhabdovirus G protein arose from the duplication of the C protein either by intra- or inter-molecular recombination in an ancestral rhabdovirus goes against conventional wisdom, it provides one theory (i.e., in conjuction with chance or convergent evolution due to functional overlap) by which the DURV C protein could possibly share identity and homology to the VSINV G protein. A hypothetical scenario of the evolution of the G protein and how the C protein arose to form the ancestral G protein would be as follows: 1) the P and C genes were overlapping genes in the same ancestral rhabdovirus, a virus which also contained an SH gene, 2) the ancestral role of the derived C protein was that of a fusion protein, while the SH protein was the ancestral attachment protein - thus, the SH and C proteins functioned synergistically to mediate cell attachment and membrane fusion, respectively, in the ancestral rhabdovirus, 3) since the C and P genes were overlapping, they could not optimally adapt individually under

the constraints of overprinting, and thus in order to alleviate this constraint, duplication of the overlapping gene (P/C) occurred by template jumping within the same virus or between related viruses, 5) consequential to the duplication of the P/C gene sequence was the optimal adaptation and functional divergence of the two genes in absense of such constraints, with the original P/C gene evolved towards conserving the P gene (i.e., making the C gene dispensable), while the duplicated P/C gene evolved towards the function of the C protein, a fusion protein, 6) without the constraints of selection acting on the overprinted sequence, the C protein continued to evolve and diversify into a fusion and attachment protein, making the SH protein obsolete, 7) the C protein subsequently evolved to become the present-day G protein (a fusion/attachment protein), while the former attachment protein, the SH protein, became vestigial, as witnessed in present-day DURV and *Tupaia rhabdovirus* (TUPV).

The fact that the C protein appears to be evolving towards obsolescence in the rhabdoviruses, suggests that its function is no longer needed. The fact that the C protein is still visible in only the vesiculoviruses and not other genera within the family may suggest that, at least in the vesiculoviruses, the remnant C protein (after duplication) may have still provided some additional function other than fusogenic activity. The fact that the DURV C protein shows identity to the VSINV G protein, and not its own G protein, could suggest that the G protein of VSINV has diverged less from the ancestral C protein than that of DURV, possibly implying that VSINV circulates in a cycle most similar to the ancestral rhabdovirus from which the C protein originated. Currently, the only two rhabdoviruses that have been shown to encode both C and SH proteins are DURV and TUPV; however, based on L and G phylogenies (see Chapter 5), other related viruses

such as *Klamath virus* (KLAV), *Sandjimba virus* (SJAV), and *Kolongo virus* (KLOV), likely also encode these two proteins, although this has yet to be demonstrated. Sequence analysis of these and other related rhabdoviruses will likely help support or refute the claim of the C protein as the progenitor of the rhabdovirus G protein.

Analysis of the other viruses presented in this research also divulged a number of interesting attributes regarding how these genetic mechanisms can influence virus evolution and what makes a new virus with altered genetics not only viable, but be able to persist and become established as a new viral agent in nature. For example, genomic analysis of the reassortant EHDV indicated a number of informative aspects about its configuration: 1) the two proteins that formed the outer capsid of the virus (VP2 and VP5) were derived from the exotic EHDV-6, while 2) the major internal protein (VP7) involved in structurally interacting with those two proteins was derived from EHDV-2 (Fig. 7.2). Comparison of the VP7 proteins of the parental viruses disclosed they were almost identical to one another (much more so than other serotypes), suggesting that the structural viability of the virus (the ability of VP7 from one virus to interact properly with VP2-VP5 from another) was mediated by this compatibility. If these interactions were not compatible, the virus would likely never have become established in the U. S. Additionally, protein analysis suggested that both of the outer capsid proteins (VP2 and VP5) most likely had to be retained from a single parent, as viruses in which the two outer capsid proteins are derived from different serotypes are structurally unstable and replicate poorly, and thus would be unlikely to persist in nature (Mertens, 1999). Moreover, as VP7 in the reassortant was retained from the endemic EHDV-2, and that VP7 is the receptor protein used to infect *Culicoides sonorensis* (the primary vector of

EHDV in the U. S.), this likely afforded the virus the ability to be transmitted just as efficiently as EHDV-2 in the vector population, while at the same time being antigenically novel to vertebrate hosts (i.e., new outer capsid proteins). In other words, a number of attributes regarding the precise reassortment configuration likely facilitated its survival and presumptive establishment in nature [(i.e., based on four consecutive years of isolation, the virus is likely now endemic in the U. S. (Fig. 7.2)]. Although the specific amino acids that are involved in these structural protein interactions are unknown, it will be of interest to compare the nucleotide and amino acid sequences of all three proteins (VP2, VP5, and VP7) of the reassortant from consecutive years in which it was isolated, as this may disclose certain mutations that may have occurred over time in order to compensate for any structural differences.

