

**Expanding Genomic Surveillance of Eastern Equine Encephalitis Virus in Canada:
Evolutionary Dynamics, Diversity, and Geographic Spread**

By

Daniel Sullivan

A Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
of The University of Manitoba
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biological Sciences

University of Manitoba

Winnipeg, MB

Copyright © 2026 by Daniel Sullivan

ABSTRACT

Eastern equine encephalitis virus (EEEV) is an important mosquito-borne alphavirus associated with severe neurological disease in humans and horses. Despite recurring detections in Canada, the genomic diversity, evolutionary relationships, and geographic history of Canadian EEEV remain poorly understood, with only one historical Canadian genome previously available. This research developed a targeted tiling-amplicon and rapid amplification of cDNA ends (RACE) workflow to recover complete coding sequences and terminal genomic regions from Canadian surveillance samples. Forty-one novel Canadian EEEV genomes were generated from equine, avian, mosquito, and wildlife samples and analyzed alongside 543 publicly available North American sequences using comparative genomic, phylogenetic, phylogeographic, discrete-trait, and haplotype analyses. North American EEEV genomes were highly conserved, with most pairwise comparisons exceeding 99% nucleotide identity, low genome-wide entropy, and widespread purifying selection. Host-state reconstructions were broadly consistent with mosquito-borne transmission but represented statistical associations rather than direct transmission chains. Canadian genomes were distributed across multiple North American lineages and haplotypes rather than forming a single geographically restricted population, consistent with multiple introductions into Canada or broader regional circulation. However, limited sampling, temporal gaps in available genomes, and high genomic conservation prevented definitive discrimination between short-term local persistence and repeated introductions of closely related lineages. Overall, this research substantially expands the genomic representation of EEEV in Canada and demonstrates the value of integrating sequencing with surveillance to improve the interpretation of arboviral diversity, evolution, and movement.

ACKNOWLEDGMENTS

I want to thank my supervisor, Dr. Oliver Lung, and co-supervisor, Dr. Steve Whyard, for the opportunity to complete my thesis research under their supervision and for the support they have provided me throughout my degree. I would also like to thank my committee members, Dr. Kenneth Jeffries and Dr. Peter Pelka, for their feedback and insight. I am also grateful to Dr. Davor Ojkic, Dr. Heather Coatsworth, and Dr. Brad Pickering for providing the surveillance samples used in this research. Next, I would like to thank the members of the Genomics Unit, both past and present, for supporting me along my journey. We shared many laughs and smiles, and it made going to the lab each day something I looked forward to. I want to express my sincerest gratitude to my parents (Michael and Joanne Sullivan), my brother (Myles Sullivan), and my partner (Shalyn Falloon) for their continued love, support, and encouragement throughout my degree. Finally, I would like to thank my cat, Mia, for continually reminding me to take breaks by sleeping on my keyboard or blocking my computer screen.

TABLE OF CONTENTS

TITLE	i
ABSTRACT.....	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES.....	vi
LIST OF SUPPLEMENTARY TABLES	vi
LIST OF FIGURES	vii
LIST OF SUPPLEMENTARY FIGURES.....	vii
CHAPTER 1: BACKGROUND AND RELEVANCE	1
1.1 Introduction	1
1.2 Biology of Alphaviruses Including Eastern Equine Encephalitis Virus.....	3
1.2.1 Alphavirus Classification and Genome Organization	3
1.2.2 Clinical Significance and Host Range of EEEV	4
1.2.3 EEEV Genome Organization and Ecology	6
1.2.4 Geographic Distribution and Avian Hosts	9
1.2.5 Bridge Vectors and Spillover.....	11
1.3 History and Surveillance of EEEV in Canada	13
1.4 Viral Evolution and Phylodynamics.....	17
1.4.1 RNA Virus Evolution	18
1.4.2 Molecular Epidemiology and Phylogenetics.....	19
1.4.3 Temporal Evolution and Phylodynamics.....	20
1.4.4 Phylogeography and Viral Dispersal	21
1.4.5 Genetic Diversity and Population Structure	22
1.5 Genomic Surveillance Technologies.....	23
1.5.1 Conventional Viral Detection and Surveillance	23
1.5.2 Next-Generation Sequencing and Targeted Enrichment	25
1.6 Objectives of Research.....	27
1.6.1 Characterization of Complete EEEV Genomes from Canadian Surveillance Samples	27

1.6.3 Evolutionary History, Geographic Dispersal, and Host-State Associations of EEEV in North America	29
CHAPTER 2: MATERIALS AND METHODS.....	30
2.1 <i>In silico</i> EEEV tiling amplicon PrimalScheme and 5'/3' RACE primer design	30
2.2 RNA extraction, cDNA synthesis, PCR, and RACE	32
2.3 Next-generation sequencing and genome assembly	33
2.4 EEEV dataset construction and phylogenetic analysis	34
2.5 Time-scaled Bayesian phylogenetic inference.....	36
2.6 Spatiotemporal dissemination and host-transition network analysis	37
2.7 Sequence variation and selection pressure analysis	38
2.8 Population genetic structure and haplotype networks	40
CHAPTER 3: RESULTS.....	42
3.1 Dataset Characteristics and Novel Canadian EEEV Genomes	42
3.2 EEEV Genome-Wide Variability and Selection Pressure.....	47
3.3 Phylogenetic Reconstruction and Temporal Signal Evaluation.....	50
3.4 Time-Resolved Phylogenetic Analysis of EEEV Genome Sequences	53
3.5 Inferred EEEV Spread in North America	57
3.6 Phylogeographic and Host-State Transition Dynamics	60
3.7 Geographic Distribution of North American EEEV Haplotypes.....	63
CHAPTER 4: DISCUSSION	66
4.1 Targeted Genome Sequencing Expands Canadian EEEV	66
4.2 Evolutionary Constraint Shapes EEEV Genomic Diversity	67
4.3 Canadian EEEV Diversity Is Integrated Within the Broader North American Population	70
4.4 Repeated Incursions Are Predominant but Short-Term Persistence Possible.....	71
4.5 North American Dispersal and Northeastern U.S. Circulation	74
4.6 Host-State Reconstruction Reflects Ecological Associations, Not Direct Transmission Chains.....	76
CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS	79
APPENDICES	82
LITERATURE CITED.....	88

LIST OF TABLES

Table 1.1. Selected reports of EEEV activity in Canada.	13
Table 2.1. Primer information for EEEV tiling amplicon sequencing.	30
Table 2.2. Primer sequences for EEEV 5'/3' RACE.	32
Table 3.1 Novel Canadian EEEV genome epidemiological metadata.	44

LIST OF SUPPLEMENTARY TABLES

Supplementary Table S1.1. Significant viral spatial transmission pathways estimated by Bayesian Stochastic Search Variable Selection (BSSVS).	82
Supplementary Table S1.2. Significant viral host-state transition pathways estimated by Bayesian Stochastic Search Variable Selection (BSSVS).	83
Supplementary Table S1.3. Geographic regions used in haplotype analysis.	83
Supplementary Table S1.4. Metadata for EEEV sequences analyzed in this study.	83

LIST OF FIGURES

Figure 1.1. Enzootic maintenance and spillover transmission of eastern equine encephalitis virus.....	7
Figure 3.1. Geographical and ecological distribution of North American EEEV sequences.	43
Figure 3.2. Pairwise nucleotide identity among North American EEEV genomes.	46
Figure 3.3. Genomic diversity and site-specific selection across North American EEEV genomes.	49
Figure 3.4. Phylogenetic relationships and temporal signal evaluation of North American EEEV.	52
Figure 3.5. Discrete time-calibrated phylogeographic analysis of North American EEEV dispersal.	56
Figure 3.6. Spatiotemporal reconstruction of EEEV dispersal in North America.	59
Figure 3.7. Spatiotemporal diffusion and host-state transition dynamics of EEEV.....	62
Figure 3.8. Haplotype network of North American EEEV.....	64

LIST OF SUPPLEMENTARY FIGURES

Supplementary Figure S1.1. Schematic representation of EEEV PrimalScheme and 5'/3' RACE.....	82
---	----

CHAPTER 1: BACKGROUND AND RELEVANCE

1.1 Introduction

Viruses are a diverse and widespread group of biological entities found across aquatic and terrestrial environments (Suttle, 2007). While the vast majority remain silent and unnoticed across diverse hosts, a subset can cause notable diseases in plants, animals, and humans.

Arthropod-borne viruses (arboviruses) are an important group of human and animal pathogens, and their associated vector-borne diseases represent an emerging global health threat. It is estimated that vector-borne diseases account for more than 700,000 deaths annually and account for more than 17% of all infectious diseases globally (World Health Organization, 2024). This threat is exacerbated by human activity, including urbanization, animal husbandry, habitat destruction, and the development of global transportation networks (De Souza & Weaver, 2024). As human populations expand into wildlife habitats, increased contact among humans, animals, and arthropod vectors can increase opportunities for the spillover of zoonotic arboviruses (Plowright et al., 2017). It is estimated that approximately half the world's population will be at risk of arbovirus infection by the year 2050 (Jones et al., 2021).

Many arboviruses are maintained in sylvatic transmission cycles, where viruses circulate naturally between wild vertebrate hosts and arthropod vectors, often without causing severe disease in reservoir species (Weaver & Barrett, 2004). Severe disease is often most apparent following spillover into incidental hosts, including humans and other animals (Gould et al., 2017). Arthropods capable of transmitting arboviruses include mosquitoes, ticks, blackflies and biting midges. Furthermore, these hematophagous (blood-feeding) species can significantly affect health by transmitting diseases (Perumalsamy et al., 2024). Mosquitoes (Diptera: Culicidae) are the most important invertebrate disease vectors for maintaining these sylvatic

cycles. They are also major contributors to global disease burdens because they carry high-consequence and emerging diseases (Franklin et al., 2019) and are often considered the most dangerous disease vector globally (Rakotoarinia et al., 2022). Enzootic transmission cycles occur after successful viral colonization of the host mosquito and dissemination to the salivary glands (Conway et al., 2014). In Canada, mosquito-borne viruses of medical and veterinary importance include West Nile virus (WNV), Eastern equine encephalitis virus (EEEV), Cache Valley virus (CVV), and members of the California serogroup viruses (CSGV), such as Jamestown Canyon virus (JCV) and snowshoe hare virus (SHV) (Ludwig et al., 2019). Mosquito vectors for these viruses include *Culex spp.* and non-*Culex* species, while reservoir hosts vary by mosquito species and can include white-tailed deer, small mammals and birds.

Concerns about climate change and its impact on the expansion of competent invertebrate disease vectors and their associated vector-borne diseases remain. Increasing ambient temperatures and changes in precipitation patterns can influence the spatiotemporal distribution of many disease vectors (Bakhiyi et al., 2024). Temperature changes also play an important role in developmental and reproductive rates, influencing vector persistence across locations and prolonging seasonal activity (Khan et al., 2020). For diseases already endemic to an area, climate change may increase variability and unpredictability in their seasonal and geographic distribution (Mitri et al., 2026). Furthermore, climate change may facilitate the introduction of exotic species, establishing new vectors for arboviruses and vector-borne diseases (Shahhosseini et al., 2020).

Importantly, there is currently no formal national surveillance system for many endemic mosquito-borne diseases (Mitri et al., 2026). Therefore, there is a need to deepen our understanding of arbovirus movement and evolution over time, and to develop future mitigation strategies to protect both human and animal health.

1.2 Biology of Alphaviruses Including Eastern Equine Encephalitis Virus

1.2.1 Alphavirus Classification and Genome Organization

Alphaviruses belong to the *Togaviridae* family and are primarily transmitted by mosquitoes (Luo et al., 2025). There are currently over 30 members in the genus, organized into seven species complexes (Atkins, 2013). Alphaviruses are enveloped viruses with single-stranded, positive-sense RNA genomes approximately 10-12 kb in length, allowing the genomic RNA to be translated directly (Fros & Pijlman, 2016). Their genome contains two open reading frames (ORFs). One ORF encodes the four non-structural proteins (nsP1-4), which are essential for viral replication. The other ORF encodes the structural proteins (SP), which include the capsid protein and envelope glycoproteins (E1-3) (Jose et al., 2009). Alphaviruses can be broadly divided into New and Old World viruses (Luo et al., 2025) based on their geographic distribution and disease presentation. Old World alphaviruses are mainly found in Africa and Asia and are often associated with arthritic diseases in humans. Notable members of these Old World lineages include Chikungunya virus (CHIKV), O'nyong'nyong virus (ONNV), Sindbis virus (SINV), Ross River virus (RRV), and Semliki Forest virus (SFV) (Weaver et al., 2012). New World lineages, predominantly found in the Americas, are often associated with neurological disease and include viruses such as Venezuelan equine encephalitis virus (VEEV), Western equine encephalitis virus (WEEV), and Eastern equine encephalitis virus (EEEV) (Weaver et al., 2012; Zacks & Paessler, 2010). Alphaviruses have been shown to infect both vertebrate hosts and invertebrate vectors and can cause significant disease. In humans, this can range from mild acute febrile illness to severe encephalitis. In animals, diseases can range from pancreatic disease in Atlantic salmon (*Salmo*

salar) and rainbow trout (*Oncorhynchus mykiss*) (Deperasińska et al., 2018) to neuroinvasive diseases in equines (Corrin et al., 2021).

1.2.2 Clinical Significance and Host Range of EEEV

Eastern equine encephalitis virus is of particular importance. Despite relatively low annual case numbers, the high virulence and potential to cause disease make understanding its movement especially important for surveillance and disease prevention. In humans, EEEV is associated with case-fatality rates of approximately 30-70%, and up to 90% of survivors experience long-term neurological complications, making it one of the most virulent human encephalitic viruses in North America (Atkins, 2013; Williamson et al., 2020). Disease symptoms can range from febrile illness, characterized by fever, chills, and body and joint pain, to neurologic disease, which may include encephalitis and meningitis (U.S. CDC, 2026b). Resulting long-term health complications can cost an estimated \$400,000 USD per individual per year (Kehn-Hall & Bradfute, 2022). Approximately 438 human cases have been reported in the United States from 1964 to 2025 (U.S. CDC, 2026a; Zacks & Paessler, 2010). One of the largest outbreaks on record occurred in the United States in 2019. A total of 38 cases were reported nationally, with 23 occurring in the northeastern states (Hill et al., 2023). Environmental conditions during that outbreak may have supported higher mosquito abundance and viral amplification (Armstrong & Andreadis, 2022; Lindsey et al., 2020). Reported human case counts have increased in recent decades, although annual activity remains episodic. For example, the average number of cases in the northeastern United States increased from approximately 1.2 per year until the early 2000s to 4.9 per year over the last two decades, representing a fourfold increase and indicating a potential resurgence and expansion of the virus (Armstrong & Andreadis, 2022). While the exact mechanisms are likely complex and remain unknown,

environmental changes that support increased mosquito abundance and virus amplification have been suggested as potential contributors to this increase (Armstrong & Andreadis, 2022).

Historically, an outbreak of encephalomyelitis in 1831 may have represented the first occurrence of EEEV in Massachusetts (Zubair et al., 2024). However, it was not until 1933 that the virus was first identified in a horse and subsequently isolated from an avian host in 1937 (Broeck & Merrill, 1933; Calisher, 1994). Subsequent outbreaks have occurred across Canada, the Caribbean, and the United States (Calisher et al., 1979; Chénier et al., 2010; Corrin et al., 2021; Letson et al., 1993). Because of its high virulence potential and the lack of a human vaccine, EEEV is also considered a potential bioterrorism agent, making it important to establish methods to understand its movement. It has also been reported that EEEV can be transmitted via aerosolization, which has led government agencies, including the United States Department of Agriculture (USDA) and the Centers for Disease Control and Prevention (CDC), to classify EEEV as a select agent (Hasan et al., 2021). In unvaccinated equids, mortality can be as high as 90%, and they tend to be the first hosts to show signs of disease, thus serving as potential indicators for outbreaks (Zacks & Paessler, 2010). While an equine vaccine is available, a study of historical cases found that although most horses had received at least one vaccine in their lifetime, only 24% had been vaccinated within the recommended 6-month window, and vaccination history did not significantly influence survival outcomes (Gomez et al., 2026). EEEV has also been shown to infect a variety of animals, including passerine and non-passerine birds; domestic farm animals such as equines, pheasants, dogs, and swine; and other mammals, including bats, white-tailed deer, alpacas, and opossums (Schmitt et al., 2007). In contrast to most incidental mammalian hosts, emus (*Dromaius novahollandiae*) can develop high levels of viremia, acting as amplifying hosts that shed the virus into the environment, thereby potentially

causing horizontal transmission to other emus in the absence of mosquito vectors (Tengelsen et al., 2001).

1.2.3 EEEV Genome Organization and Ecology

As with other alphaviruses, EEEV is a mosquito-borne, single-stranded, positive-sense RNA virus with a genome size of approximately 11.7 kb, primarily transmitted between mosquitoes and vertebrate hosts. Publicly available North American EEEV genomes exhibit approximately 99% average pairwise sequence identity, indicating relatively low genomic diversity (Arrigo et al., 2010). Historically, EEEV was divided into four lineages (Groups I-IV), with Group I encompassing strains circulating in North America and the Caribbean and accounting for most human cases (Corrin et al., 2021). Groups II–IV, primarily associated with Central and South American strains, were distinguished from North American EEEV by substantial genetic divergence, including approximately 23–24% nucleotide divergence, as well as differences in evolutionary history and ecology (Arrigo et al., 2010). These differences supported their reclassification as Madariaga virus (MADV) in 2010, a distinct species within the EEE complex, rather than representing high diversity within North American EEEV itself (Arrigo et al., 2010). The EEEV genome consists of a 5' cap and UTR, which contains conserved sequence elements essential for promoter recognition and viral RNA synthesis (Hyde et al., 2015). This is followed by ORF1, which encodes the four non-structural proteins, nsP1-4, translated as a single polyprotein that supports viral RNA synthesis and polyprotein processing (Aguilar et al., 2007; Hasan et al., 2021). Translation of the nsP1-4 polyprotein relies on an opal stop codon between nsP3 and nsP4, where low-frequency read-through allows the translation of the nsP4 RNA-dependent RNA polymerase (RdRp). A 26S subgenomic promoter controls the second ORF, which encodes the structural polyprotein C-E3-E2-6K/TF-E1, containing the

capsid, E1-E2 glycoproteins, E3 protein, and 6K/transframe protein (Carey et al., 2019). The genome terminates in a 3' UTR featuring regulatory elements and a poly(A) tail essential for viral RNA replication and stability (Hyde et al., 2015). Successful replication of the virus requires receptor-mediated endocytosis, followed by the release of the viral RNA into the host cell cytoplasm, where the non-structural proteins involved in viral genome replication are translated. This leads to transcription of subgenomic RNA, a shorter RNA transcript used to produce the structural proteins, which are processed in the endoplasmic reticulum and Golgi before assembly and packaging of viral RNA into new virions (Carey et al., 2019; Hasan et al., 2021).