The fact that the reassortment of the gene segment encoding a protein involved in antigenicity and/or host cell binding in different virus families (see *Literature review* section) has been shown to lead to an increase in pathogenicity and host range is of direct interest to the EHDV reassortant in this research, as VP2 and VP5 are responsible for antigenicity and mammalian receptor binding. It is also of interest that pathogenesis studies using the Australian prototype strain of EHDV-6 (CSIRO 753) [which, retrospectively, was shown to share a 98% nucleotide identity to the EHDV reassortant in both VP2 and VP5 sequences] demonstrated that it was not pathogenic in cattle (St. George *et al*., 1983), which may suggest that if a phenotype of a particular EHDV strain can be transferred solely by reassortment of the outer capsid protein(s), then the EHDV reassortant may have a similar pathogenicity phenotype (in cattle) of the Australian strain of EHDV-6. However, although reassortment of the outer capsid undoubtedly alters the

antigenicity of the virus, it is unlikely that VP2 and VP5 are the sole genetic determinants in EHDV pathogenicity, and extrapolation of results from one system to another should be interpreted with caution.

Currently, the ruminant host range of different EHDV serotypes has not been studied, but as exotic serotypes may exist in transmission cycles involving many different species of ungulates (e.g., endemic African serotypes), it is possible that different serotypes may have differential abilities to infect certain ruminants. Whether the acquirement of the surface proteins of an exotic serotype into an endemic backbone may potentially allow such a reassortant to infect and cause disease in ruminant species (e.g., those in zoological collections) not previously (or less) susceptible to endemic strains is unknown. The ability of the novel EHDV reassortant reported in this research to infect and cause disease in an endemic wild ungulate species (i.e., white-tailed deer) has been circumstantially demonstrated in free-ranging animals (i.e., isolation of the virus from animals exhibiting hemorrhagic clinical signs) and also reproduced experimentally by our laboratory. Viremia titers and clinical disease in deer infected with the reassortant were comparable to previous experiments with endemic serotypes, suggesting that the exchange of the surface proteins responsible for mammalian cell infection did not have any deleterious effects on the ability of the virus to replicate and cause disease in a North American ungulate species. The long-term ramifications of the introduction of the reassortant into the United States, in terms of the economic impact to the agricultural or private ruminant industries, whether negligible or significant, remain to be determined.

If one of the viruses described in this research was a novel recombinant derived from two known North American flaviviruses, e.g., *West Nile virus* (WNV) and *St. Louis*

encephalitis virus (SLEV), what would be the potential epidemiological consequences of such a recombination event? Would one be able to predict that this recombinant flavivirus might diversify into three novel flaviviruses with distinct ecological niches and geographical ranges that stretch from Canada to South America? Or that one of these flaviviruses will become a widespread pathogen that will be responsible for large-scale human epidemics, while the other two viruses will not cause a single documented case of human disease? It is unlikely that anyone could predict such a scenario of events. As for the two alphaviruses presented in this research, *Highlands J virus* (HJV) and *Fort Morgan virus* (FMV), the future is essentially the present, as the recombination event leading to the emergence of these two viruses, along with WEEV, is estimated to have occurred 1,600 years ago (Weaver *et al*., 1997). As insinuated with the hypothetical flaviviruses, the ancestral recombinant alphavirus of the past has evolved into three genetically and antigenically distinct viruses. Although all three viruses circulate under the same generalized ecological cycle involving invertebrates as vectors and passerine birds as amplifying hosts, each virus has specialized to using a specific enzootic vector, which has allowed each to adapt to a specific ecological niche and has likely fostered their genetic divergence.

As detailed in this research, a single recombination event (or in theory, two cross-overs) led to the establishment of three new alphavirus species. From a geographical standpoint, these three viruses collectively stretch from Canada, across the entire continental United States, and south to Argentina. Therefore, this single recombination event has been highly successful not only in terms of the genetic diversity that has arisen from the event, but also in term of the geographical expansion and ecological territory

that has now been occupied by new viruses. In a similar regard to the sequence and structural similarities recognized between the parental viruses of the EHDV reassortant, sequence and secondary structure analysis of all the recombinant alphaviruses (HJV, FMV, WEEV) and their parental viruses (EEEV, SINV) disclosed that there were a number of interesting similar sequence motifs in the parental viruses that may have contributed to the initial survival and replication of the ancestral recombinant (see Chapter 3). Almost invariably, viruses derived from recombination from two viruses that are genetically divergent cannot survive in nature due to structural incompatibilities (see Strauss and Strauss, 1997); however, in the case of EEEV and SINV, an extraordinary set of circumstances (i.e., two cross-over events) (Fig. 7.3) (Hahn *et al*., 1988), along with unique conserved sequence elements in the 3' UTR (found only between EEEV and SINV), likely led to the survival of the ancestral virus. This ancestral virus then diversified into three novel alphavirus species (HJV, FMV, WEEV) that have differential levels of importance in terms of their impacts on human health or wild and domestic animal populations (i.e., unlike WEEV, neither HJV nor FMV are normally associated with disease in humans or other mammals, although all three viruses are avian pathogens to varying degrees) (Waters *et al*., 1976; Ficken *et al*., 1993; Randolph *et al*., 1994; Huyvaert *et al*., 2008).

The final example of viral evolution in this research was the detection and characterization of novel genetic variants of *canine parvovirus* (CPV) isolated from raccoons (*Procyon lotor*) (Fig. 7.4). CPV emerged as a recognized pathogen in 1978, apparently as a host range variant of *feline panleukopenia virus* (FPV) (Parrish, 1999). This species jump of a feline virus into dogs was facilitated by only a few amino acid

changes in the VP2 capsid protein, which is the protein responsible for binding to the transferrin receptor of the host, thereby mediating cellular entry of the virus (Parker *et al*., 2001). Within two years of its emergence, CPV spread globally, initiating a pandemic of hemorrhagic gastroenteritis in dogs (Parrish *et al*., 1988). Although spillover of viruses into aberrant species is a common occurrence, a species jump followed by the adaptation and establishment of a novel virus in a new host is a rare event. The CPV isolates obtained from raccoons in the last few years in the southeastern U. S. represent such an event. Previously, raccoons have been demonstrated not to be susceptible to CPV (Appel and Parrish, 1982). How the emergence of this novel pathogen will affect raccoon populations or other, less abundant, procyonid species in the U. S. [e.g., ringtails (*Bassariscus astutus*)] is currently unknown. Nor is it known if this variant will spread southward with *P. lotor* to Central and South America, which has an abundance of procyonid species. Additionally, whether these novel variants still have the ability to infect dogs or cats *in vivo* remains to be determined. If the host range of these variants is wide (e.g., procyonids, cats, dogs), and if they are antigenically divergent to the point of vaccine failure, the impacts of this species jump of CPV could be far-reaching.

The normal epidemiological consequences of these evolutionary mechanisms (recombination, reassortment, overprinting, and mutation) are for the most part either neutral or debilitating to the virus, such that most variants are functionally or structurally unstable and are removed from circulation by negative selection and, hence, are never observed. However, in certain cases, as shown here, they can lead to a virus that has a new-found ability to infect a novel host or vector, subvert the immune response, or be transmitted more efficiently, leading to the evolution of a new pathogen.