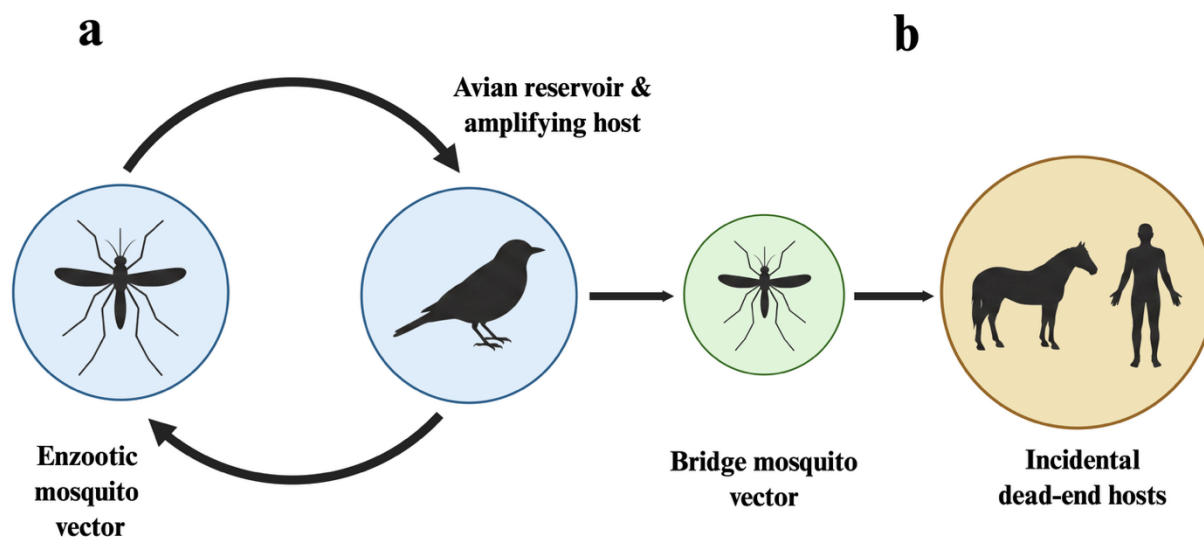


Figure 1.1. Enzootic maintenance and spillover transmission of eastern equine encephalitis virus. (a) EEEV is maintained in an enzootic transmission cycle between avian reservoir and amplifying hosts and enzootic mosquito vectors. (b) Bridge mosquito vectors may transmit EEEV from the enzootic cycle to incidental dead-end hosts, including horses and humans. Created with <https://BioRender.com>.

EEEV is maintained in an enzootic cycle in which the primary vector, *Culiseta melanura* (Diptera: Culicidae), transmits the virus to passerine (perching) birds, which act as the main reservoirs and amplifying hosts (Figure 1.1a). Spillover to incidental hosts is facilitated by bridge vectors that feed on both birds and mammals (Figure 1.1b). Inland swamps, such as those in

Tennessee, may have a different primary vector, such as *Culex erraticus*, the most abundant mosquito species, rather than *Cs. melanura*, which is low in abundance or absent at potential outbreak sites (Cohen et al., 2009; Mukherjee et al., 2012). The ecological network of EEEV is extensive, encompassing over 35 mosquito species, including *Aedes vexans*, *Aedes sollicitans*, *Ochlerotatus canadensis*, and *Coquillettidia perturbans*.

Furthermore, the enzootic landscape of this virus is vast, encompassing a host range of over 200 distinct species across avian, mammalian, reptilian, and amphibian taxa (Bingham et al., 2014; Corrin et al., 2021; Mundis et al., 2022). Frequent transmission occurs in regions characterized by woodlands and marshlands, which provide suitable habitats for mosquito larvae development (Parashar et al., 2025). In northern climates, larvae overwinter, and adults emerge from late April to June to feed on birds before oviposition. The second generation can emerge from June through August, continuing the EEEV amplification cycle (Elias et al., 2017). In the southeastern United States, these environments provide year-round habitats for vector feeding and serve as preferred habitats for migrating birds in both the spring and fall (Burkett-Cadena et al., 2022). Migratory birds are assumed to facilitate the long-distance northward spread of EEEV, while human activity may also contribute by transporting infected mosquitoes and altering landscapes that shape vector, host, and human interactions (Armstrong & Andreadis, 2022). The presence of competent birds that enable enzootic establishment is important; however, large bird populations may also have a negative effect, diluting the virus (Hill et al., 2023). Migratory birds can also play a major role in the dissemination of EEEV, given their seasonal movement patterns and ability to cover large distances (Sparagano et al., 2015). Serological studies of birds have shown that many species have antibodies indicating extensive enzootic circulation of EEEV within local ecosystems and resulting in amplification between

reservoir hosts and competent mosquito vectors (Armstrong & Andreadis, 2022; Pedersen et al., 2016). Serum samples from both white-tailed deer and moose have indicated that the landscape for EEEV prevalence may be larger than previously described and may serve as useful surveillance markers (Downs et al., 2015; Mutebi et al., 2012, 2015).

1.2.4 Geographic Distribution and Avian Hosts

EEEV can be found in almost every eastern U.S. state, including those that border the Gulf of Mexico or the Atlantic Ocean, as well as the eastern Canadian provinces (Stobierski et al., 2022). The epidemiology of the virus remains complex, however. In northern latitudes, EEEV prevalence and cases follow trends in mosquito populations and bird migratory activity, with increases in the warmer summer months and declines in the colder months (Armstrong & Andreadis, 2022). By contrast, in the southeastern United States, particularly Florida, transmission occurs year-round (Heberlein-Larson et al., 2019). This makes Florida an important EEEV enzootic focus because it supports year-round enzootic circulation and geographic mixing, which can lead to seasonal outbreaks in northern states, facilitated by migratory bird movement (Tan et al., 2018).

Passerines represent the most diverse group of avian species, with nearly 6400 species, and include sparrows, starlings, thrushes, magpies, crows, swallows, and finches (Williams et al., 2023). Bird species diversity and relative abundance are also important factors, as putative EEEV vectors do not feed opportunistically on birds and instead show preferences for some species over others (Estep et al., 2011). For example, in some regions *Cs. melanura* uses the Northern Cardinal as a preferred host, demonstrating that regional differences in host abundance can drive local EEEV ecology (Molaei, Armstrong, Abadam, et al., 2015). The enzootic amplification of EEEV relies heavily on a diverse avian reservoir comprising at least 44

passerine species (Estep et al., 2011; Molaei, Armstrong, Graham, et al., 2015). Among these, some have been identified as potentially important amplification hosts because of their abundance, mosquito-feeding associations, and transmission competence (Molaei et al., 2016). In fact, the Wood Thrush appears to be one of the most frequent hosts for *Cs. melanura* across multiple Northeastern U.S. states, including New York, Connecticut, New Jersey, and Massachusetts (Molaei et al., 2006, 2013, 2016; Molaei & Andreadis, 2006). However, high rates of EEEV have also been found in Wood Thrushes in southern Alabama (Stamm, 1968). Given their large habitat and breeding range, which can span from eastern Canada along the eastern coast of the United States to Central America, Wood Thrushes may play a potential role in the spread of EEEV (Stanley et al., 2015). American Robins also appear to be important, frequently serving as vertebrate hosts for mosquito blood meals and demonstrating evidence of EEEV infection through molecular or serological detection (Molaei et al., 2015). They are also frequently found near wetland breeding sites for *Cs. melanura* (Komar et al., 1999). Similarly, Robins have a large geographic range and tend to migrate south after the breeding period, which may increase the chances of human or animal infections by reducing the availability of preferred hosts (Brüssow & Figuerola, 2025).

In Florida, mosquito host preferences may shift from passerines to wading birds (Ciconiiformes), such as white ibis (*Eudocimus albus*), egrets, and herons (*Ardea spp.*), as well as other enzootic hosts (Bingham et al., 2014). Some wading-bird species are highly susceptible to EEEV and may develop clinical disease (Kissling et al., 1954; McLean et al., 1995). Early research demonstrated that natural or experimental EEEV infection occurs in at least 48 avian species across seven orders, including various waterfowl, wading birds, gamefowl, and 29 distinct species of passerines (Herman, 1962). Beyond avian species, seasonality may also

influence the preference for *Cs. melanura* to feed on reptiles as day length and temperature increase in the summer (Blosser et al., 2017). Cottonmouth and copperhead snakes from Alabama have demonstrated evidence of EEEV infection through the detection of viral RNA, while 35.4% of sampled cottonmouths showed serological evidence of exposure (Bingham et al., 2012). Experimental studies further suggest that snakes can maintain viremia through hibernation, supporting a potential role in EEEV overwintering and early-season amplification (Bingham et al., 2012). Given the vast diversity of avian species, determining a competent host can be quite difficult and relies heavily on serological surveys. Serological evidence alone does not establish host competence, which also requires sufficient viremia and survival for onward mosquito infection (Burkett-Cadena et al., 2022).

1.2.5 Bridge Vectors and Spillover

Putative bridge vectors also play an important role in EEEV transmission. They act as competent vectors that opportunistically feed on various animals, potentially transmitting disease from a reservoir host to an incidental host and inducing disease spillover (Corrin et al., 2021). While many studies have investigated the detection of EEEV in potential bridge vectors, they often fail to consider viral titre and whether it is sufficient for transmission (Armstrong & Andreadis, 2010). While *Cs. melanura* predominantly feeds on birds, blood-meal analyses show that up to 10% of meals are derived from equines and humans, occasionally allowing it to act as a bridge vector (Molaei et al., 2006, 2013).

During epizootic cycles and potential outbreaks, *Aedes*, *Culiseta*, *Culex*, and *Coquillettidia* species can act as important bridge vectors for EEEV transmission to incidental hosts due to their more aggressive feeding behaviours (Zubair et al., 2024). Increasingly, *Cq. perturbans* is recognized as one of the main bridge vectors for transmission to humans and

horses. This is due to its similar geographic range and its tendency to feed opportunistically on both birds and mammals, making it better suited as a bridge vector (Andrews et al., 2014; Batzer et al., 2020). It has also been demonstrated that *Cq. perturbans* can carry sufficient viral titres for transmission (Nasci & Mitchell, 1996). However, its low abundance in late August and September, when EEEV activity peaks in the same region, suggests that other vectors remain important (Armstrong & Andreadis, 2010).

A twenty-year surveillance program for EEEV in New York State concluded with over 629 positive detections across 11 mosquito species and 7 genera (Oliver et al., 2018). As expected, *Cs. melanura* was the predominant vector (87% of positive results), while the next-largest share (10% of total) came from *Cq. perturbans*, *Cs. morsitans*, and *Ae. canadensis*. Another study summarizing EEEV findings in mosquito species also found similar results, with *Cs. melanura* being the predominant vector, but other bridge vectors such as *Cq. perturbans*, *Ae. canadensis*, *Ae. vexans*, and *Cs. morsitans* were also common (Corrin et al., 2021). In other northeastern U.S. states, such as Connecticut, New Jersey, and Massachusetts, EEEV-positive pools included over a dozen mosquito species, with *Cq. perturbans*, *Cx. pipiens/restuans*, and various *Aedes spp.* among the frequently detected mosquito species (Armstrong & Andreadis, 2022).

How EEEV persists through winter and re-emerges each season remains unresolved. Proposed mechanisms include annual reintroduction by migratory birds, vertical transmission and larval overwintering in *Cs. melanura*, and cryptic persistence within hibernating reptiles or amphibians (Armstrong & Andreadis, 2022; Young et al., 2008).

1.3 History and Surveillance of EEEV in Canada

Research on EEEV in Canada remains limited, leaving significant knowledge gaps regarding its true prevalence and whether recurring detections reflect sustained local transmission or repeated introduction. EEEV case reporting has been sporadic, with fewer than 10 human cases and roughly 108 equine cases across Ontario, Quebec, New Brunswick and Nova Scotia (Canadian Food Inspection Agency, n.d.; Stephens et al., 2026). To date, surveillance studies have largely focused on routine viral detection rather than collecting sequence data needed for genetic characterization. Before this research, only one Canadian EEEV genome was publicly available: a virus isolated from a horse in Ontario in 1961 (GenBank accession KJ469595.1). However, the first reported EEEV cases in Canada occurred in horses from St. George and St. Catharines, Ontario, in 1938 (Carman et al., 1995; Nelder et al., 2014). Subsequent activity has included sporadic equine cases, outbreaks involving equines and emus, evidence of exposure through equine serosurveillance, and a small number of human cases (Table 1.1).

Table 1.1. Selected reports of EEEV activity in Canada.¹

Year	Province	Host	Reported activity	Source
1938	Ontario	Horse	First reported Canadian EEEV in equines from St. George and St. Catharines.	(Carman et al., 1995; Nelder et al., 2014)
1961	Ontario	Horse	Genome sequence from an equine sample; source of the only publicly available Canadian EEEV genome before this study.	GenBank ²
1972	Quebec	Horse	EEEV outbreak during September and October in Quebec equines.	(Sellers, 1989)
1992	Ontario	Horse	A single equine case was reported in southwestern Ontario following the acute onset of neurological symptoms.	(Carman et al., 1995)

1999	Quebec	Horse	A single equine case was reported in December.	(Desjardins et al., 2000)
2001–2009	Ontario	Horse	Twenty-four sporadic equine EEEV cases were reported over this period, including 2003 outbreak.	(Government of Ontario, 2022)
2003	Ontario	Horse, emu	Eleven equine and one emu case represented the largest reported EEEV outbreak in Canada at the time.	(Government of Ontario, 2022)
2008	Quebec	Horse, emu	Nineteen equine and a flock of emus tested positive during the fall season; all 19 horses were euthanized and 16 of 35 emus died.	(Chénier et al., 2010)
2009–2024	Nova Scotia	Horse	The first equine case was reported in 2009, with 10 cases reported by 2024.	(Government of Nova Scotia, 2026)
2012	Quebec	Horse	An equine serosurvey in southern Quebec identified regional neutralizing-antibody seroprevalence of 6–14%.	(Rocheleau et al., 2013, 2017)
2016	Ontario	Human	The first reported human EEEV case in Canada occurred during the summer.	(Public Health Agency of Canada, 2017)
2016–2021	Ontario	Horse	Approximately 39 equine cases were reported during this time period.	(Public Health Agency of Canada, 2021)
2022	Ontario	Human	A second human case was reported.	(Public Health Agency of Canada, 2024a)
2023	Ontario, New Brunswick, Quebec	Horse	A record 22 equine cases were reported: Ontario (18), New Brunswick (2), and Quebec (2).	(Public Health Agency of Canada, 2023)
2024	Ontario, Quebec, Nova Scotia	Human, horse	EEEV was detected in 38 equines: Ontario (23), Quebec (14), and Nova Scotia (1); four human cases were also reported.	(Public Health Agency of Canada, 2024b)
2025*	Ontario, Quebec	Human, horse	Three human cases and 10 equine cases were reported as of November 1, 2025.	(Public Health Agency of Canada, 2025)

*The 2025 case counts were provisional at the time of reporting.

(1) Only published and publicly documented reports are listed. Internal diagnostic submissions and novel sequences are evaluated in Chapter 3.

(2) GenBank; Eastern equine encephalitis virus strain EEEV/*Equus ferus caballus*/CAN/32-14-989/1961, complete genome (<https://www.ncbi.nlm.nih.gov/nuccore/KJ469595.1>)

Notably, a 2003 Ontario outbreak, involving 11 horses and one confirmed emu, was the largest reported Canadian EEEV outbreak until the 2008 outbreak in Quebec involving 19 horses and 16 emus (Chénier et al., 2010; Government of Ontario, 2022). During the 2008 outbreak, an entomological survey was conducted; however, freezing temperatures late in the season resulted in low mosquito trap numbers ($n = 118$), and no specimens tested positive by PCR (Chénier et al., 2010). A 2012 equine EEEV serosurvey in southern Quebec found an overall neutralizing-antibody seroprevalence of 6–14% in some regions, suggesting that non-lethal EEEV infections in horses, potentially influenced by high vaccination rates, could serve as useful early indicators of viral activity within a region (Rocheleau et al., 2013, 2017). Reported activity has increased in recent transmission seasons. Canada observed a record number of EEEV-positive horses in 2023, with 22 cases reported across Ontario, New Brunswick, and Quebec (Public Health Agency of Canada, 2023). This record was exceeded in 2024, when EEEV was detected in 38 horses, primarily in Ontario and Quebec, along with four reported human cases, the highest annual number of human cases reported in Canada to date (Public Health Agency of Canada, 2024b). The most recent 2025 seasonal update reported three human cases and 10 equine cases as of November 1, 2025 (Public Health Agency of Canada, 2025). Despite these detections, the limited availability of sequence information has prevented a detailed understanding of whether EEEV is repeatedly introduced into Canada or maintained locally between transmission seasons.

Until recently, Canadian EEEV surveillance remained minimal and primarily reactive to sporadic detections in horses or mosquito vectors. Furthermore, despite wild birds' role in viral amplification and spread, no formal active avian surveillance program currently exists for endemic Canadian arboviruses, even though similar frameworks were successfully used during

the early West Nile virus outbreaks (Mitri et al., 2026). Human infections are likely underreported, as EEEV infection is not a nationally notifiable disease in Canada and therefore lacks a formal national surveillance system (Public Health Agency of Canada, 2025). Mosquito vector surveillance programs have also not been routinely used to assess the presence of EEEV. It was not until 2010 that the Ontario Ministry of Health and Long-term Care recommended that public health units begin monitoring for EEEV, coinciding with their larger West Nile virus surveillance programs (Nelder et al., 2014). This followed a detection of EEEV in the province in 2009, from a positive pool of *Cs. melanura* mosquitoes (Nelder et al., 2014). Because *Cs. melanura* is a multivoltine species, it can undergo at least two generations per season, increasing opportunities for vector exposure to EEEV-positive hosts (Iranpour et al., 2009).

Canada is home to over 80 mosquito species, but their distribution across the provinces can vary greatly, and shifting environmental conditions, including rainfall and temperature, may alter local transmission dynamics (Ludwig et al., 2019). The net effect could be increased mosquito activity, longer activity periods, and the introduction of new species into Canada, such as *Aedes albopictus*, which has already been found to carry and experimentally transmit EEEV (Mitchell et al., 1993; Scott et al., 1990). Existing West Nile virus surveillance frameworks provide an established template to enhance EEEV monitoring across diverse mosquito populations (Todoric et al., 2022). Eastern migratory flyways that connect Canadian breeding grounds with southern overwintering foci may serve as important corridors for viral reintroduction via species such as the American Robin and Wood Thrush (Armstrong & Andreadis, 2022; Hobson et al., 2015; Oliver et al., 2020). Therefore, future spatiotemporal modelling will be important to understand how these migratory dynamics influence the re-emergence and dissemination of the virus across distant regions (Hort et al., 2023). Increased

concentrations of breeding passerines can also provide a large host reservoir for *Cs. melanura* mosquitoes, which can remain reproductively active late into the fall (Mahmood & Crans, 1998). Despite recent increases in reported EEEV activity and case counts, major operational gaps remain in Canadian vector surveillance (WDG Public Health, 2025). Specifically, vector surveillance is limited by our understanding of mosquito bloodmeals across host species, which is essential for determining feeding habits and identifying active bridge vectors (Nelder et al., 2014). Furthermore, conventional CDC light traps baited with CO₂ can exhibit low attraction rates for ornithophilic vectors such as *Cs. melanura*, leading to underestimation of primary local vector populations (Fagre et al., 2024; Nelder et al., 2014).

1.4 Viral Evolution and Phylodynamics

Technological advancements and the increasing availability of whole-genome sequencing have transformed the study of viral evolution and epidemiology. Researchers can compare genetic variation among viral genomes, enabling the reconstruction of evolutionary relationships, estimates of molecular evolution, patterns of geographic spread, and potential signatures of natural selection (Baele et al., 2016). Together, these methods form the basis of molecular epidemiology and phylodynamics, which provide powerful tools for understanding the ecological and evolutionary processes that shape viral populations (Featherstone et al., 2022). For arboviruses such as EEEV, this type of genomic analysis can offer valuable insight into transmission dynamics, viral dispersal, host associations, and the emergence of genetically distinct lineages.