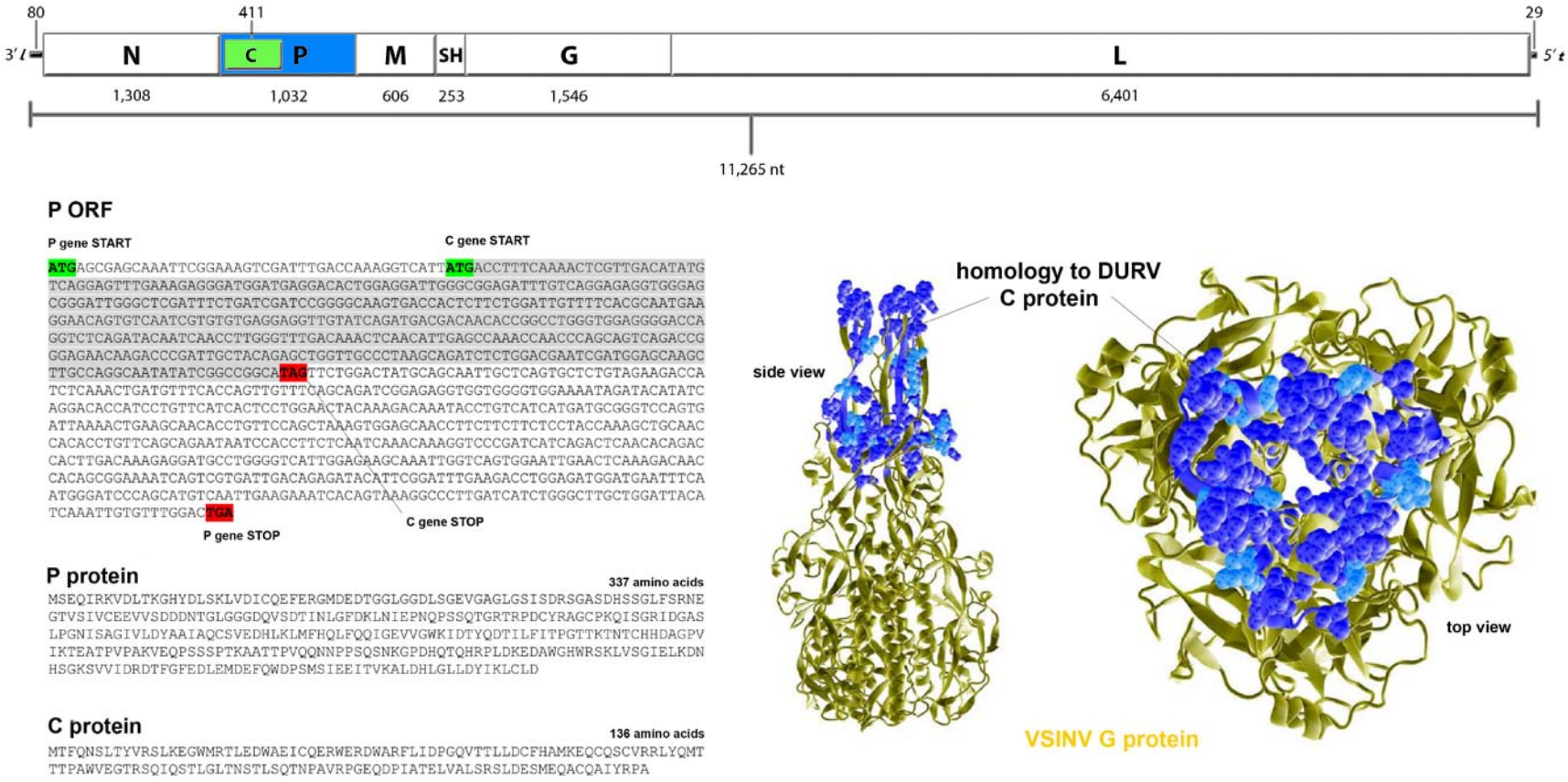


Figure 7.1. Overprinting in rhabdoviruses. (Upper): Schematic organization of the genome of *Durham virus* (DURV). The phosphoprotein (P) gene is highlighted in blue, while the C protein gene contained within the P gene is shown in green. (Lower left): Nucleotide sequence of the P open reading frame (ORF) and the amino acid sequence of the two proteins (P and C) encoded within the P ORF. Start and stop codons for the P and C genes are noted. The C gene sequence is shaded in gray. (Lower right): Identity/homology of the DURV C protein with the fusion domain of the G protein of *Vesicular stomatitis Indiana virus* (VSINV). See chapter 5 for details.