1.4.1 RNA Virus Evolution

The Baltimore classification framework organizes viruses according to how genetic information is transferred from the viral genome to messenger RNA (mRNA), reflecting differences in genome type and organization (Baltimore, 1971; Koonin et al., 2021). Advances in genomic and metagenomic sequencing have since revealed extensive viral diversity and complex evolutionary relationships that are not fully represented by these classes alone (Koonin et al., 2021). Within this broader diversity, RNA viruses represent a particularly complex group and can possess positive-sense, negative-sense, or double-stranded RNA genomes. Their replication relies on using virally encoded RNA-dependent RNA polymerase (RdRp), which generally lacks proofreading capacity. This can lead to high mutation rates during replication, ranging from 10^{-4} to 10^{-6} mutations per nucleotide per cell infection, driven by variations in polymerase fidelity (Sanjuán et al., 2010). These high mutation rates can generate quasispecies, or large communities of closely related mutants, enabling rapid adaptation to changing environments and promoting antigenic drift that can complicate host immune responses (Domingo et al., 2021; Payne, 2017). However, a delicate balance must be maintained. As a subset of mutations becomes fixed in a viral population, excessive mutation can have deleterious effects, leading to reduced transmission and fitness (Patterson et al., 2020).

Arboviruses face unique evolutionary challenges, given their need to replicate in different hosts and the distinct selection pressures they face (Rozen-Gagnon et al., 2014). Despite high mutation potential, many arboviruses exhibit low sequence variability, as maintaining fitness across diverse host environments imposes strong purifying selection (Weaver et al., 2021). Furthermore, arboviruses encounter population bottlenecks during vector transmission that limit selection and allow deleterious mutations to accumulate via genetic drift faster than genetic

reversions can correct them, a process known as Muller's ratchet (Muller, 1964; Weaver et al., 2021). Additionally, for arboviruses such as EEEV that circulate in sylvatic transmission cycles, incidental hosts typically fail to develop sufficient viremia to enable onward transmission. Thus, while within-host selection may still occur, dead-end hosts contribute minimally to long-term lineage evolution (Weaver et al., 1991).

When measured at the population level, these combined mechanisms result in measurable evolutionary substitution rates. RNA viruses commonly exhibit high but variable rates of nucleotide substitution, often on the order of 10^{-3} to 10^{-4} substitutions/site/year (Jenkins et al., 2002). Interestingly, EEEV exhibits a slower evolutionary rate than many other arboviruses, which may reflect strong evolutionary constraints imposed by its long-standing enzootic transmission cycle between mosquito vectors and avian reservoir hosts (Tan et al., 2018). Specifically, EEEV displays a consistent, clock-like evolutionary substitution rate of approximately 1.7×10^{-4} substitutions/site/year (Tan et al., 2018). This is relatively slow compared to several other arboviruses such as West Nile virus with 4.9×10^{-4} substitutions/site/year (Bialosuknia et al., 2019), Western equine encephalitis virus with 2.8×10^{-4} substitutions/site/year (Bergren et al., 2014), and Chikungunya virus with 2.23×10^{-4} to 4.09×10^{-4} substitutions/site/year depending on the genotype (Hao et al., 2025). It is more comparable to, or slightly higher than, rates observed in Jamestown Canyon virus at 3×10^{-5} substitutions/site/year (Bourgikos et al., 2026), and Venezuelan equine encephalitis virus at 9.8×10^{-5} substitutions/site/year (Pisano et al., 2014).

1.4.2 Molecular Epidemiology and Phylogenetics

Whole-genome sequencing is a valuable tool for understanding how viruses are related to one another (Grubaugh et al., 2018). It captures the complete viral genome, enabling the

reconstruction of historical relationships and the genetic variation accumulated among viral sequences. By applying phylogenetic methods to these sequence data, researchers can construct evolutionary trees whose branching topologies reflect patterns of shared ancestry, divergence, and historical lineage diversification over time (Holmes, 2009; Lam et al., 2010; Tan et al., 2018; Wadas & Domingues, 2025). When integrated with epidemiological metadata—such as sampling date, host species, and geographic origin, these phylogenetic structures form the backbone of molecular epidemiology, offering key insights into viral transmission dynamics and the emergence of distinct lineages (Belák et al., 2009; Hill et al., 2021; Ritsch et al., 2023; Volz et al., 2013). For arboviruses such as EEEV, these approaches have provided valuable insights into the virus's overwintering capacity in New York (Young et al., 2008), patterns of dispersal along the eastern United States coastline (Tan et al., 2018), and the dynamics behind a 2019 outbreak that included 38 human cases (Hill et al., 2023).

1.4.3 Temporal Evolution and Phylodynamics

While standard phylogenetics describes the evolutionary relationships among viruses, phylodynamic approaches extend these relationships by integrating genomic data with spatiotemporal and epidemiological metadata, such as sampling dates and locations, to reconstruct viral evolutionary dynamics, circulation, and population changes over time (Dellicour et al., 2016). This approach can be especially well suited for arboviruses, where it can infer evolutionary and epidemiological events to help address challenges associated with incomplete sampling and surveillance gaps (Dellicour et al., 2016). An important requirement for any time-scaled evolutionary analysis is sufficient temporal signal within the dataset (Rambaut et al., 2016). This is a measurable relationship between genetic divergence and sampling time that demonstrates the clear accumulation of nucleotide substitutions over the sampling window

(Duchene et al., 2020). In heterochronous datasets, sufficient temporal sampling allows the use of sampling dates to calibrate molecular clock models, providing a framework for estimating viral evolutionary rates and divergence times (Biek et al., 2015; Holmes, 2003). Conversely, a weak temporal signal can produce unreliable evolutionary rate estimates and model misspecification that reflects random noise rather than true evolutionary structure (Rambaut et al., 2016). Molecular clock models convert genetic distances into time, dating internal phylogenetic nodes and enabling the reconstruction of when lineages diverged and major evolutionary events may have occurred (Ho & Duchêne, 2014). For EEEV, this temporal framework is particularly valuable for investigating how diversity is generated, maintained, and replaced across North America. Specifically, it can help determine whether EEEV detections within Canada reflect repeated introductions or sustained local persistence.

1.4.4 Phylogeography and Viral Dispersal

Phylogeography can be considered an extension of phylodynamics, in which phylogenetic relationships are combined with discrete traits, such as sampling location or host source, to investigate patterns of viral movement and dispersal (Faria et al., 2011). By placing time-calibrated phylogenies into a geospatial context, phylogeographic approaches can reveal how viral lineages move across landscapes and how spatial structure shapes viral diversity (Dellicour et al., 2020). For example, these analyses have been used to investigate the emergence and dispersal of Powassan virus within the United States (Vogels et al., 2023), track H5N1 avian influenza lineage replacement across North America (Signore et al., 2025) and assess the spread of major SARS-CoV-2 clades during the COVID-19 pandemic (Gómez-Carballa et al., 2020). Ultimately, tracking these long-distance dispersal events can provide important insight for epidemic response and outbreak preparedness (Mencattelli et al., 2023). In addition to

geographic reconstruction, discrete trait phylogeographic approaches treat metadata such as sampling location or host species as categorical traits that can be reconstructed across a phylogeny. These traits can then be used to reconstruct changes in sampled geographic or host states along a phylogeny (Hill et al., 2026; Zhao et al., 2023). This framework is particularly useful for arboviruses with limited surveillance, where historical transmission pathways must be inferred directly from sampled genomes. In Canada, where genomic data have historically been limited to a single sequence, discrete phylogeographic analysis can clarify whether viral detections result from repeated introductions into Canada or from sustained local persistence. Resolving this distinction is important for defining the regional ecology of EEEV and determining the mechanisms driving recurring viral detections across Canadian provinces.

1.4.5 Genetic Diversity and Population Structure

Within viral populations, patterns of genetic diversity reflect the combination of mutation, selection, genetic drift, transmission dynamics, and geographic structure. Next-generation sequencing (NGS) technologies have improved the characterization of these processes, providing important information for genomic surveillance, virulence assessment, antiviral monitoring, and vaccine strategies (Beerenwinkel et al., 2012; Posada-Céspedes et al., 2021). By evaluating synonymous and nonsynonymous substitution rates across coding regions, codon-based selection analyses can identify specific genomic regions that are subject to evolutionary constraint or that undergo purifying versus diversifying selection (Kosakovsky Pond & Frost, 2005). This is particularly important for RNA viruses like EEEV, which exhibit low genetic diversity, as even limited variation can highlight regions of functional constraint or sites of adaptation. Additionally, to complement phylogenetic trees, pairwise sequence identity and haplotype networks can provide finer-scale resolution of genomic variation and population

structure (Li et al., 2023). Pairwise comparisons provide quantitative measures of sequence similarity among viral genomes, coding regions, or specific genes (Muhire et al., 2014). Haplotype networks further enable exploration of viral population structure across broad landscapes by visualizing shared haplotypes and the mutational steps that separate them. By mapping these shared mutational networks, haplotype-based approaches can identify subtle patterns of genetic connectivity and regional clustering that may not be fully resolved by phylogenetic trees alone (Khan et al., 2025).

1.5 Genomic Surveillance Technologies

Advances in genomic surveillance technologies have become important to infectious disease monitoring, enabling the detection and characterization of viral pathogens across diverse sample types. While conventional diagnostics remain critical for confirmatory testing, they offer limited insight into viral diversity, evolution, and spatial dissemination. NGS addresses these gaps by recovering complete or near-complete viral genomes, providing the information needed to determine whether detections result from repeated introductions or from sustained local persistence. In Canada, where EEEV genomic data have historically been limited, expanding the current NGS capabilities is essential for tracking viral movement, informing public health and veterinary responses, and understanding pathogen emergence, even as diagnostic limitations remain.

1.5.1 Conventional Viral Detection and Surveillance

Traditional diagnostic and surveillance methods focus on determining whether a pathogen is present or absent in a sample. These approaches offer several benefits, including scalability, automation, rapid turnaround times, and relatively low costs compared with many

NGS methods (Dronina et al., 2021). Conventional diagnostic methods can be broadly divided into direct and indirect approaches (Cassedy et al., 2021). Direct methods target viral nucleic acids, antigens, or structural components, whereas indirect approaches require viral isolation and propagation in competent cell lines to observe cytopathic effects (CPE) or cell disruption (Cassedy et al., 2021; Pretorius & Venter, 2017). Although virus isolation can provide evidence of infectious virus, it is inherently labour-intensive, time-consuming, and requires a suitable growth medium (Pretorius & Venter, 2017). Additionally, observing CPE generally requires further testing to confirm the identity of the isolated virus (Leland & Ginocchio, 2007).

Direct immunoassay-based methods, including immunoblots, enzyme-linked immunosorbent assay (ELISA), and lateral flow immunoassay (LFIA), are designed to detect viral antigens in a sample (Leland & Ginocchio, 2007). Although cost-effective and scalable, these approaches are limited by antibody availability, long runtimes, and poor specificity among closely related strains. Furthermore, unlike nucleic acid amplification methods, they lack the flexibility needed to rapidly adapt should novel variants emerge (Boonham et al., 2014). In practice, these limitations restrict the insight provided by routine EEEV surveillance. For example, in Northern New England, cervid serosurveys conducted on deer and moose samples aimed to understand the geographic extent of EEEV activity but were unable to address the viruses evolutionary history (Mutebi et al., 2022). Similarly, Florida's approach relies on sentinel chicken seroconversion rates for EEEV surveillance; however, this method should be integrated with other approaches to support more robust, data-informed decision-making (Campbell et al., 2022). From a diagnostic standpoint, serological assays, such as immunoglobulin M (IgM) detection, remain important methods for detecting EEEV (Sahu et al., 1994). However, their utility is limited among vaccinated horses, as vaccination can influence IgM seroprevalence.

Finally, while tissue-based techniques such as immunohistochemistry can visualize localized viral presence, they lack the broad genomic or evolutionary context needed to trace viral dissemination (Pennick et al., 2012).

To overcome the constraints of immunoassay-based approaches, nucleic acid-based detection methods have become central to routine viral diagnostics and surveillance (Kumar et al., 2026). Among these is conventional PCR, which uses target-specific primers to amplify a genomic region of interest, prior to product visualization on a gel. Reverse-transcription quantitative PCR (RT-qPCR) expands on conventional PCR by using fluorescent dyes or sequence-specific probes to monitor amplification in real time, enabling rapid and highly sensitive detection of viral RNA (Faye et al., 2013). By leveraging multiple fluorophores, multiplex RT-qPCR assays can detect several viral targets or internal controls within a single reaction (Ding et al., 2020), while high-density microplate formats offer highly scalable and comparatively inexpensive screening (Xu et al., 2022). Consequently, RT-qPCR remains the gold standard for EEEV detection and is widely adopted for confirmatory testing (Lambert et al., 2003). Despite these advantages, targeted diagnostic assays are limited to the genomic regions for which they are designed (Carman et al., 1995). While effective for confirming the presence of a known viral pathogen, they provide limited information about whole-genome diversity, lineage relationships, or novel mutations. These limitations restrict their utility for detailed evolutionary and epidemiological investigations, highlighting the need for complementary NGS approaches.

1.5.2 Next-Generation Sequencing and Targeted Enrichment

Unlike conventional surveillance methods, NGS recovers whole-genome data, enabling detailed mutation characterization, lineage assignment, and evolutionary analyses (Hill et al., 2026). However, the effectiveness of NGS can be limited by poor sample integrity, low viral

target abundance, and high host or microbial genetic background (Wylie et al., 2015). These challenges can reduce sequencing coverage and genome completeness, which need to be overcome using enrichment strategies such as random amplification, targeted amplicon tiling, or probe-based hybridization (Johnson et al., 2016). Ultimately, the performance of each approach depends on the design specificity and sequence similarity between the viral target and the primers or probes used (Deng et al., 2020). Random-amplification methods, such as sequence-independent single-primer amplification (SISPA), use tagged random hexamers during cDNA synthesis, followed by primer-specific PCR, to amplify total RNA without requiring prior sequence knowledge (Chrzastek et al., 2017). Although versatile, random approaches lack target specificity and remain vulnerable to high host background. Conversely, targeted PCR methods generate overlapping amplicons that tile across a specific genomic region, offering a rapid, economical, and highly sensitive alternative (Quick et al., 2017); however, they may require periodic primer updates if target genomes undergo significant divergence (Freed et al., 2020). Alternatively, hybridization-based enrichment approaches use complementary biotinylated probes to capture viral nucleic acids prior to sequencing, enriching for pathogen-derived reads in samples with abundant host background or degraded material (Wylie et al., 2018). Compared to PCR approaches, probe-based methods can generally tolerate higher levels of sequence divergence because hybridization capture can accommodate some probe-target mismatches (Bonsall et al., 2015). However, this approach remains more costly, workflow-intensive and dependent on viral load (Calle-Prieto et al., 2024; Kantor & Jiang, 2024). Given the high sequence conservation among EEEV genomes, targeted PCR amplification could be an efficient and cost-effective strategy for recovering viral genomes from Canadian surveillance samples. Complementing existing diagnostic workflows rather than replacing them, this approach would

provide the genomic data needed for downstream phylogenetic, phylodynamic, and phylogeographic analyses to understand EEEV activity in Canada.

1.6 Objectives of Research

Advances in NGS technologies and the development of increasingly accessible methods for recovering viral genomic information have created new opportunities to investigate the genomic diversity, evolutionary history, and geographic dispersal of arboviruses such as EEEV across North America. However, substantial gaps remain in the Canadian EEEV landscape due to a lack of available complete genomes. Additionally, important questions remain regarding the movement of EEEV across North America, the frequency of viral introductions into Canada, and the potential for sustained regional persistence. To address these gaps, an EEEV-specific targeted sequencing workflow was developed and applied to recover and characterize complete viral genomes from Canadian surveillance samples, thereby expanding the genomic representation of EEEV in Canada. These genomic data were then used to reconstruct the evolutionary and geographic history of EEEV across North America and evaluate whether Canadian detections were consistent with repeated independent incursions, broader regional circulation, or sustained local persistence. The overall goal of this research was to apply genomic approaches to improve the characterization and interpretation of emerging arboviruses within Canadian surveillance systems, ultimately improving the understanding of arboviral activity, risk assessment, and outbreak preparedness to inform strategies that protect human and animal health.

1.6.1 Characterization of Complete EEEV Genomes from Canadian Surveillance Samples

Research Question: Can complete or near-complete EEEV genomes be characterized from Canadian surveillance samples to improve understanding of North American EEEV diversity?

To expand the limited genomic representation of Canadian EEEV, complete or near-complete genomes were generated from Canadian surveillance samples using a novel targeted tiling-amplicon sequencing workflow. Given the high genomic conservation of EEEV and the single previously available Canadian sequence, it was hypothesized that newly characterized Canadian genomes would be closely related to existing North American lineages while also capturing previously unsampled genetic variation among Canadian surveillance samples. The resulting genomes provided the foundation for subsequent comparative analyses of EEEV diversity and evolutionary relationships across North America.

1.6.2 Genetic Diversity and Population Structure of North American EEEV

Research Question: How are genetic diversity and population structure distributed among North American EEEV genomes, and where do Canadian sequences fit within this broader diversity?

To assess patterns of EEEV genetic diversity across North America, it was hypothesized that EEEV populations would exhibit strong evolutionary constraints due to long-term maintenance within established transmission cycles, while still retaining detectable geographic structure associated with regional circulation patterns. Given the limited availability of Canadian genomic data, Canadian sequences could represent either previously unsampled regional variation or genetic diversity embedded within broader North American lineages. To evaluate these patterns, sequence identity, codon-level selection, and haplotype relationships were characterized among North American EEEV genomes. These analyses determined whether Canadian viruses represented distinct genetic lineages or reflected broader patterns of EEEV diversity across North America.

1.6.3 Evolutionary History, Geographic Dispersal, and Host-State Associations of EEEV in North America

Research Question: What temporal, geographic, and host-state association patterns shape EEEV evolution in North America, and do Canadian detections reflect repeated introductions or sustained local persistence?

To assess the evolutionary history and geographic dispersal of EEEV across North America, it was hypothesized that recurring detections in Canada may reflect repeated introductions rather than sustained local persistence, with migratory avian hosts providing a plausible mechanism for long-distance virus dispersal. Bayesian phylogenetic and discrete-trait approaches were used to reconstruct temporal relationships and geographic and host-state associations among North American EEEV genomes. These analyses evaluated whether Canadian detections were consistent with repeated introductions, broader regional circulation, or sustained local persistence.

CHAPTER 2: MATERIALS AND METHODS

2.1 *In silico* EEEV tiling amplicon PrimalScheme and 5'/3' RACE primer design

To enrich for the complete coding sequence (CDS), an EEEV-specific tiling amplicon approach (Table 2.1) was designed using PrimalScheme v3.3.0 (<https://primalscheme.com/>) with a desired amplicon length of 550 bp, 2 non-overlapping primer pools, and annealing temperatures ranging from 60 °C – 65 °C. A reference genome (NCBI accession: EF151502.1) was used as the template reference. For the initial design, a total of 446 complete or near-complete EEEV genomes were obtained from the NCBI Virus database (02 March 2025) and uploaded to the PrimalScheme tool to optimize amplicon design and primer binding sites towards conserved primer-binding sites across the genome.

Table 2.1. Primer information for EEEV tiling amplicon sequencing.

Pool	Primer	Sequence	Amplicon Size
1	EEEV_1_LEFT	AGGCAACACCCTATTTCCACCTAT	556
	EEEV_1_RIGHT	AACCCGATCCAATAGATAGTTCGTACA	
	EEEV_3_LEFT	GGATTTACGGGAAAGTGGATAATCTTGC	573
	EEEV_3_RIGHT	CGGCCTCAGTAATAATTGGAGCG	
	EEEV_5_LEFT	CAGGACGATACAAGGTAGAGCCAT	552
	EEEV_5_RIGHT	CATTAGCAGCACTGAATCCACTGT	
	EEEV_7_LEFT	AGAGGGTGGGTAAACAACCTGCA	575
	EEEV_7_RIGHT	GGGCTATTATCCCAGTGTTTCATTGG	
	EEEV_9_LEFT	ATTAACGTGAGGACTCCCTACCG	516
	EEEV_9_RIGHT	GTGCGCTTTACCAGTTGCTACC	
	EEEV_11_LEFT	GTTATTCAACCACCGAGGGCAAG	501
	EEEV_11_RIGHT	GGATTCTTGCAGGTATGGGCACA	
	EEEV_13_LEFT	TGGAGGATTTCAGTCGTTTCGCC	570
	EEEV_13_RIGHT	TTGCAGTACTTGAATACATGGGAGC	
	EEEV_15_LEFT	CTACAAGACAAACCCGATAAGACTCAC	572
	EEEV_15_RIGHT	CGGATTTCATCATTGCCCGAA	
	EEEV_17_LEFT	GAAAGCAGTTGAATCACGCTACGA	572
	EEEV_17_RIGHT	GCACGCGTAACCATTCACCTGT	
	EEEV_19_LEFT	TGTGCGTCCTGGCCAATATCA	574
	EEEV_19_RIGHT	CCCACTGGCCTGAATTCTACCTTA	
	EEEV_21_LEFT	CGAGGAGAGGGCGACACTTTTA	576
	EEEV_21_RIGHT	GTTGTTCCACAGATAATTCAGCACTTG	

	EEEV_23_LEFT	CATTCATGTGGGGAGGAGCCTA	
	EEEV_23_RIGHT	TTCCCACTGCACAATTTTCTGCA	570
	EEEV_25_LEFT	TCATTTACATTCCACTTCTCCACTGC	
	EEEV_25_RIGHT	GCATTATGCGCTGCCTATAGTGT	561
2	EEEV_2_LEFT	ACTGACTCAACTTGCAGGTACCA	
	EEEV_2_RIGHT	AGACCCTCTCCCCTCTTAACGT	571
	EEEV_4_LEFT	AAGAAGGTGCCCCGCCGTATT	
	EEEV_4_RIGHT	GTTGCACTTTCACTCAGAGCCTG	554
	EEEV_6_LEFT	TGGTGGTCAGTGCAAAGAAAGAAA	
	EEEV_6_RIGHT	ACCCCTTTTCTAGTAAGCCCCTG	553
	EEEV_8_LEFT	CGACATCGACAGTGGGTTGTTCT	
	EEEV_8_RIGHT	GTTCAAGTGGTCCACTGCTTTTCT	559
	EEEV_10_LEFT	TGTGTGGCGCTCTTTACAGGAA	
	EEEV_10_RIGHT	GCGTAAATTTTCAGCTATGTCTTTGGC	556
	EEEV_12_LEFT	TGAAAGATTGGTGCCTAGGCGA	
	EEEV_12_RIGHT	TGGTTTGATCGTGTCTGCGCT	541
	EEEV_14_LEFT	CTACTGCAGGGGATAGGCTCATAC	
	EEEV_14_RIGHT	AAAAGGGCAGCTGCTTTGGG	569
	EEEV_16_LEFT	AACATCTGTGCACTTACCAACAGG	
	EEEV_16_RIGHT	TGACCTCTTATGTGTTTAAAGTTGCTAACT	573
	EEEV_18_LEFT	CCTAAACCAGGTAAACGACAGCG	
	EEEV_18_RIGHT	TGGCATGCAGGGTGGTTGAT	567
	EEEV_20_LEFT	GGGACACGGTTACAGTTGGGTT	
	EEEV_20_RIGHT	GTCCGGGTGCAGGTGTAAAATC	545
	EEEV_22_LEFT	CCCAATACTCCTGGCGTTACTTTG	
	EEEV_22_RIGHT	CACTCTTCCGAGCGCTCTACAT	558
	EEEV_24_LEFT	CTTCGAACGATGGAAAAAGGACAAAG	
	EEEV_24_RIGHT	TGCTGATGTTTCCTCCTACCAGC	541

To recover the 5-prime (5') and 3-prime (3') untranslated regions (UTR) of the EEEV genome not captured by the tiling amplicon scheme, gene-specific rapid amplification of cDNA ends (RACE) primers were designed to bind to conserved regions of the EEEV genome approximately 1 kb downstream the 5' end and 1.3 kb upstream the 3' end for use in a RACE PCR (Table 2.2). The same 446 complete or near-complete EEEV genomes used for the PrimalScheme design were aligned using MAFFT v7.511 (Katoh & Standley, 2013) and the resulting alignment was visualized in Geneious Prime v2025.0.3. Primers were manually designed according to the manufacturer's instructions for the SMARTer RACE 5'/3' Kit with a target T_m of 68 °C (Takara Bio, USA).

Table 2.2. Primer sequences for EEEV 5'/3' RACE.

Primer Name	Sequence	T _m (°C)	Product Size
EEEV_5_RACE	CTTCTGGGCGTCATCGACACTGACG	67.4	~1kb
EEEV_3_RACE	TCCACTGTCATCTGCGTGGTCCCC	68.3	~1.3kb

2.2 RNA extraction, cDNA synthesis, PCR, and RACE

A total of 41 EEEV-positive samples, comprising equine (n = 28), avian (n = 7), mosquito (n = 4), white-tailed deer (n = 1), and wolf (n = 1), were submitted to the Canadian Food Inspection Agency's National Centre for Foreign Animal Disease. Initial EEEV-positive status was confirmed by submitting laboratories via RT-qPCR prior to sample receipt, with corresponding diagnostic Ct values provided in Table 3.1. Samples were received either as pre-extracted total RNA (n = 18) or as unextracted diagnostic specimens (n = 23), which included brain and spleen tissue, or swabs. For the unextracted material, total RNA was extracted using an in-house standardized protocol in which samples were initially homogenized in TRI Reagent® (Zymo Research, USA) using the BeadBug™ 6 Microtube Homogenizer (Benchmark Scientific, USA) and BeadBug™ prefilled silica bead tubes (Benchmark Scientific, USA) according to the manufacturer's instructions. Homogenized samples were centrifuged, and the supernatant was then used for subsequent RNA extraction using the Direct-zol™ RNA MiniPrep Plus Kit (Zymo Research, USA) and an on-column DNase I treatment, following the manufacturer's instructions. The eluted RNA was quantified using the Qubit RNA HS Assay Kit on the Qubit 4.0 Fluorometer (Thermo Fisher Scientific, USA) and stored at -80 °C prior to cDNA synthesis and downstream PCR amplification.

First-strand cDNA synthesis was performed using the SuperScript™ IV First-Strand Synthesis System (Invitrogen, USA), and second-strand synthesis was performed using the NEBNext® Ultra II Non-Directional RNA Second Strand Synthesis Module (New England

Biolabs, USA), both according to the manufacturers' instructions. The cDNA was finally quantified using the Qubit dsDNA 1X HS Assay Kit on the Qubit 4.0 Fluorometer (Thermo Fisher Scientific, USA). The resulting cDNA was then used for the multiplex PrimalScheme PCR with Q5® High-Fidelity 2X Master Mix (New England Biolabs, USA). Each 25 µL reaction contained 9.5 µL cDNA template, 3 µL of multiplex primer pool (Pool 1 or Pool 2; 10 µM stock concentration, yielding a 1.2 µM final concentration), and 12.5 µL Q5 2X Master Mix. The PCR was performed under the following conditions: initial denaturation at 98 °C for 30 s; 35 cycles of 98 °C for 10 s, 54 °C for 30 s, and 72 °C for 30 s; followed by a final extension step at 72 °C for 2 min. Equal volumes of Pool 1 and Pool 2 reaction products per sample were combined (50 µL total), purified via a 1X SPRI bead cleanup (Beckman Coulter, USA) and quantified on the Qubit 4.0 Fluorometer (Thermo Fisher Scientific, USA) prior to downstream library preparation and sequencing.

To recover UTR terminal regions, extracted RNA was also subjected to 5' and 3' RACE using the SMARTer® RACE 5'/3' Kit (Takara Bio, USA) according to the manufacturer's protocol and using "Program 2" PCR with the following conditions: initial denaturation at 94 °C for 60 s; followed by 20 cycles of 94 °C for 30 s, 68 °C for 30 s, 72 °C for 3 min; followed by a final extension step at 72 °C for 5 min. Equal volumes of 5' and 3' RACE products were pooled for each sample (100 µL total), purified with a 1X SPRI bead cleanup, and quantified prior to downstream library preparation and sequencing.

2.3 Next-generation sequencing and genome assembly

Sequencing libraries for both the amplified EEEV PrimalScheme and the 5' and 3' RACE products were prepared using the Native Barcoding Sequencing Kit (SQK-NBD114.96; Oxford Nanopore Technologies, UK) according to the manufacturer's instructions and sequenced on an

Oxford Nanopore Technologies GridION instrument using an R10.4.1 SpotON Flow Cell (FLO-MIN114), targeting a final loading concentration of 100 fmol.

Raw Nanopore sequencing data were re-basecalled using Dorado v0.9.5 (<https://github.com/nanoporetech/dorado>) with the super-accurate model dna_r10.4.1_e8.2_400bps_sup@v5.0. The basecalled reads were then demultiplexed and converted from BAM to FASTQ using bedtools v2.27.1 (Quinlan & Hall, 2010). The demultiplexed reads were trimmed for adapters and barcodes using Porechop_ABI v0.5.0 (Bonenfant et al., 2023), followed by read quality and length filtering using Chopper v0.9.2 (De Coster & Rademakers, 2023), with a minimum mean quality threshold of Q10 and a minimum length cutoff of 50 bp. The trimmed and filtered reads underwent another round of trimming to remove primer sequences from the PrimalScheme and RACE PCR reactions using Cutadapt v5.0 (Martin, 2011) with three iterative rounds of trimming using -a and -g flags targeting both standard and non-internal 5' and 3' primer sequences. The cleaned Nanopore reads were aligned against an EEEV reference genome (NCBI accession: KU840359.1) using Minimap2 v2.28-r1209 (Li, 2018). The resulting alignment files were imported into Geneious Prime v2025.0.3 for manual coverage inspection and consensus sequence generation using a 75% majority threshold and a minimum 15× coverage depth cutoff required to call a base at each position. Positions failing to meet the 15× coverage threshold were masked as “N”. Manual inspection was conducted to confirm variant calls, assessing strand support, homopolymer regions, and potential primer artifacts.

2.4 EEEV dataset construction and phylogenetic analysis

To construct the whole-genome dataset, a complete set of available EEEV (Taxonomy ID: 11021) sequences were downloaded from the NCBI Virus database on 29 April 2025.

Duplicate accessions were removed prior to sequence filtering. Sample metadata can be found in Supplementary Table S1.4. Sequences were removed from the dataset if they met any of the following criteria: were less than 10 kb in length, contained >50% ambiguous bases (N's), or were not from North America or the Caribbean (retained due to their close phylogenetic relationship to North American EEEV lineages). Sequences for which required metadata (county/state/province, collection year, source host) could not be obtained were also excluded from subsequent analysis. The final whole-genome dataset contained 584 complete or near-complete genomes, including 41 newly generated Canadian genomes and 543 publicly available genomes. The public dataset comprised sequences from Canada (n = 1), the USA (n = 537), Mexico (n = 3), Jamaica (n = 1), and the Dominican Republic (n = 1).

The phylogenetic position of Canadian EEEV sequences was assessed using 584 complete or near-complete genomes. All full-genome sequences were aligned with MAFFT v.7.511 (Kato & Standley, 2013). Maximum-likelihood phylogenetic reconstruction was performed using IQ-TREE v3.0.1 (Wong et al., 2026). The optimal model of sequence evolution was selected using the Bayesian information criterion (BIC) score (Schwarz, 1978) calculated by ModelFinder (Kalyaanamoorthy et al., 2017), which identified GTR + F + R3 as the best-fit model for the dataset. Maximum-likelihood analyses were performed using the default settings, with all model parameters estimated by the program. Node support was assessed using ultrafast bootstrap (UFBoot) (Minh et al., 2013) with 1,000 replicates. The resulting phylogenetic tree was visualized with FigTree v.1.4.4 (<https://github.com/rambaut/figtree>). Node labels with UFBoot support < 70% were removed to reduce visual clutter, and major non-Canadian U.S. monophyletic groups were collapsed to highlight the phylogenetic placement of Canadian sequences.

To assess the temporal signal and clock-like behaviour of the phylogeny, the maximum-likelihood tree was imported into TempEst v1.5.3 (Rambaut et al., 2016) using heuristic residual mean squared rooting. Sampling dates specified by year were assigned to the calendar-year midpoint (YYYY.5). The correlation coefficient, coefficient of determination, regression slope, and residual mean square were used to assess the strength of the temporal signal. Visual inspection of residuals revealed no major anomalies, and no sequences were removed from the analysis.

2.5 Time-scaled Bayesian phylogenetic inference

To investigate the evolutionary dynamics, spatiotemporal dissemination patterns, and host-associated evolutionary patterns, Bayesian phylogenetic inference was performed using BEAST v1.10.4 (Drummond et al., 2012). The nucleotide substitution process was modelled with a General Time Reversible (GTR) matrix and estimated base frequencies. To account for rate heterogeneity across sites, a four-category discrete Gamma distribution plus invariant sites (GTR + G4 + I) was used. Demographic history and population fluctuations over time were modelled under a Coalescent Extended Bayesian Skyline Plot (EBSP) tree prior with a linear model, initiated from a random starting tree (Heled & Drummond, 2008). Lineage-specific evolutionary rates were constrained under a global Strict Clock model calibrated using sampling dates assigned to each tip, which was selected based on prior model testing and the strong temporal signal observed across the dataset. Discrete ancestral state reconstructions were inferred for two independent trait partitions: host and geography. For each discrete partition, an asymmetric trait substitution model was implemented to allow unequal transition rates between states, and ancestral states were reconstructed across all nodes. To identify a sparse set of supported transition pathways and minimize model over-parameterization, Bayesian Stochastic

Search Variable Selection (BSSVS) was applied to both discrete traits (Lemey et al., 2009). Five independent Markov chain Monte Carlo (MCMC) chains (Yang & Rannala, 1997) were run for 100 million generations, with parameters and trees sampled every 1,000 generations. Four convergent chains were combined using LogCombiner v1.10.4 after removing the first 10% of each run as burn-in. Convergence and mixing were evaluated in Tracer v1.7.2 (Rambaut et al., 2018), ensuring that effective sample sizes (ESS) exceeded 200 for all continuous parameters. The post-burn-in posterior distribution of trees was summarized as a single Maximum Clade Credibility (MCC) tree using TreeAnnotator v1.10.4, with keep target heights specified. The resulting MCC tree and ancestral state reconstructions were visualized in FigTree v.1.4.4 (<https://github.com/rambaut/figtree>).

A final dataset was established by excluding three novel Canadian sequences (Table 3.1; 2004_EEEV_NP_Horse_1–3) due to missing geographic location metadata. Because these three genomes lacked province-level collection information, their inclusion would introduce geographic ambiguity into the discrete trait model, potentially skewing results. Consequently, the final BEAST dataset ($n = 581$) comprised 38 novel Canadian sequences alongside 543 reference strains.

2.6 Spatiotemporal dissemination and host-transition network analysis

To evaluate spatiotemporal dissemination patterns and host-state transitions of EEEV, discrete ancestral-state reconstructions inferred from BEAST v1.10.4 were analyzed. Two discrete trait partitions were defined using alignment metadata: host species and geographical location. Geographic traits were categorized at the state or provincial level where applicable, while sequences from Mexico, Jamaica, and the Dominican Republic were analyzed at the country level. The host trait partition comprised eight ecological categories (mosquito, bird,

horse, human, alpaca, dog, deer, wolf). For each discrete partition, an asymmetric trait substitution model was implemented to allow direction-specific transition rates between states. Spatiotemporal dissemination histories and host-transition networks were visualized and statistically evaluated using SpreaD3 v0.9.7 (Bielejec et al., 2011). The post-burn-in tree and log files from the combined MCMC replicates were parsed to calculate posterior inclusion probabilities and compute Bayes Factors (BF) for all discrete transition routes.

To characterize the spatial dynamics and host-association patterns of EEEV, discrete lineage movement among states, provinces, and countries, as well as transitions between host categories, were reconstructed over time. All reconstructed transition pathways supported by a Bayes factor (BF) ≥ 10 were extracted and reported in Supplementary Tables S1.1 and S1.2, with BF ≥ 10 and BF ≥ 100 interpreted as strong and decisive support, respectively. To highlight well-supported evolutionary linkages (Figure 3.7), a dual stringency threshold requiring both posterior inclusion probability (PP) ≥ 0.70 and BF ≥ 10 was applied. Geographic diffusion pathways were rendered as interactive D3 HTML browser-based timelines, with line widths scaled by the posterior mean transition-rate parameter conditional on pathway inclusion and statistically supported host-transition networks were extracted to summarize key host-state transition dynamics.

2.7 Sequence variation and selection pressure analysis

To evaluate sequence similarity, positional variability, and codon-specific selective pressures across the EEEV genome, pairwise nucleotide identity, site-specific Shannon entropy, and codon-based selection analyses were conducted. Whole-genome sequences were used for pairwise identity and entropy analyses, whereas coding sequences corresponding to the non-

structural polyprotein regions (nsP123 and nsP4) and the structural polyprotein region (SP) were analyzed separately for codon-based selection.

Pairwise nucleotide identity was computed from the whole-genome alignment of 584 EEEV sequences using a custom Python 3 script utilizing Biopython v1.81 (Cock et al., 2009), Numba v0.57 (Lam et al., 2015), NumPy v1.24 (Harris et al., 2020), Pandas v2.0 (McKinney, 2010), and Matplotlib v3.7 (Hunter, 2007). For each unique pairwise comparison, identity was calculated as the proportion of matching nucleotide positions at sites containing unambiguous bases in both sequences. Alignment gaps, ambiguous bases, and positions with missing data were excluded. Pairwise identity values were visualized as a sequence-by-sequence heatmap, with colour intensity reflecting nucleotide identity. Pairwise identity values from all unique sequence comparisons were subsequently binned using a fixed bin width of 0.05% identity and plotted as a frequency distribution histogram in R v4.5.0 to assess overall genomic conservation across the dataset.

Site-specific nucleotide entropy across the EEEV genome alignment was performed using the Los Alamos HIV database Shannon Entropy-One tool (http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one.html). Sequences were aligned with MAFFT v7.511 (Katoh & Standley, 2013), and positions containing gaps or ambiguities were excluded from the calculation. Entropy values were plotted against the reference EEEV strain FL93-939 (GenBank accession: U01034.1) to identify conserved genomic regions and localized hot-spots of sequence variation.