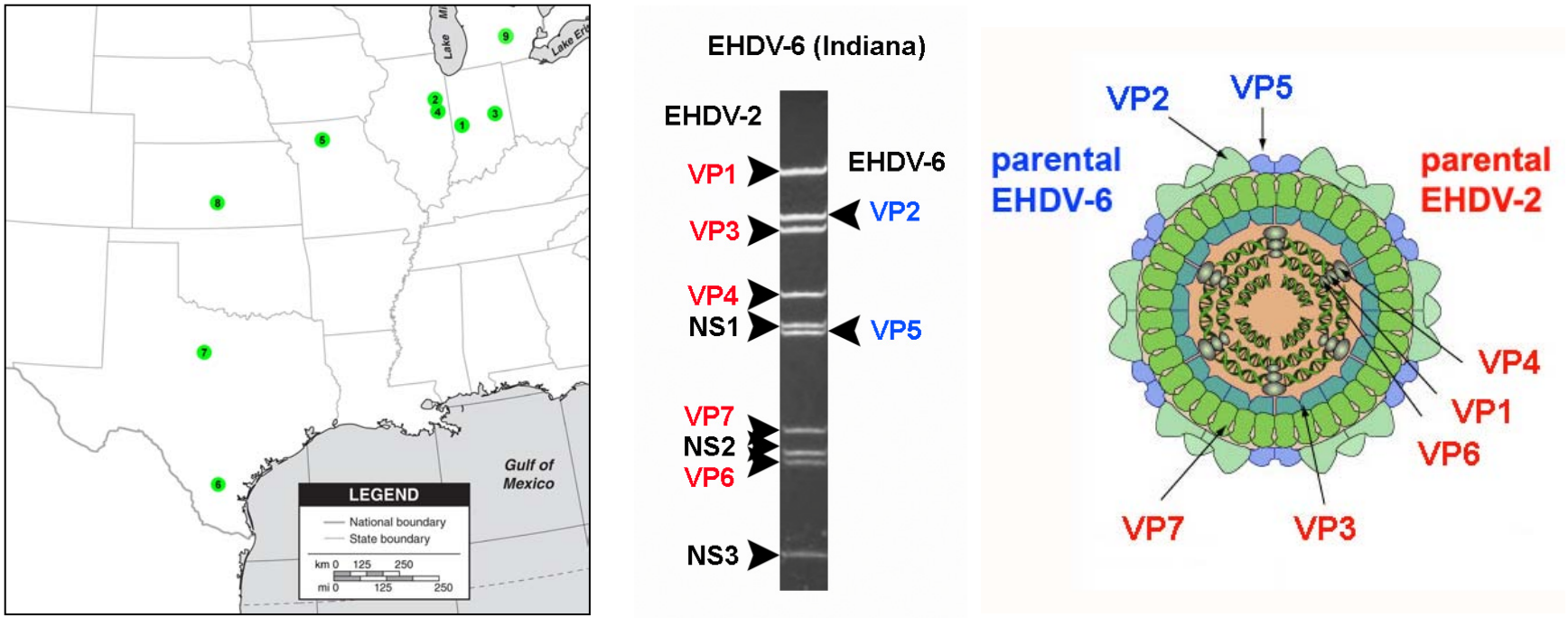


Figure 7.2. Reassortment in orbiviruses. (Left): Geographical locations of EHDV-6 (Indiana) isolates recovered from white-tailed deer in the United States during 2006-2009. Isolations are represented as cases 1-4 (Indiana and Illinois, 2006), case 5 (Missouri, 2007), cases 6-8 (Texas and Kansas, 2008), and case 9 (Michigan, 2009). (Center): Electrophoretic mobility of the 10 double-stranded RNA segments of the novel reassortant, *epizootic hemorrhagic disease virus* serotype 6, Indiana strain [EHDV-6 (Indiana)], isolated from the spleen of a moribund wild white-tailed deer in Indiana in 2006. Each of the 10 RNA segments and their corresponding genes are indicated with an arrowhead, with each arrowhead underneath the parental virus from which it was derived. Structural genes in blue indicate those RNA segments obtained from the exotic serotype, EHDV-6. Red genes (other structural genes) and black genes (non-structural genes) were derived from the endemic serotype, EHDV-2. (Right): The reconstructed protein configuration of the new EHDV reassortant. Proteins derived from EHDV-6 and EHDV-2 are shown in blue and red, respectively. Retrospective BLAST analysis of the VP2 and VP5 genes from the reassortant disclosed that it shared a 97-98% nucleotide identity to CSIRO 753, an EHDV-6 isolate from Australia (Orbivirus particle diagram courtesy of ViralZone of the Swiss Institute of Bioinformatics at http://education.expasy.org/images). See chapter 4 for details.