To identify codon sites under selective pressure, Fixed Effects Likelihood (FEL) analysis was performed on the Datamonkey web server (Kosakovsky Pond & Frost, 2005; Weaver et al., 2018). Coding sequences corresponding to the non-structural polyprotein regions (nsP123 and

nsP4) and the structural polyprotein region (SP) were manually extracted from annotated whole-genome sequences in Geneious Prime v2025.0.3 and using MAFFT v7.511 (Katoh & Standley, 2013). Terminal stop codons were removed prior to analysis, while alignment gaps, ambiguous codons, and missing data were handled natively according to Datamonkey's default filtering criteria. Codon positions were numbered sequentially relative to the first codon of each individual polyprotein region (nsP123, nsP4, and SP). FEL models were fitted independently to each polyprotein alignment to estimate rates of synonymous (α) and nonsynonymous (β) substitutions. Selection pressure was quantified as the difference between these rates ($\beta - \alpha$), with values below 0 indicating purifying selection ($\beta < \alpha$) and values above 0 indicating diversifying selection ($\beta > \alpha$). Statistical significance was assessed at $P < 0.05$.

2.8 Population genetic structure and haplotype networks

To resolve evolutionary relationships and evaluate population genetic structure, a haplotype network was constructed. A dataset of 581 EEEV genomes (with 3 Canadian genomes removed as described in Section 2.5) was aligned with MAFFT v7.511 (Katoh & Standley, 2013) and imported into PopART v1.7 (Leigh & Bryant, 2015) using default filtering settings. Nucleotide sites with more than 5% undefined states (alignment gaps or ambiguous bases) were masked during dataset parsing. No additional sequences were excluded from the final network. The haplotype network was reconstructed using the Templeton-Crandall-Sing (TCS) statistical parsimony method (Templeton et al., 1992), which is well suited for visualizing evolutionary relationships among closely related haplotypes. Alignment summary statistics, including the number of segregating (polymorphic) sites, parsimony-informative sites, haplotype counts, and overall nucleotide diversity (π), were calculated in PopART v1.7. Overall sequence identity percentage was derived from the mean pairwise distance of the alignment.

For spatial analysis, the dataset was divided into six broad geographic categories: Canada, Northeastern U.S., Southeastern U.S., Midwestern/Interior U.S., Gulf Coast U.S., and Mexico and Caribbean (Jamaica, Dominican Republic). Network nodes were displayed as pie charts coloured by the geographic composition of each haplotype, with node size proportional to haplotype frequency. Mutational steps separating haplotypes were indicated along connecting lines as hatch marks.

CHAPTER 3: RESULTS

3.1 Dataset Characteristics and Novel Canadian EEEV Genomes

To establish a comprehensive spatiotemporal framework for evaluating EEEV evolution, a genomic dataset consisting of 584 total sequences was assembled, incorporating both publicly available genome sequences from NCBI Virus ($n = 543$; Supplementary Table S1.4) and our novel Canadian EEEV genomes ($n = 41$; Table 3.1). Sample collection locations revealed a broad geographic footprint spanning key enzootic and epizootic zones (Figure 3.1a).

Geographically, the dataset primarily consisted of sequences originating from the United States ($n = 537$, 92%), with a large proportion of genomes from the states of New York ($n = 211$, 36.1%) and Florida ($n = 112$, 19.2%). The addition of newly characterized sequences substantially expanded the representation of Canadian occurrences ($n = 42$, 7.2%; 41 novel and one previously available) (Figure 3.1b). Ecologically, the dataset demonstrated a diverse vector-host distribution across the eight source hosts (Figure 3.1c). EEEV detections were heavily dominated by mosquito vectors ($n = 357$, 61.1%), with bird reservoirs ($n = 71$, 12.2%) and incidental dead-end hosts, including humans ($n = 25$, 4.3%) and horses ($n = 125$, 21.4%), accounting for the majority of the remaining EEEV cases.

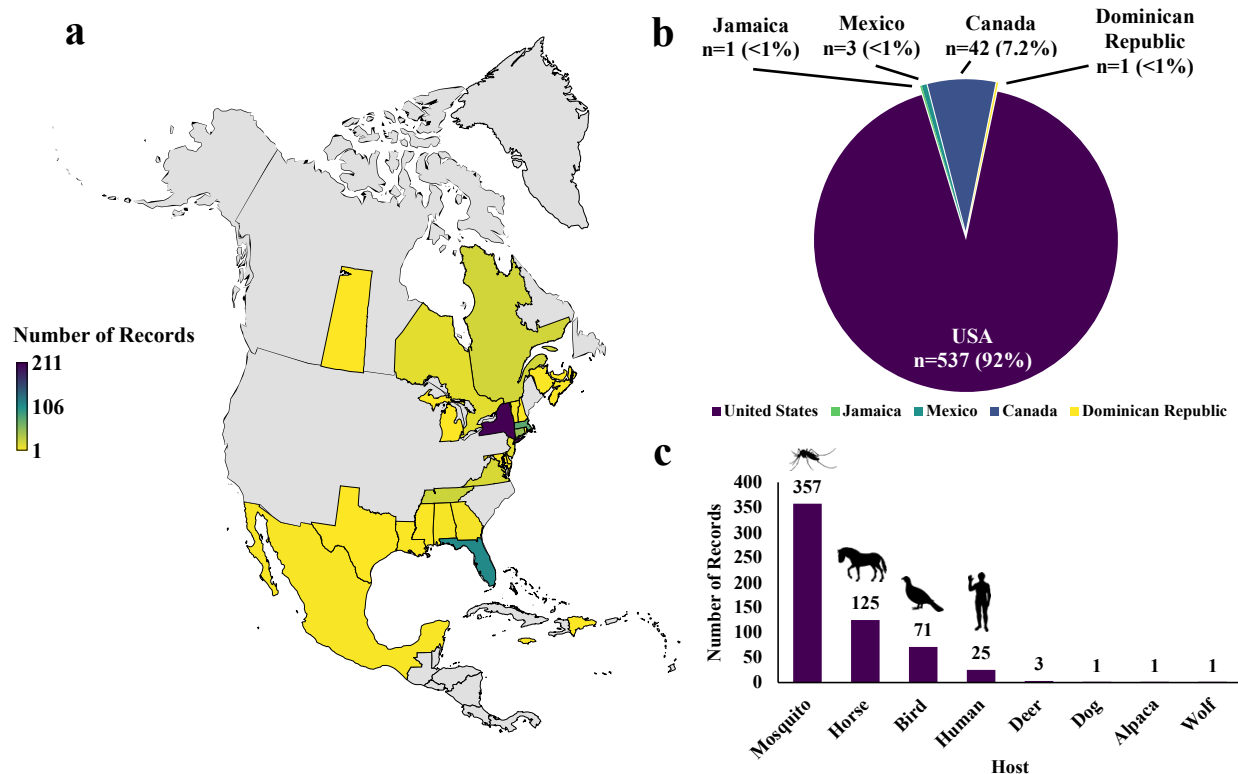


Figure 3.1. Geographical and ecological distribution of North American EEEV sequences. (a) Geographic map of sample collection locations in North America. (b) Total sequence (n = 584) distribution by country of origin (n = 5). (c) Ecological distribution of viral sequences across eight distinct host and vector categories.

To address critical surveillance gaps in the Canadian EEEV landscape, a newly designed, targeted amplicon sequencing approach was applied to collected surveillance samples. The targeted amplicon and RACE sequencing workflow recovered 41 novel Canadian EEEV genomes across multiple host categories and diagnostic Ct values ranging from 13.4–34.2 (Table 3.1). Sequencing yielded an average of approximately 650,000 reads per sample with a mean depth of coverage of approximately 58,000x (median: 28,339x; IQR: 18,981x–92,712x). The generated consensus sequences spanned the complete genome length up to 11,703 nucleotides. Relevant sample metadata, including the specific collection year, sampling locations, and source hosts for these newly characterized genomes, are detailed in Table 3.1.

Table 3.1 Novel Canadian EEEV genome epidemiological metadata.

Strain ID	Location	Collection Year	Host	Ct Value
2004_EEEV_QC_Horse_1	QC	2004	Horse	13.4
2004_EEEV_QC_Horse_2	QC	2004	Horse	13.5
2004_EEEV_QC_Horse_3	QC	2004	Horse	26.2
2004_EEEV_QC_Horse_4	QC	2004	Horse	14.3
2004_EEEV_QC_Horse_5	QC	2004	Horse	14.5
2004_EEEV_QC_Horse_6	QC	2004	Horse	19.8
2004_EEEV_QC_Horse_7	QC	2004	Horse	13.5
2004_EEEV_QC_Horse_8	QC	2004	Horse	15.3
2004_EEEV_QC_Horse_9	QC	2004	Horse	20.1
2004_EEEV_QC_Horse_10	QC	2004	Horse	17.1
2004_EEEV_NP_Horse_1*	Not Provided	2004	Horse	17.4
2004_EEEV_NP_Horse_2*	Not Provided	2004	Horse	18.0
2004_EEEV_NP_Horse_3*	Not Provided	2004	Horse	18.2
2018_EEEV_ON_Horse_1	ON	2018	Horse	17.0
2020_EEEV_ON_Horse_1	ON	2020	Horse	18.2
2020_EEEV_ON_Horse_2	ON	2020	Horse	20.1
2021_EEEV_ON_Horse_1	ON	2021	Horse	22.3
2022_EEEV_ON_Bird_1	ON	2022	Bird	32.9
2023_EEEV_QC_Horse_1	QC	2023	Horse	21.9
2023_EEEV_NB_Bird_1	NB	2023	Bird	25.6
2023_EEEV_QC_Horse_2	QC	2023	Horse	26.4
2023_EEEV_QC_Horse_3	QC	2023	Horse	16.4
2023_EEEV_ON_Horse_1	ON	2023	Horse	21.7
2023_EEEV_ON_Horse_2	ON	2023	Horse	19.3
2024_EEEV_QC_Deer_1	QC	2024	Deer	19.7
2024_EEEV_QC_Bird_1	QC	2024	Bird	26.0
2024_EEEV_SK_Horse_1	SK	2024	Horse	Not Provided
2024_EEEV_QC_Horse_1	QC	2024	Horse	23.8
2024_EEEV_ON_Horse_1	ON	2024	Horse	16.4
2024_EEEV_ON_Horse_2	ON	2024	Horse	18.2
2024_EEEV_ON_Horse_3	ON	2024	Horse	15.3
2024_EEEV_ON_Bird_1	ON	2024	Bird	29.0
2024_EEEV_QC_Bird_2	QC	2024	Bird	23.9
2024_EEEV_NS_Horse_1	NS	2024	Horse	15.2
2024_EEEV_QC_Wolf_1	QC	2024	Wolf	26.5
2024_EEEV_ON_Bird_2	ON	2024	Bird	26.5
2024_EEEV_QC_Bird_3	QC	2024	Bird	26.2
2024_EEEV_NB_Mosquito_1	NB	2024	Mosquito	22.6

2025_EEEV_NS_Mosquito_1	NS	2025	Mosquito	34.0
2025_EEEV_NS_Mosquito_2	NS	2025	Mosquito	31.5
2025_EEEV_PEI_Mosquito_1	PEI	2025	Mosquito	34.2

*NP = Not Provided

Pairwise nucleotide identity was high among the EEEV genome dataset (Figure 3.2). Most unique sequence comparisons exceeded 99% identity, with the average near 99.1% (Figure 3.2a). Comparisons with pairwise nucleotide identities below 98.5% were relatively uncommon, accounting for a small minority of total sequence pairs. These lower identity values were primarily driven by temporal divergence involving early historical strains and reference genomes, such as the 1935 New Jersey strain (KJ469566.1), the 1940 Maryland PE6 reference strain (AY722102.1), and other 1950s–1970s sequences. Additionally, the 1978 Dominican Republic strain (KJ469636.1) exhibited pairwise nucleotide identity values ranging from 97.74% to 99.10% across the dataset. The heatmap similarly shows the same overall pattern, with most pairwise comparisons displaying high nucleotide identity across the alignment (Figure 3.2b). Taken together, these results demonstrate that North American EEEV genomes are highly conserved despite the nearly 100-year sampling period.

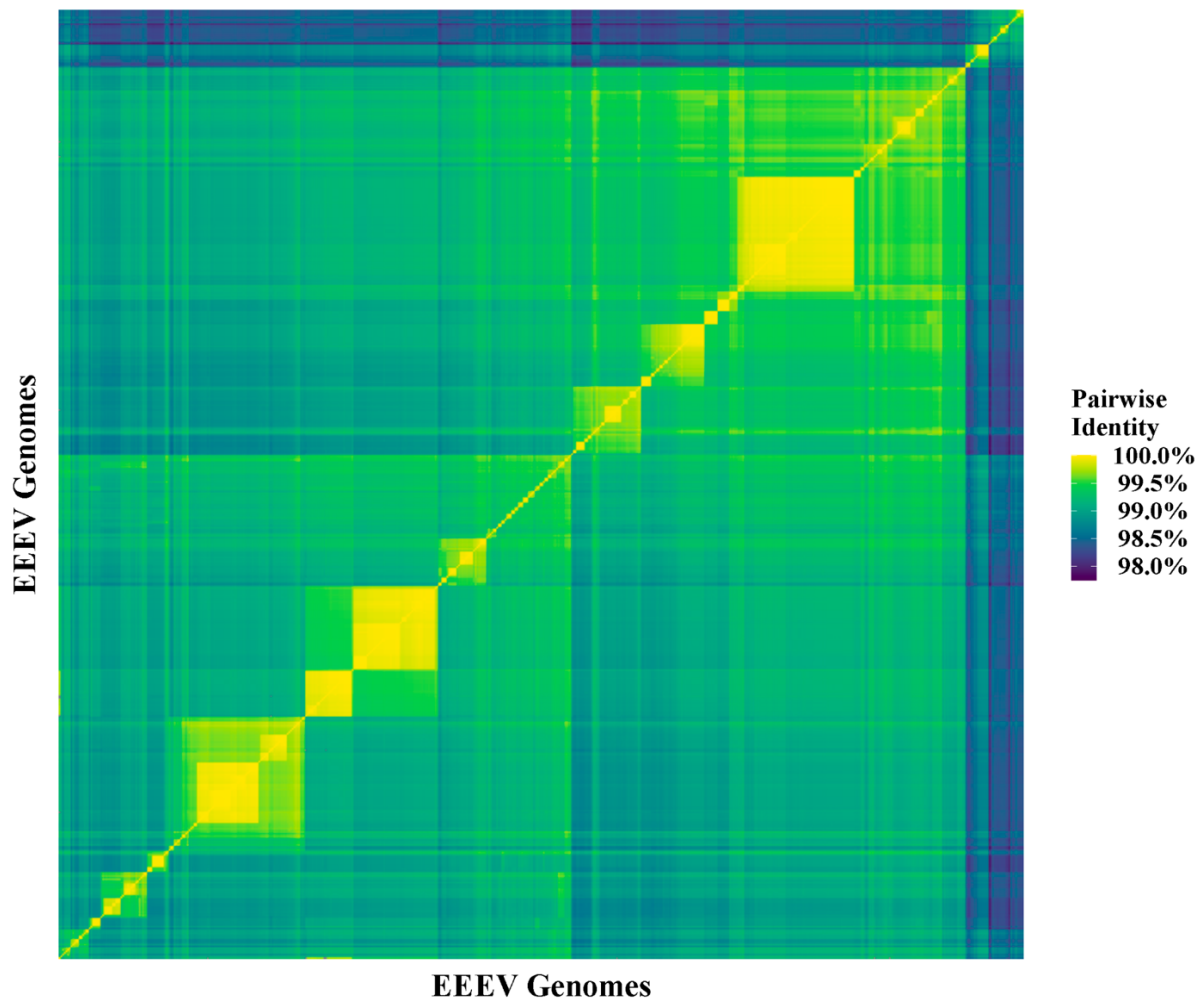
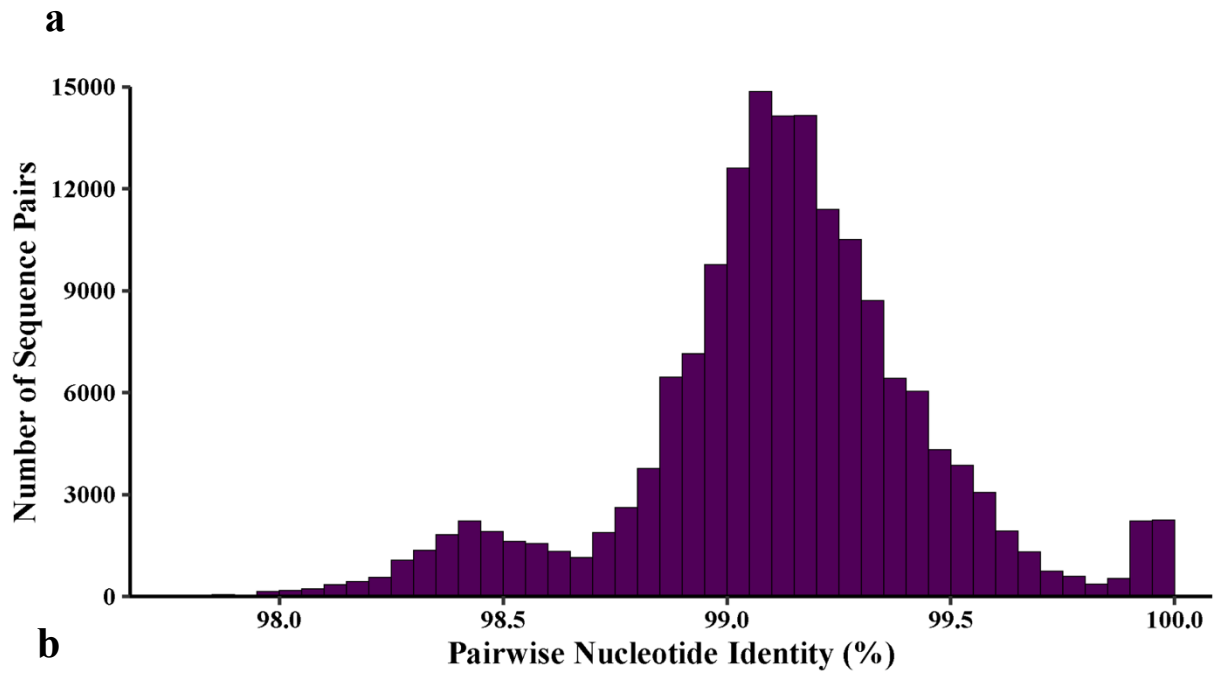


Figure 3.2. Pairwise nucleotide identity among North American EEEV genomes. (a) Distribution of pairwise nucleotide identity values across all unique sequence comparisons. (b) Heatmap showing pairwise nucleotide identity between aligned EEEV genome sequences. Identity was calculated as the proportion of matching positions among sites containing unambiguous nucleotides in both sequences. Alignment gaps, ambiguous nucleotide characters, self-comparisons, and reciprocal duplicate comparisons were excluded.

3.2 EEEV Genome-Wide Variability and Selection Pressure

We next wanted to investigate the mechanisms influencing the low diversity observed among EEEV genomes. To assess positional variation, site-specific Shannon entropy was calculated across the dataset (Figure 3.3a). Entropy values were low across the genome (mean entropy: 0.028 bits; median: 0 bits), with 61.3% of nucleotide positions exhibiting complete sequence conservation ($H = 0$ bits). The maximum entropy observed at any single position was 0.722 bits. While low entropy was maintained relatively consistently across both non-structural and structural coding regions, localized clusters of elevated entropy were concentrated near the extreme 3' UTR due to non-coding variation and reduced sequence coverage at the terminal genome end. These results reflect low genomic diversity and strong evolutionary constraints across the EEEV genome, thereby maintaining sequence stability.

To characterize the pattern of evolutionary constraints across coding regions, Fixed Effects Likelihood (FEL) analysis was used to infer synonymous (α) and non-synonymous (β) substitution rates across the EEEV genome (Figure 3.3b). Of the 3,735 codons tested across the open reading frames, 899 sites (24.07%) showed statistically significant selection ($P < 0.05$). Most of these significant positions (895 sites, 99.56% of significant sites; 23.96% of all tested codons) were under purifying selection ($\beta < \alpha$). Only four codon positions (0.44% of significant sites; 0.11% of all tested codons) exhibited significant diversifying selection ($P < 0.05$, $\beta > \alpha$). Two of these sites are located within the non-structural polyprotein regions. The first occurs at

nsP123 codon position 691 ($\beta - \alpha = +2.078$, $P = 0.0422$) within nsP2, driven by a non-synonymous substitution changing aspartic acid to alanine ($D \rightarrow A$) in a small subset of sequences ($n = 4$). The second is found at nsP123 codon position 1703 ($\beta - \alpha = +1.907$, $P = 0.0402$) within nsP3, in which aspartic acid is predominant and mutates into glutamic acid, tyrosine, glycine, or asparagine ($D \rightarrow E/Y/G/N$) in a minority of sequences ($n = 6$). The remaining two diversifying sites are found in the structural polyprotein region (SP). Here, SP codon positions 789 ($\beta - \alpha = +6.743$, $P < 0.0001$) and 1149 ($\beta - \alpha = +1.728$, $P = 0.0294$) are driven by a non-synonymous substitution in which the predominant alanine is changed to valine ($A \rightarrow V$) in a small subset of samples ($n = 21$) and alanine into multiple alternative residues, including threonine, serine, or valine ($A \rightarrow T/S/V$), in a few remaining samples ($n = 5$), respectively. Conversely, strong purifying selection was observed at two codon positions: 1246 ($\beta - \alpha = -21.776$, $P < 0.0001$) within the nsP123 polyprotein and 236 ($\beta - \alpha = -17.585$, $P = 0.0001$) within the nsP4 polyprotein, indicating sites under particularly strong evolutionary pressure. Overall, these results demonstrate that EEEV is subject to widespread evidence of purifying selection across the genome, with only a few sites exhibiting positive diversification.

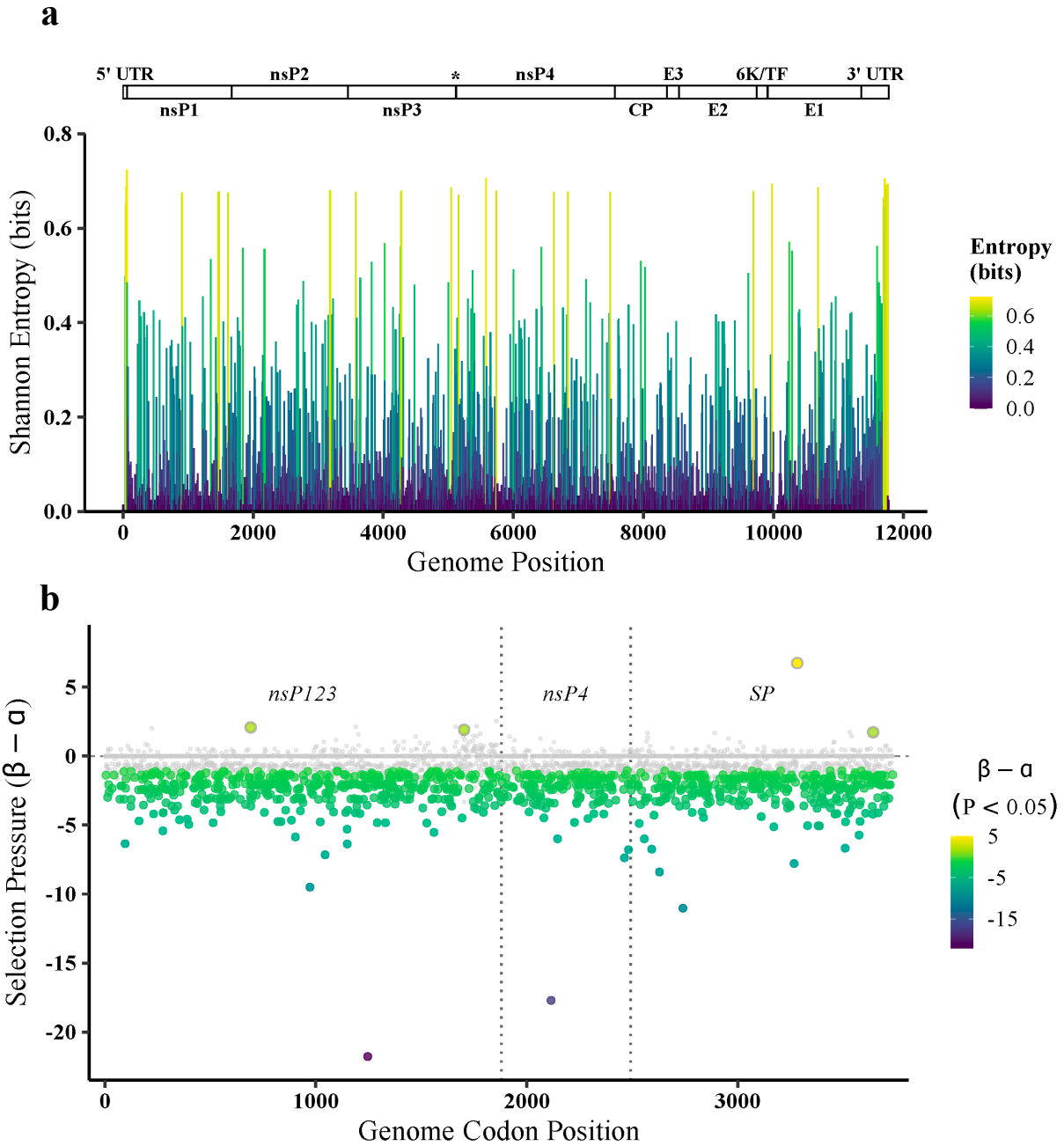


Figure 3.3. Genomic diversity and site-specific selection across North American EEEV genomes. (a) Site-specific Shannon entropy plot across a multiple-sequence alignment of EEEV genomes ($n = 584$) with a schematic of the EEEV genome structure positioned above the plot. The scale colour indicates positions of elevated sequence diversification in bits. (b) Fixed Effects Likelihood (FEL) analysis of site-specific non-synonymous (β) and synonymous (α) substitution rate differences ($\beta - \alpha$). Negative values ($\beta - \alpha < 0$) indicate sites experiencing purifying (negative) selection, while positive values ($\beta - \alpha > 0$) indicate sites under diversifying (positive) selection. Statistically significant sites ($P < 0.05$) are highlighted and separated by their respective polypeptide regions.

3.3 Phylogenetic Reconstruction and Temporal Signal Evaluation

Maximum likelihood (ML) phylogenetic reconstruction of the EEEV dataset was performed to assess baseline genetic relationships and branching architecture among the 584 EEEV genomes prior to time-scaled analysis (Figure 3.4a). The resulting ML tree topology features a structured branching pattern anchored at deep basal nodes by early historical sequences, including the earliest Canadian sequence (KJ469595.1; 1961), which branches alongside historical U.S. strains. The structural arrangement of the novel Canadian EEEV genomes indicates multiple introduction events rather than a single-source outbreak model, as the sequences did not form a single monophyletic group. Instead, Canadian sequences were polyphyletic and distributed across multiple distinct phylogenetic positions throughout the tree, comprising 5 isolated singletons alongside 5 multi-sequence clades. Relevant major ancestral nodes and key Canadian clusters received high support (UFBoot $\geq 95\%$), while weakly supported nodes (UFBoot $< 70\%$) were left unlabeled. To facilitate visual interpretation of Canadian lineages, large non-Canadian U.S. monophyletic clades were collapsed.

To evaluate the dataset's suitability for time-scaled phylogeographic analysis, root-to-tip regression analysis was performed using TempEst (Figure 3.4b). Across the 584 EEEV genomes sampled over a 93-year window, the analysis exhibited a strong relationship between sampling year and genetic divergence ($R^2 = 0.91$), demonstrating clock-like substitution dynamics. The regression yielded an estimated evolutionary rate (slope) of 1.69×10^{-4} substitutions/site/year and the x-intercept projected the time of the most recent common ancestor (TMRCA) to approximately 1924, immediately upstream of the earliest physical sampling baseline in the dataset (1933). Importantly, these values were interpreted as exploratory evidence of a sufficient temporal signal rather than as formal posterior estimates of the evolutionary rate or root age,

which were subsequently inferred using Bayesian molecular-clock modelling in BEAST.

Furthermore, the 41 novel Canadian genomes fell tightly along the regression line with minimal residual variations ($< 1.0 \times 10^{-3}$), indicating that the novel Canadian sequences did not appear as outliers relative to the broader North American EEEV population.

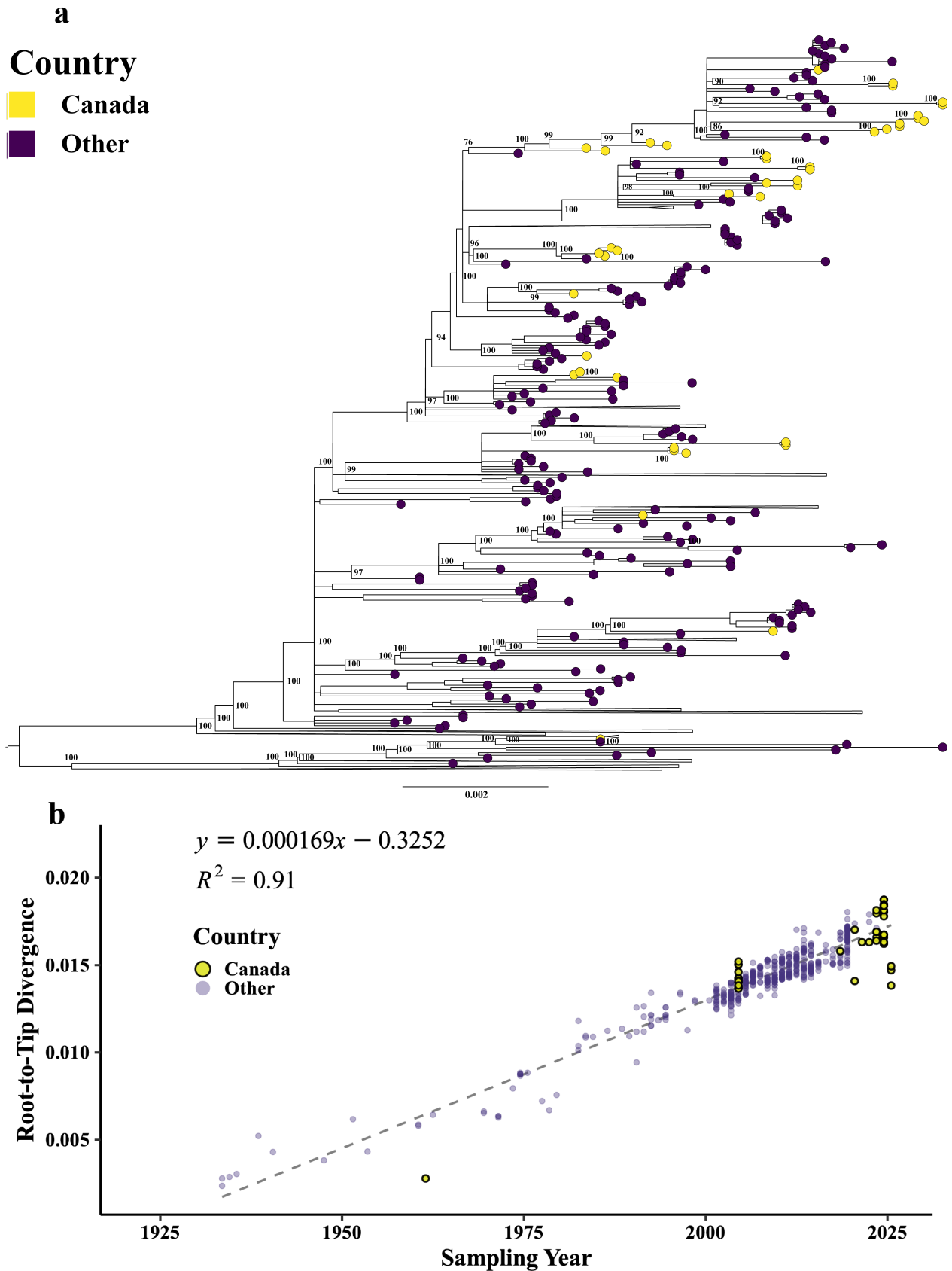


Figure 3.4. Phylogenetic relationships and temporal signal evaluation of North American

EEEV. (a) Maximum likelihood phylogenetic tree of North American EEEV. Continuous, non-Canadian lineages are collapsed into macro-level clades (triangles). Canadian genomes are highlighted with yellow node shapes, while non-Canadian (Other) sequences are represented in purple. Branch lengths are measured in nucleotide substitutions per site. Node support values (UFBoot) are displayed only for major ancestral splits and key lineage introductions. **(b)** Root-to-tip genetic distance regression plot assessing the temporal signal and molecular clock suitability within the North American EEEV dataset. Canadian genomes are highlighted with yellow node shapes, while non-Canadian (Other) sequences are represented in purple.

3.4 Time-Resolved Phylogenetic Analysis of EEEV Genome Sequences

A time-scaled phylogeographic reconstruction was performed using 581 genomes, including 543 publicly available genomes and 38 novel Canadian genomes with province-level metadata (Figure 3.5). Overall, the resulting topology was consistent with previously published phylogenies based on smaller datasets (Hill et al., 2023; Tan et al., 2018). Under the strict molecular-clock model, the median time to the most recent common ancestor (TMRCA) of the dataset was estimated as 1911 (95% Highest Posterior Density [HPD]: 1909–1919), and the posterior mean substitution rate was 1.602×10^{-4} substitutions/site/year (95% HPD: 1.514×10^{-4} – 1.692×10^{-4}), consistent with our root-to-tip regression slope. The inferred TMRCA predates the earliest sampled genomes in the dataset (1933; NCBI Accessions KJ469600.1 and KX000089.1) by approximately two decades; however, the exact location, host, and circulation during this interval remain unobserved. The 38 novel Canadian genomes did not form a single monophyletic group and instead were distributed among several clusters and individual phylogenetic positions.

To explore the spatial dynamics of EEEV across North America, discrete-trait phylogeographic reconstruction was used to infer ancestral geographic states across the time-scaled tree. The resulting ancestral state reconstructions indicated that the United States consistently served as the primary source of viral introductions into neighbouring regions,

consistent with prior phylogeographic analyses (Hill et al., 2023; Tan et al., 2018). Within this context, the earliest recorded Canadian sequence from 1961 forms a sister clade to a 1933 U.S. reference strain (NCBI Accession KJ469600.1), with their most recent common ancestor (MRCA) estimated at 1933 (95% HPD: 1932–1933). However, because node dates reflect evolutionary divergence rather than the precise timing of geographic entry, the introduction into Canada could have occurred at any point along the intervening descendant branch prior to 1961. Separate inferred ancestries are also evident among Canadian sequences from the 2000s. For example, two 2004 Quebec horse sequences (2004_EEEV_QC_Horse_6 and 2004_EEEV_QC_Horse_7), share an inferred MRCA dated to 2002 (95% HPD: 2001–2003). A separate cluster of 2004 Quebec genomes (2004_EEEV_QC_Horse_2, 2004_EEEV_QC_Horse_8, and 2004_EEEV_QC_Horse_9) shared an inferred MRCA dating back to approximately 2001 (95% HPD: 2000–2003) with a Florida strain (NCBI Accession KU840315.1). These distinct phylogenetic positions are consistent with more than one viral incursion into Canadian regions during this period.

Similarly, Canadian EEEV sequences collected during the 2020s were distributed across several distinct portions of the phylogeny. A single 2023 Ontario equine sequence, 2023_EEEV_ON_Horse_2, was placed within a lineage containing viruses from New York, Connecticut, and Massachusetts and shared an inferred ancestral node dated to 2009 (95% HPD: 2009–2012). A separate Maritime–Ontario monophyletic cluster included 2025 mosquito-pool sequences from Nova Scotia (2025_EEEV_NS_Mosquito_1, 2025_EEEV_NS_Mosquito_2) and Prince Edward Island (2025_EEEV_PEI_Mosquito_1), together with a 2020 Ontario equine sequence (2020_EEEV_ON_Horse_2). This clade was most closely related to U.S. sequences collected between 2010 and 2019 and descended from an inferred ancestral node dated to 2003

(95% HPD: 2001–2005). A larger Canadian cluster contained 12 sequences collected between 2020 and 2024 from avian, equine, and mosquito hosts in Ontario, Quebec, and New Brunswick. This group forms a sister clade to Massachusetts sequences collected in 2018, with an inferred MRCA estimated at 2010 (95% HPD: 2009–2011). Additional Canadian sequences formed smaller, geographically localized clusters near the tips of the tree, several of which were associated with U.S. lineages rather than with deeper Canadian ancestral branches. Collectively, the occurrence of Canadian sequences in multiple phylogenetic positions is consistent with repeated introductions followed by localized transmission. However, the clustering of some Canadian sequences across years also leaves open the possibility of short-term regional persistence or overwintering, particularly given the long intervening branches and limited sampling.



Figure 3.5. Discrete time-calibrated phylogeographic analysis of North American EEEV dispersal. A Bayesian MCC phylogenetic tree with internal branches coloured according to ancestral discrete geographic locations inferred using an asymmetric trait substitution model with BSSVS in BEAST v1.10.4. Node heights represent median age estimates. Tip nodes are colored according to the sampling location. The time scale is shown in years.

3.5 Inferred EEEV Spread in North America

Based on the discrete trait phylogeographic reconstruction, the United States was inferred as the most probable origin for the North American EEEV dataset (Figures 3.5 and 3.6a).

However, these spatial reconstructions reflect the most probable ancestral states inferred from the available data and are sensitive to sampling composition and density. Following the emergence of the virus, reconstructed ancestral states remained concentrated along the eastern U.S. coast for several decades. The model placed Florida at the base of several coastal lineages, with subsequent transitions extending northward along the Atlantic coast to states such as New Jersey and Maryland by approximately 1928. These inferred transitions predate formally reported cases in 1933. By 1960, the reconstructed geographic range included additional U.S. coastal regions such as Massachusetts, Delaware, and New York, as well as southern states such as Georgia and Louisiana (Figure 3.6a).