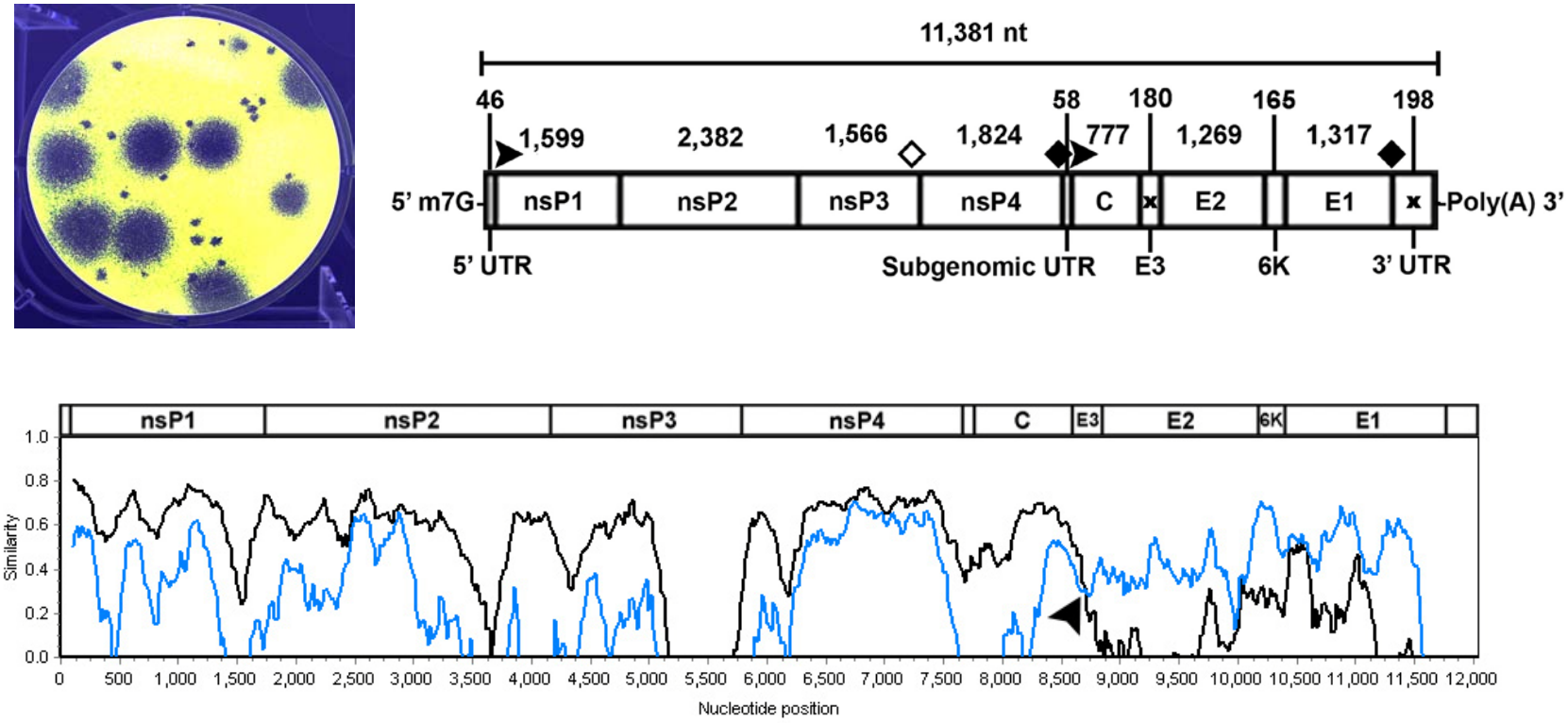
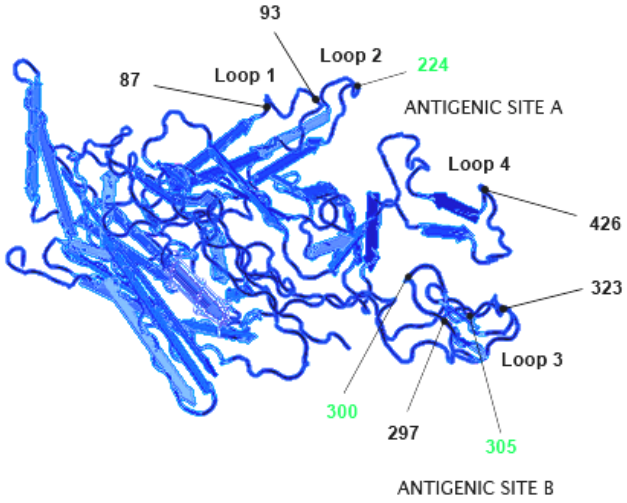


Figure 7.3. Recombination in alphaviruses. (Upper left): An *in vitro* co-infection of two *Highlands J virus* (HJV) isolates is shown, demonstrating alternate plaque morphologies induced by different E2 amino acid configurations (Upper right): Schematic of the recombinant alphavirus genome. The genome of *Fort Morgan virus* (FMV) is shown. The approximate E3 and 3' UTR recombination cross-over sites of the ancestral progenitor of HJV, FMV, and *western equine encephalitis virus* (WEEV) are indicated with an "X". (Lower): Recombination analysis of FMV to the parental viruses from which it was derived. A schematic of the FMV genome is shown to scale above the graph. Nucleotide similarity of EEEV (black line) and SINV (blue line) to the FMV genome is shown. The approximate location of the coding region recombination site (in the E3 gene) between EEEV and SINV is indicated with an arrowhead. The second cross-over site, which occurred at the terminal end of the 3' UTR, is not detectable by this method due to the high genetic variability of the region and its short sequence. See chapter 3 for details.




```
                                    ●●▼ ●  ●●  ●●   ●●●   ●▼  ●●●
Canine    RFREFQLSKSWRDEHFVEIQVKSSNAQNTVTIVDMESDLVYLAESPEGYVAYSKATTVTG 240
Raccoon   QLRTYLLSRVWHDEHFVKIQVKSSSARNSVTIVGTTSGVVYLVESPEGYVAYSKAATVTG 240
Feline    RFRELQLSKAWHDEHFVKVQVKG-SASNSVTIVGTNSGMVYLVESPEGYVAYSKAATVTG 239
          ::*   **: *:****.*::***. .* *:****. :*.:***.************:****
                          ■NXS/T
Canine    RLVHVNFGTKKDFENLKSPVNGSLVIARAGKITFAEKVANAQSYNALGVLIYMDQARFPI 300
Raccoon   RLVHANFGTKKDFENLNSRANGSLVIVRAGKITFAEKVANAQSFNAIGVLIYKDQSKFPI 300
Feline    RLVHANFGTKKDFENLNSPVNGSLVIVRAGKITFAEKVANAESFNAIGVLIYMDQAKFPI 299
          ****.***********.* .*******.***************:*:**:****** **:.***
            ● ●             ■NXS/T                                    ●
Canine    VNARIPFFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPSIPVQTISRAAAEKLFENME 360
Raccoon   INAEIPYFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPSIPVQTISRAAAEKLFENME 360
Feline    INAEIPFFGHAHLGTGDPYTPGFPSFNHTQFPPSQSSGLPNIPVQTISRANAEKLFGNME 359
           **.**.**********************************.********* ***** ***
          ●                  ●●▼ ●
Canine    GDCPSAWEIDPSCRLETSSNKNVNLTVNNVLKEIRIFNVFGVIKGFEEPDRYVVIGAQRD 420
Raccoon   GDCPSAWETDSSCRLETSQNRNVKLTVNNVLKEIRIFNIFGVIKGFEEPDRYVVIGAQRD 420
Feline    GDCPSAWETDSSCRLETSRNWNVKLSVNNVLKEIRIFNVFGVIKGFEEPDHYVVVGAQRD 419
          ******** *.******* * **:*:********#####:##########:***:*****
```

Virus	VP2 residue 80	87	93	103	224	232	297	300	305	323	426	564	568
RPV	K	M	K	V	G	V	S	A	D	D	D	N	A
FPV	K	M	K	V	G	V	S	A	D	D	N	N	A
MEV	K	M	K	V	G	V	S	V	D	D	N	N	A
CPV-2	R	M	N	A	G	I	S	A	D	N	N	S	G
CPV-2a	R	L	N	A	G	I	S/A	G	Y	N	N	S	G
CPV-2b	R	L	N	A	G	I	S/A	G	Y	N	D	S	G
CPV-2c[a]	R	L	N	A	G	I	A	G	Y	N	E	S	G
CPV-2c(a)[b]	R	L	N	A	G	I	A	D	Y	N	N	S	G
CPV-2c(b)	R	L	N	A	G	I	A	D	Y	N	D	S	G
CC 287-08 (GA)	R	L	N	A	G	I	A	D	H	N	N	S	G
CC 289-08 (GA)	R	L	N	A	G	I	A	D	H	N	N	S	G
CC 349-08 (GA)	R	L	N	A	G	I	A	D	H	N	N	S	G
CC 118-A-09 (VA)	R	L	N	A	R	I	A	D	D	N	N	S	G
CC 278-09 (VA)	R	L	N	A	G	I	A	D	H	N	N	S	G
CC 351-09 (TN)	R	L	N	A	R	I	A	D	D	N	N	S	G
CC 358-B-09 (KY)	R	L	N	A	R	I	A	D	D	N	N	S	G
CC 381-09 (FL)	R	L	N	A	G	I	A	D	H	N	N	S	G
CC 37-10 (WI)	R	L	N	A	G	T	S	D	D	N	N	S	G

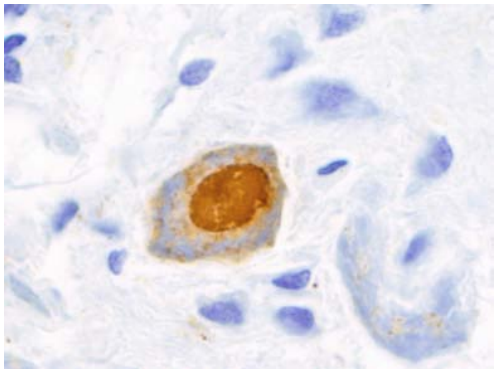


Figure 7.4. Mutation in parvoviruses. (Upper left): Ribbon diagram of the VP2 protein of canine parvovirus (CPV) (PBD ID 1IJS) showing amino acid substitutions of importance in defining the host range of the raccoon CPV isolates. Amino acids under apparent positive selection in the raccoon isolates are highlighted in green. (Upper right): A comparative amino acid alignment of the apical domain (highlighted in turquoise) of the canine, raccoon and feline transferrin receptors is shown. (Lower left): Comparison of the phylogenetically informative VP2 amino acid positions responsible for the differential host range and antigenicity observed among the carnivore parvoviruses. (Lower right): Immunohistochemical detection of CPV antigen in the nucleus of a Purkinje cell in the cerebellum (1000X) of a raccoon. See chapter 6 for details.