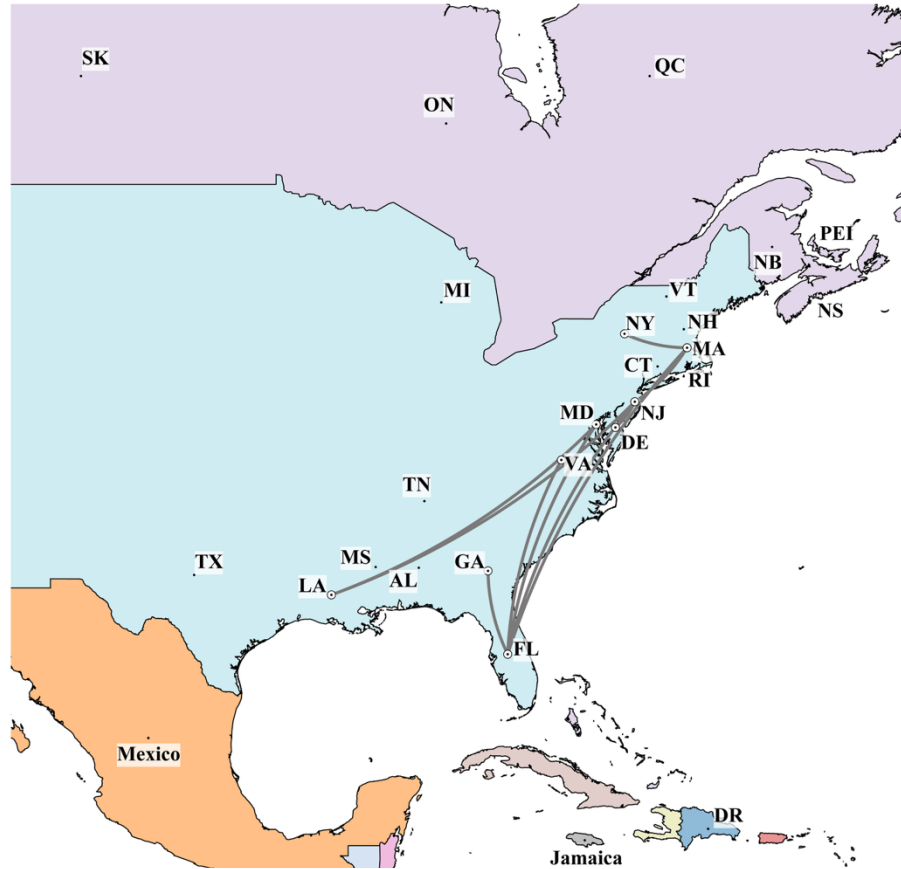
The period from 1961 to 2004 records the earliest reconstructed transitions linked to Canadian detections, including an inferred transition associated with an Ontario detection in 1961, in which New Jersey was reconstructed as the most probable ancestral state. Over the next four decades, reconstructed spatial paths spanned additional regions across North America and the Caribbean, including Jamaica, the Dominican Republic, and several additional U.S. states (Michigan, Mississippi, Rhode Island, Connecticut, Texas, and Tennessee) (Figure 3.6b). This interval concludes with additional Canadian detections in Quebec in 2004, which traced back to inferred node origins in Florida.

The period from 2005 to 2025 is characterized by expanded EEEV surveillance, encompassing more locations and a broader network of lineages (Figure 3.6c). During this time, reconstructed spatial states expanded to include Vermont, New Hampshire, and Alabama, alongside inferred transitions into Mexico. Because increased surveillance effort can mimic geographic expansion, these additional locations reflect a broader sampled range rather than definitive evidence of recent viral spread. Reconstructed transitions associated with 2024 Ontario detections were linked to inferred ancestral states in New York and Massachusetts.

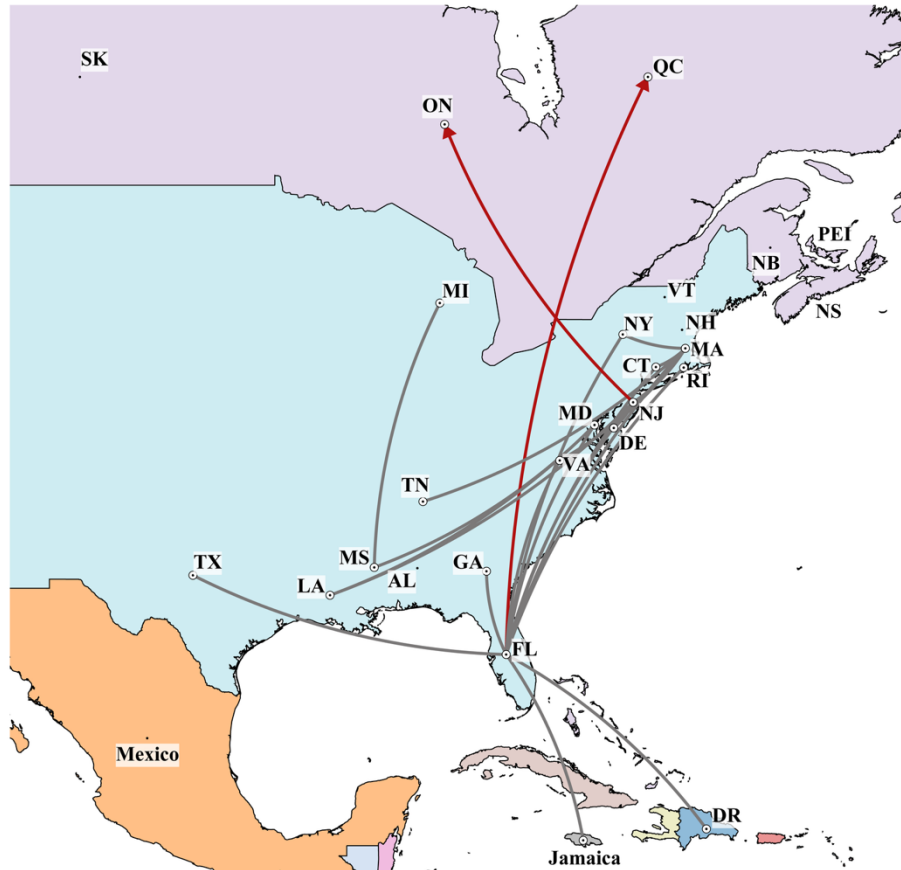
Additionally, during this period, the model reconstructed transitions between sampled Canadian geographic locations. An initial inferred transition between Quebec and Ontario was resolved in 2011, preceding detections in 2018 and 2020. Subsequent transitions between Quebec and Ontario were inferred in 2015, 2020, and 2024. Eastward transitions were also reconstructed from Quebec to Nova Scotia (2020, 2025) and New Brunswick (2024), as well as from Nova Scotia to Prince Edward Island (2025). Finally, a reconstructed transition was inferred from Ontario to Saskatchewan in 2024.

Because many Canadian sequences exhibited high genetic similarity, the reconstructed pathways could not distinguish direct dissemination among Canadian locations from independent introductions of closely related viruses from unsampled source populations. The inferred Ontario-to-Saskatchewan transition should therefore be interpreted as a discrete trait reconstruction rather than evidence of direct viral or host movement between the two provinces.

a



b



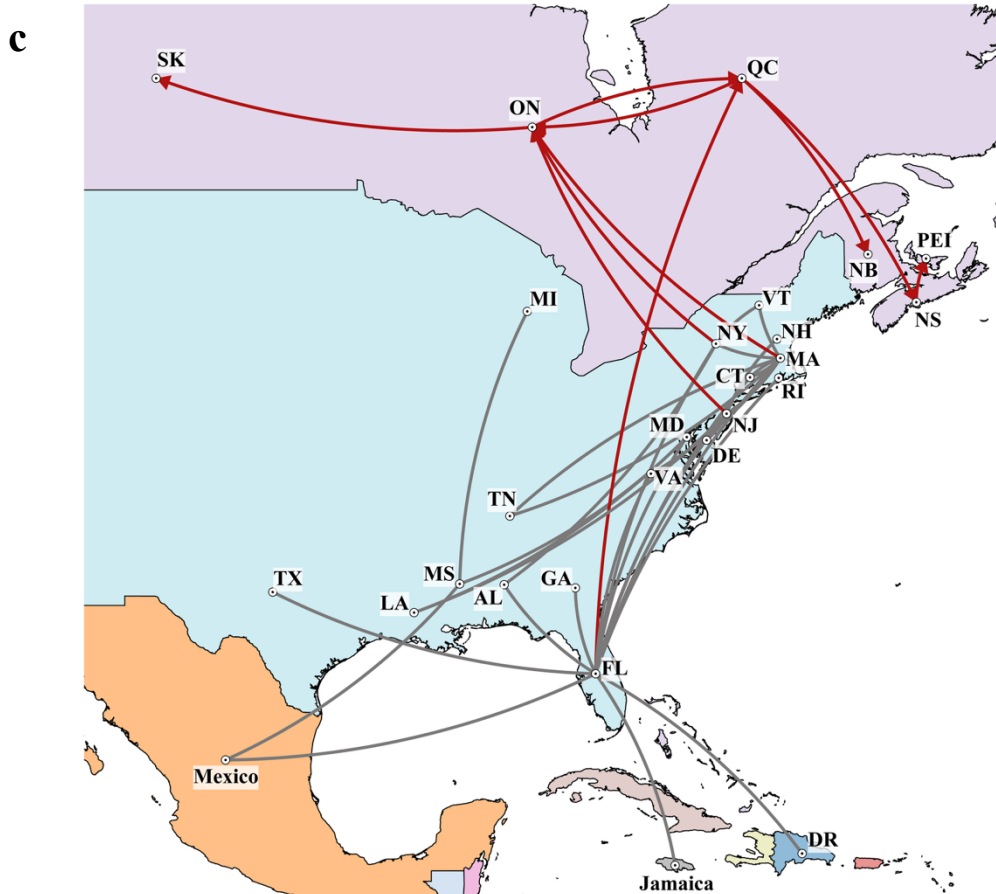


Figure 3.6. Spatiotemporal reconstruction of EEEV dispersal in North America. Dispersal pathways are coloured in red for movement within Canada and across the Canada-U.S. border. (a) Inferred viral emergence and spread within the U.S., 1911–1960. (b) Inferred transitions into Canada, Jamaica, and the Dominican Republic alongside reconstructed regional circulation within the U.S., 1961–2004. (c) Reconstructed viral dispersal characterized by an increased number of sampled locations and transitions, reflecting a highly diverse network of viral genomes across eastern North America, 2005–2025.

3.6 Phylogeographic and Host-State Transition Dynamics

To better characterize patterns of viral movement for EEEV, discrete-trait ancestral reconstruction using BSSVS identified 22 strongly supported geographic transition pathways with posterior inclusion probability (PP) ≥ 0.70 and Bayes factor (BF) ≥ 10 (Figure 3.7a; Supplementary Table S1.1). Florida and Massachusetts were prominent nodes in the inferred diffusion network. Transitions from Florida to Massachusetts, New York, and Quebec received maximal posterior support (PP = 1.00), while additional strongly supported pathways included

Florida to Georgia and Connecticut. Massachusetts was also associated with several well-supported outgoing transitions, including movement to Virginia, Tennessee, New Jersey, and Connecticut. Bidirectional exchange was inferred between Massachusetts and Tennessee and between Massachusetts and Connecticut. Within Canada, strongly supported regional movement was identified from Quebec to Ontario and from Nova Scotia to Prince Edward Island. Additional supported pathways included transitions from Mississippi to Michigan and to Mexico. Because sampling intensity varied substantially among locations, the prominence and direction of some inferred transitions may partly reflect uneven sequence representation.

Discrete host-state reconstruction across eight host categories identified strongly supported directional transitions among mosquitoes, birds, horses, humans, deer, alpacas, dogs, and wolves (Figure 3.7b; Supplementary Table S1.2). The most strongly supported pathways originated from mosquitoes and included transitions to birds, horses, and humans, each with a PP = 1.00. Additional supported transitions included horse to deer, bird to alpaca, horse to mosquito, and bird to horse. Because these reconstructed parameters reflect discrete character-state transitions along tree branches rather than observed physical transmission events, small sample sizes in certain host categories can bias inferred ancestral states. Only transitions with PP $\geq$ 0.70 and BF $\geq$ 10 were displayed in Figure 3.7 to emphasize the most consistently supported geographic and host-transition pathways. Complete transition statistics are provided in Supplementary Tables S1.1 and S1.2.

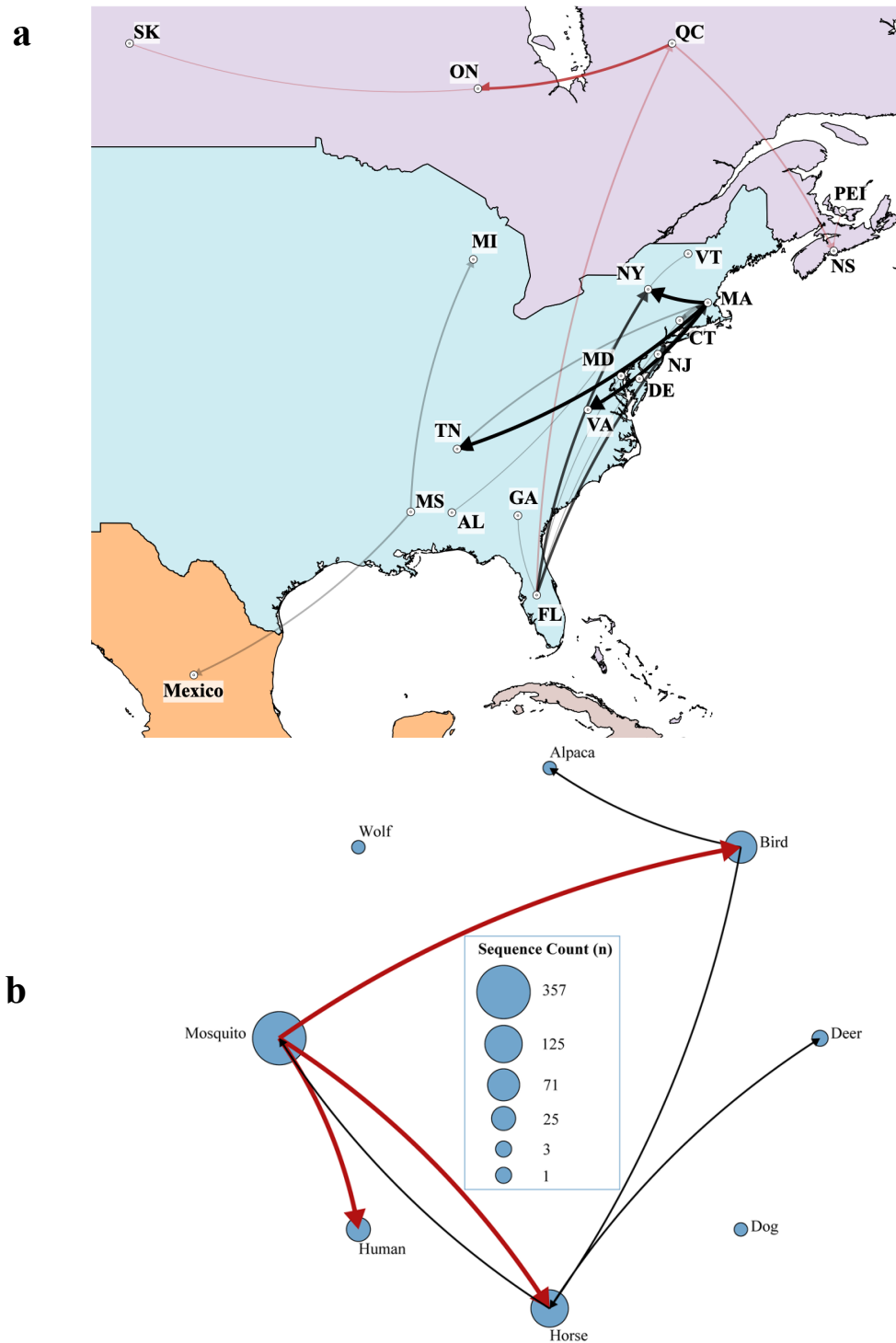


Figure 3.7. Spatiotemporal diffusion and host-state transition dynamics of EEEV. (a) Inferred geographic diffusion pathways reconstructed using Bayesian Stochastic Search Variable Selection (BSSVS). Only transitions with posterior inclusion probability (PP) ≥ 0.70 and Bayes factor (BF) ≥ 10 are shown. Line width is proportional to the posterior mean transition-rate

parameter, with solid black lines representing higher relative transition rates (> 1.5) and transparent lines indicating lower relative transition rates (< 1.5). Canadian pathways are highlighted in red. **(b)** Directional host-state transition network reconstructed using BSSVS, showing only transitions with $PP \geq 0.70$ and $BF \geq 10$. Node size is proportional to the number of genomes within each host category. Ecologically expected mosquito-associated transitions are highlighted in red; other statistically supported host-state transitions are shown in black. Arrows represent reconstructed changes in sampled host state and should not be interpreted as direct transmission events.

3.7 Geographic Distribution of North American EEEV Haplotypes

To investigate haplotype relationships among North American EEEV, 581 genomes were analyzed using a Templeton, Crandall, and Sing (TCS) network based on six geographic location groupings: Northeastern U.S., Southeastern U.S., Gulf Coast U.S., Midwest/Interior U.S., Canada, and Mexico and Caribbean (Figure 3.8; Supplementary Table S1.3). The analysis identified 223 haplotypes, of which 143 (64.1%) were represented by a single sequence and 80 (35.9%) were shared by two or more sequences, with the largest comprising 56 genomes. The alignment contained 243 segregating sites, including 160 parsimony-informative sites, and exhibited low overall nucleotide diversity ($\pi = 0.00747$; corresponding to a mean pairwise sequence identity of 99.25%). Neighbouring haplotypes were generally separated by few mutational steps, reflecting high sequence similarity. The network showed limited segregation by broad geographic region; the largest, high-frequency haplotypes were geographically mixed and predominantly composed of co-circulating Northeastern U.S. and Southeastern U.S. sequences. Lineages from the Gulf Coast U.S., Midwest/Interior U.S., and Mexico and Caribbean were less frequent, appearing as peripheral variants or within mixed clusters.

The 39 Canadian genomes were distributed across 27 haplotypes spanning multiple areas of the network, rather than forming a single geographically exclusive cluster. Of these, 36 Canadian genomes occurred in haplotypes composed exclusively of Canadian sequences ($n=24$; designated here as "Canada-only" haplotypes), while three occurred in haplotypes shared with

U.S. sequences. The three shared haplotypes contained sequences from Ontario and Nova Scotia, together with sequences from several U.S. states, including Massachusetts, New York, New Jersey, Delaware, Florida, and Virginia. Thus, Canadian genomes showed previously uncharacterized haplotype diversity while also sharing identical haplotypes with the broader U.S. dataset.