References

Appel, M. J., Parrish, C. R., 1982. Raccoons are not susceptible to canine parvovirus. J. Am. Vet. Med. Assoc. 181, 489.

Chen, W., Calvo, P. A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., Basta, S., O'Neill, R., Schickli, J., Palese, P., Henklein, P., Bennink, J. R., Yewdell, J.W., 2001. A novel influenza A virus mitochondrial protein that induces cell death. Nat. Med. 7, 1306-1312.

Conenello, G. M., Zamarin, D., Perrone, L. A., Tumpey, T., Palese, P., 2007. A single mutation in the PB1-F2 of H5N1 (HK/97) and 1918 influenza A viruses contributes to increased virulence. PLoS Pathog. 3, 1414-1421.

Ficken, M.D., Wages, D.P., Guy, J.S., Quinn, J.A., Emory, W.H., 1993. High mortality of domestic turkeys associated with Highlands J virus and eastern equine encephalitis virus infections. Avian Dis. 37, 585-590.

Gibbs, M. J., Gibbs, A. J., 2006. Was the 1918 pandemic caused by a bird flu? Nature 440, E8.28.

Hahn, C. S., Lustig, S., Strauss, E. G., Strauss, J. H., 1988. Western equine encephalitis virus is a recombinant virus. Proc. Natl. Acad. Sci. U. S. A. 85, 5997-6001.

Huyvaert, K. P., Moore, A. T., Panella, N. A., Edwards, E. A., Brown, M. B., Komar, N., Brown, C. R., 2008. Experimental inoculation of house sparrows (Passer domesticus) with buggy creek virus. J. Wildl. Dis. 44, 331-340.

Mertens, P. P. C., 1999. Orbiviruses and Coltiviruses (*Reoviridae*): General Features, p. 1043-1061 *In* Granoff, A., Webster R. G. (ed.), Encyclopedia of Virology, Second ed., Elsevier Academic Press, San Diego, California.

Pappas, C., Aguilar, P. V., Basler, C. F., Solórzano, A., Zeng, H., Perrone, L. A., Palese, P., García-Sastre, A., Katz, J.M., Tumpey, T. M., 2008. Single gene reassortants identify a critical role for PB1, HA, and NA in the high virulence of the 1918 pandemic influenza virus. Proc. Natl. Acad. Sci. U. S. A. 105, 3064-3069.

Parker, J. S., Murphy, W. J., Wang, D., O'Brien, S. J., Parrish, C. R., 2001. Canine and feline parvoviruses can use human or feline transferrin receptors to bind, enter, and infect cells. J. Virol. 75, 3896-3902.

Parrish, C. R., Have, P., Foreyt, W. J., Evermann, J. F., Senda, M., Carmichael, L. E., 1988. The global spread and replacement of canine parvovirus strains. J. Gen. Virol. 69, 1111-1116.

Parrish, C. R., 1999. Host range relationships and the evolution of canine parvovirus. Vet. Microbiol. 69, 29-40.

Randolph, K.D., Vanhooser, S.L, Hoffman, M., 1994. Western equine encephalitis virus in emus in Oklahoma. J. Vet. Diagn. Invest. 6, 492-493.

St. George, T. D., Cybinski, D. H., Standfast, H. A., Gard, G. P., & Della-Porta, A. J., 1983. The isolation of five different viruses of the epizootic haemorrhagic disease of deer serogroup. Aust. Vet. J. 60, 216-217.

Strauss, J. H., Strauss, E. G., 1997. Recombination in alphaviruses. Semin. Virol. 8, 85-94.

Taubenberger, J. K., Reid, A. H., Lourens, R. M., Wang, R., Jin, G., Fanning, T. G., 2005. Characterization of the 1918 influenza virus polymerase genes. Nature 437, 889-

Tumpey, T. M., Basler, C. F., Aguilar, P. V., Zeng, H., Solórzano, A., Swayne, D. E., Cox, N. J., Katz, J. M., Taubenberger, J. K., Palese, P., García-Sastre, A., 2005. Characterization of the reconstructed 1918 Spanish influenza pandemic virus. Science 310, 77-80.

Waters, J.R., 1976. An epidemic of western encephalomyelitis in humans-Manitoba, 1975. Can. J. Public Health 67, S1:28-32.

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