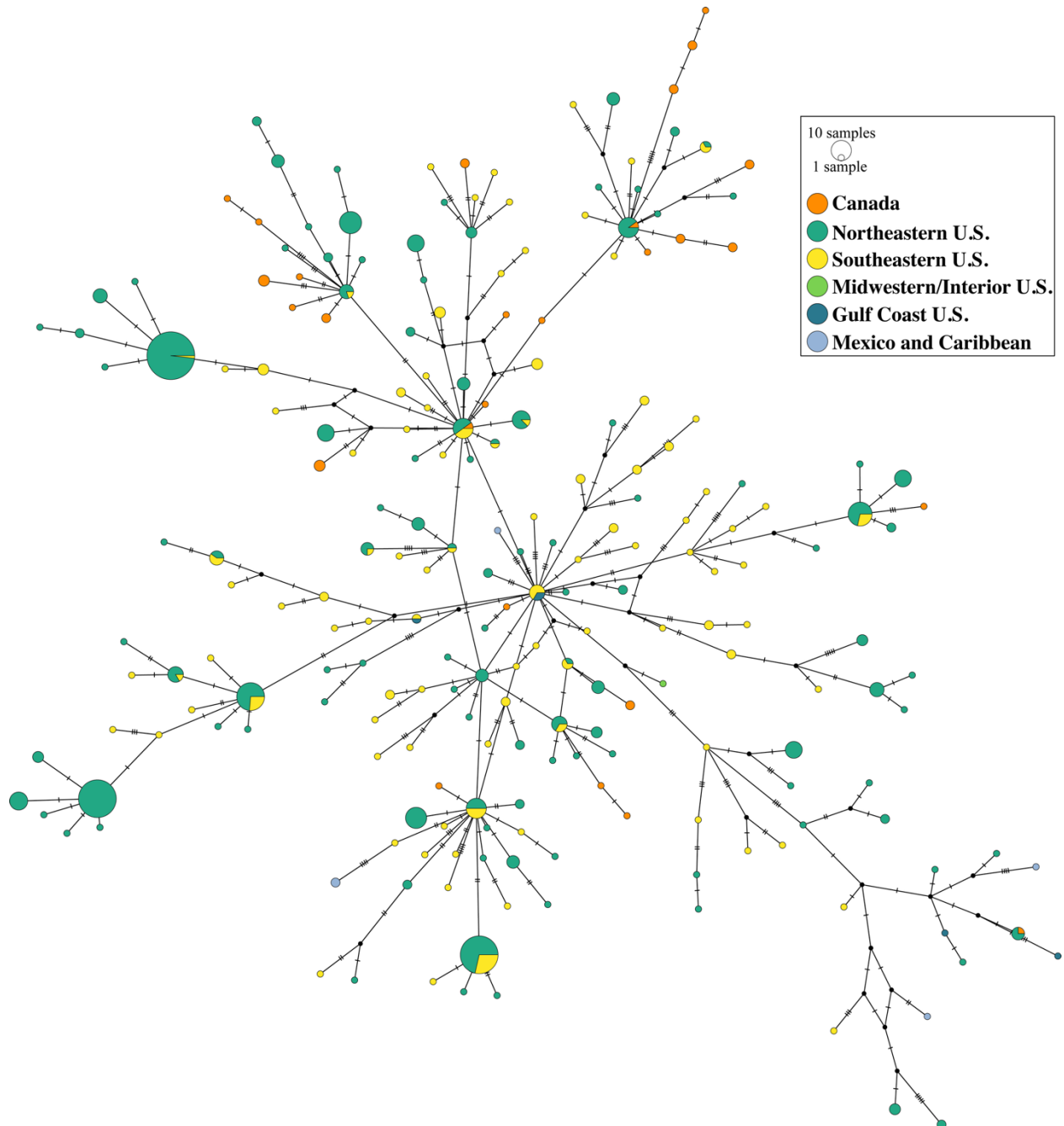


Figure 3.8. Haplotype network of North American EEEV. Network reconstruction was performed using the TCS method from a multiple-sequence alignment of 581 genomes. Each circle represents a unique haplotype, with circle size proportional to sequence frequency and colours corresponding to the six geographic regions of origin. Connecting lines illustrate mutational pathways, with transverse hatch marks representing individual nucleotide substitution steps. Geographically mixed pie charts denote identical viral genotypes detected across multiple regions. Small black nodes represent inferred, unsampled ancestral intermediates.

CHAPTER 4: DISCUSSION

4.1 Targeted Genome Sequencing Expands Canadian EEEV

Eastern equine encephalitis virus (EEEV) is an important mosquito-borne, neuroinvasive alphavirus with recurring detections in Canada. This research substantially expanded the genomic characterization of EEEV in Canada and provided the first broad assessment of Canadian EEEV diversity within the wider North American landscape. This study generated 41 novel Canadian EEEV genomes from equine ($n = 28$), avian ($n = 7$), mosquito ($n = 4$), white-tailed deer ($n = 1$), and wolf ($n = 1$) surveillance samples spanning multiple geographic locations, collection years, and viral loads using a novel targeted tiling-amplicon sequencing approach (Table 3.1). The recovery of complete genomes demonstrated that targeted sequencing can expand genomic surveillance of EEEV across diverse sample types, although sensitivity across samples and cycle-threshold values was not formally evaluated. The high genomic conservation among EEEV genomes likely contributed to the success of this approach and its suitability for characterizing consensus-level genomic variation (Figure 3.2; Supplementary Figure S1.1). In public health and arboviral surveillance settings, tiling-amplicon schemes offer practical operational advantages due to their lower cost per sample, scalability, and multiplexing capabilities (Grubaugh et al., 2019; Quick et al., 2017). Future integration would allow surveillance and diagnostic detections to be converted into epidemiologically informative data. Ultimately, the workflow converted surveillance detections into complete sequence data that could be used to investigate EEEV diversity, evolutionary relationships, and geographic movement in Canada.

Amplicon sequencing approaches can have limitations, however. They remain prone to dropout or uneven sequencing coverage depth, especially in samples containing low viral

template, degraded genetic material, or mutations at primer-binding sites (Quick et al., 2017; Freed et al., 2020). Furthermore, while the current EEEV landscape exhibits high genomic conservation, periodic evaluation and primer optimization for the tiling-amplicon scheme may be required as novel diversity is characterized. Although the tiling-amplicon workflow successfully recovered the complete coding sequence of EEEV, amplicon-based methods typically experience coverage dropouts at the extreme 5' and 3'UTRs, requiring specialized techniques such as 5'/3' RACE to capture terminal non-coding sequences, and thus the complete genome. Multiple amplification steps can introduce PCR artifacts, although the use of a high-fidelity polymerase with proofreading abilities and reducing the number of PCR cycles in high-viral-load samples would serve as suitable methods to reduce this (Orton et al., 2015). Additionally, primer sequences must be trimmed prior to consensus sequence generation to avoid artificial masking of true variation at primer binding sites (Grubaugh et al., 2019). Finally, because this study relied on reference-based consensus generation, the workflow was designed to characterize consensus-level genomic variation across samples rather than minority variants or intrahost diversity. Consensus sequences represent the dominant nucleotide at each genomic position and therefore capture population-level variation between samples but not low-frequency intrahost variants. Characterizing intrahost diversity would require deeper sequencing and variant-frequency analyses, which were beyond the scope of this study.

4.2 Evolutionary Constraint Shapes EEEV Genomic Diversity

North American EEEV genomes remained highly conserved despite the large temporal span and sample size. The average pairwise nucleotide identity was approximately 99.1%, with most comparisons exceeding 99% identity (Figure 3.2). Site-specific entropy was similarly low across the genome, with no regions showing consistently elevated variability (Figure 3.3a).

These findings indicate that EEEV diversity is distributed across relatively few nucleotide positions rather than concentrated within highly variable genomic regions. This conservation is consistent with the complex transmission cycle of EEEV, which requires viral replication in biologically distinct avian and vector hosts (Armstrong & Andreadis, 2022). Alternating between hosts may restrict evolutionary change because mutations that increase fitness in one host may be deleterious in another. Widespread evidence of purifying selection identified by Fixed Effects Likelihood (FEL) analysis ($P < 0.05$, $\beta < \alpha$; 895 of 3,735 codon sites or 23.96% of total) further demonstrated strong evolutionary constraint across the EEEV genome (Figure 3.3b). Particularly strong purifying selection was observed at non-structural sites nsP123 codon 1246 ($\beta - \alpha = -21.776$, $P < 0.0001$) and nsP4 codon 236 ($\beta - \alpha = -17.585$, $P = 0.0001$), highlighting critical functional constraints within the viral RNA-dependent RNA polymerase and replication machinery (Zhang et al., 2026). This widespread constraint limits overall nucleotide diversity across the dataset, causing the observed haplotype variation to be driven primarily by low-frequency single-nucleotide differences rather than by deeply divergent lineages.

Despite strong purifying selection, four codon positions showed significant evidence of diversifying selection ($P < 0.05$, $\beta > \alpha$). Within nsP123, codon 691 located in the mature nsP2 protein, was characterized by an aspartic acid-to-alanine substitution ($D \rightarrow A$, $n = 4$; $\beta - \alpha = +2.078$, $P = 0.0422$). Alphavirus nsP2 plays central roles in viral replication, host polyprotein processing, and suppression of host antiviral responses (Adeyinka et al., 2025; Kim et al., 2013; Parashar et al., 2025). However, this substitution occurred at low frequencies, and its specific molecular or functional effect on nsP2 performance remains unknown.

A second diversifying site occurred at nsP123 codon 1703 within the hypervariable domain of nsP3 ($n = 6$, $\beta - \alpha = +1.907$, $P = 0.0402$), showing substitutions to multiple amino

acids (D → E/Y/G/N). This region functions as a hub involved in the assembly of the viral replication complex and host protein interactions (Kril et al., 2024; Meshram et al., 2018). While variable residues at this locus indicate structural tolerance, no consistent pattern was observed across host species, collection locations, or sampling years, and the functional consequences of these substitutions are uncharacterized.

Two additional sites under diversifying selection were identified in the structural polyprotein. Codon 789 in the 6K viroporin protein showed an alanine-to-valine substitution (A → V, $n = 21$; $\beta - \alpha = +6.743$, $P < 0.0001$), consistent with previous EEEV analyses (Tan et al., 2018). The 6K protein is involved in membrane permeability, glycoprotein trafficking, and viral budding (Negi et al., 2025). Additionally, structural codon 1149 in the E1 envelope glycoprotein exhibited variation (A → T/S/V, $n = 5$; $\beta - \alpha = +1.728$, $P = 0.0294$), a protein that mediates virus-cell attachment and membrane fusion during entry (Kielian et al., 2010). As with the non-structural sites, the impact of variation at these structural positions has not been experimentally determined.

These four positions represent reasonable candidate sites for future investigation, but the FEL results identify candidate variable positions rather than demonstrating functional adaptation. The underlying substitutions were uncommon; two sites exhibited p-values near the 0.05 significance threshold across thousands of codon positions tested, and no experimental validation was performed. Without formal testing correction across the entire genome, the possibility of false-positive site identification must be acknowledged. Rather than relying on sequence counts, FEL estimates substitution rates relative to phylogenetic topology, meaning a small number of nonsynonymous changes on terminal branches can yield statistical evidence of diversifying selection (Kosakovsky Pond & Frost, 2005). In highly conserved viral genomes such as EEEV,

these terminal substitutions can represent transient or mildly deleterious variants that fail to become established in the broader population (Tan et al., 2018). Consequently, candidate genomic sites for future study are those supported by independent replication in prior work, such as 6K codon 789, which may represent genomic regions where variation could influence viral fitness or host–virus interactions.

4.3 Canadian EEEV Diversity Is Integrated Within the Broader North American Population

A TCS haplotype network reconstructed from 581 North American EEEV genomes was used to investigate whether Canadian sequences formed a single geographically distinct group or were distributed throughout the North American population (Figure 3.8). The 39 Canadian sequences were distributed among 27 haplotypes, of which 36 genomes occurred in Canada-only haplotypes ($n = 24$), while the remaining three occurred in haplotypes shared with sequences from the northeastern or southeastern United States. Although the expanded Canadian dataset revealed previously unsampled haplotype diversity, Canadian genomes did not form a single geographically distinct group and instead occurred across the broader North American network, including several haplotypes shared with U.S. sequences.

Although 24 haplotypes were observed only among Canadian genomes, this should not be interpreted as evidence of an endemic presence or evolutionary restriction to Canada. Many Canadian haplotypes were represented by one or a small number of genomes, and many were separated by only a few mutational steps from haplotypes dominated by northeastern U.S. genomes. While the northeastern United States was well represented numerically, sampling remained uneven across states and collection years. Uneven sampling from neighbouring U.S. regions and from the Canadian enzootic transmission cycle may therefore cause widely

circulating haplotypes to appear geographically restricted, even if they circulate more broadly than the current dataset indicates (Frost et al., 2015). Therefore, the designation of a Canada-only haplotype reflects its occurrence in the current dataset rather than its long-term persistence or local establishment.

Taken together, the haplotype network and phylogenetic analyses provide little support for a single, independently evolving Canadian population. Canadian genomes were genetically connected to the broader North American population, while the Canada-only haplotype labels describe their distribution in the current dataset rather than demonstrating geographic restriction or local establishment.

4.4 Repeated Incursions Are Predominant but Short-Term Persistence Possible

Given the relatively small number of Canadian genomes, high pairwise sequence identity, and absence of population differentiation, repeated independent introductions provide a stronger explanation for Canadian EEEV detections than the maintenance of a single endemic lineage. The placement of Canadian EEEV genomes across multiple branches of both the maximum-likelihood and time-scaled phylogenies indicates that the Canadian detections cannot be explained by the endemic persistence and evolution of a single lineage (Figure 3.4a, Figure 3.5). Instead, the tree topology supports repeated incursions of EEEV into Canada from the broader North American population. Canadian genomes collected during similar periods often shared different common ancestors and were distributed separately across the tree, indicating that multiple viral lineages circulate or enter Canada concurrently. Although small clusters of Canadian sequences were present in the maximum-likelihood tree, their occurrence does not demonstrate sustained local persistence, as repeated introductions from a highly conserved source population can produce patterns resembling continued local circulation.

The analysis of 38 novel Canadian genomes with sufficient metadata as part of the broader 581-genome phylogeographic dataset enabled a comprehensive study of EEEV evolution across North America (Figure 3.5). The dataset exhibited a strong temporal signal, supporting the use of molecular clock analysis and yielding rate estimates consistent with existing studies (Figure 3.4b) (Hill et al., 2023). The time-scaled maximum clade credibility (MCC) tree supported multiple independent incursions of EEEV into Canada and revealed several Canadian clusters, suggesting that some EEEV introductions may have been followed by localized transmission. Closely related genomes collected across multiple provinces, hosts, and collection years formed a Maritime–Ontario cluster and a larger group of 12 sequences from Ontario, Quebec, and New Brunswick. The detection of closely related genomes across provinces, hosts, and successive transmission seasons is compatible with localized circulation or short-term persistence, but repeated introductions from an unsampled, genetically similar source population could produce the same pattern (Tan et al., 2018).

The Saskatchewan genome represented an isolated detection and did not provide evidence of sustained western circulation. Its phylogenetic placement was consistent with a separate incursion, but the absence of regional sampling and detailed case-history information prevented definitive identification of the exposure location or source population. The inferred Ontario-to-Saskatchewan transition should therefore not be interpreted as direct viral movement between the two provinces. Nevertheless, given EEEV’s multi-day incubation period, exposure during domestic animal transport or travel from an EEEV-positive zone such as Ontario remains a plausible alternative explanation alongside unobserved vector movement (World Organisation for Animal Health, 2021).

When evaluating EEEV movement into Canada, sampling heterogeneity across the dataset must be considered. The majority of Canadian genomes were collected after 2020, leaving major temporal gaps between the historical 1961 sample and our earliest modern sample in 2004. Furthermore, U.S. sequences account for 92% of the dataset, with over half originating from the northeastern U.S. region (Figure 3.1). Uneven sampling across states, years, hosts, and transmission cycles, with Connecticut, New York, and Massachusetts representing the most heavily sampled jurisdictions (Figure 3.1a), means that closely related viruses circulating in unsampled locations or years could repeatedly enter Canada and produce phylogenetic patterns resembling local persistence. Crucially, sequence representation in the dataset does not necessarily reflect ecological importance in viral maintenance or spread; heavily sampled states may simply reflect intensive regional surveillance rather than true primary-source hubs (De Maio et al., 2015). Additionally, discrete phylogeographic models are sensitive to sampling composition; overrepresentation of particular U.S. states can make models overconfident in inferring heavily sampled locations as source nodes while obscuring transmission pathways in locations with limited sampling, such as Canada (Hill et al., 2023; Layan et al., 2023).

Consequently, phylogenetic topology alone is insufficient to conclusively establish sustained local persistence. In several Canadian clusters, detections were separated by long unsampled intervals in the time-scaled tree and substantial temporal gaps, in some cases almost 15 years. For example, the Maritime-Ontario cluster contained samples from 2020-2025, yet its inferred ancestral node was dated to 2003 (95% HPD: 2001-2005). This unsampled interval could represent cryptic circulation within Canada, repeated introductions from an unsampled external source population, or a combination of both. High genomic conservation further complicates this distinction, as separate transmission events may remain nearly identical over

epidemiologically meaningful periods. Together, these findings indicate that independent incursions are the primary driver of Canadian EEEV diversity, though temporally and geographically clustered lineages suggest that certain introductions lead to localized transmission and potential persistence across seasons. Denser seasonal sampling of primary enzootic vectors (e.g., *Cs. melanura*), opportunistic bridge species (e.g., *Cq. perturbans*, *Ae. canadensis*), and migratory or resident passerine reservoirs (e.g., American Robins, Wood Thrushes) will be required to resolve the frequency of reintroduction versus sustained local maintenance.

4.5 North American Dispersal and Northeastern U.S. Circulation

The movement of EEEV across North America was reconstructed using discrete phylogeographic analysis. Several strongly supported transitions (posterior inclusion probability (PP) ≥ 0.70 and Bayes factor (BF) ≥ 10) linked Florida, the eastern United States, and eastern Canada (Figures 3.6 and 3.7a). Florida and Massachusetts were reconstructed as important ancestral or intermediate locations, whereas New York, Quebec, Ontario, Nova Scotia, and Prince Edward Island were part of a broader regional dispersal network. These findings indicate that Canadian EEEV detections occur within a larger circulating system along the Atlantic coast, rather than representing geographically isolated transmission.

Spatiotemporal reconstruction highlighted Florida as an important reconstructed geographic state, associated with several strongly supported transitions towards Massachusetts, New York, and Quebec. This pattern aligned with previous studies implicating Florida as an important enzootic focus and source region for EEEV, supported by its favourable climate, abundant vector habitats, and role as a wintering region for migratory birds (Bingham et al., 2014; Burkett-Cadena et al., 2022; J. Downs et al., 2024; Tan et al., 2018). However, this reconstruction should not be taken as definitive evidence that Florida was the exclusive

geographic origin of North American EEEV lineages. As one of the most heavily sampled regions in the dataset, ancestral state reconstruction can be sensitive to sequence counts, temporal distribution, and tree topology (Wright et al., 2015).

Similar to Florida, the northeastern United States appeared to serve as an important regional link between southern EEEV circulation and Canadian detections. Massachusetts was associated with multiple strongly supported transitions involving New York, Connecticut, New Jersey, Virginia, and Tennessee, while Canadian genomes frequently shared recent common ancestors with sequences from Massachusetts, New York, and neighbouring states. The continuity of coastal wetlands and forested habitats across the northeastern United States and eastern Canada likely supports similar mosquito-avian transmission cycles (Armstrong & Andreadis, 2022; Molaei et al., 2016). Dispersal of migratory avian hosts along Atlantic flyways offers a biologically plausible hypothesis for long-distance viral dissemination (Young et al., 2008). However, discrete phylogeographic analysis models geographic location alone; it cannot identify the specific host species, vector dynamics, or transport mechanisms underlying any reconstructed state transition.

Reconstructed viral movements also indicated transitions from Quebec to Ontario and Nova Scotia, from Nova Scotia to Prince Edward Island, and from Ontario to Saskatchewan. However, these supported pathways represent ancestral-state changes inferred along phylogenetic branches rather than routes of direct dispersal. Inferred state transitions frequently span unsampled intermediate geographic locations or unobserved transmission cycles, particularly when sampling density is limited, meaning a single branch transition does not imply direct physical movement between the two locations. This distinction is particularly crucial for long-distance inferences such as the Ontario-to-Saskatchewan transition, which most likely

reflects unsampled intermediate steps or hosts rather than a direct interprovincial transmission event.

The impact of sampling heterogeneity is particularly pronounced for heavily represented jurisdictions such as New York, Florida, and Massachusetts, which contributed 211, 112, and 78 genomes, respectively ($n = 401$, 69% of dataset). Because these locations dominate the dataset, discrete phylogeographic models are more likely to reconstruct them as primary ancestral nodes or intermediate hubs, whereas viral movement through sparsely sampled adjacent regions remains obscured (De Maio et al., 2015). Consequently, while the analysis supports a pattern of broad regional EEEV connectivity along the eastern North American corridor, the inferred importance of specific states reflects a combination of true ecological connectivity and surveillance intensity, leaving the precise source locations, pathways, and mechanisms for individual incursions uncertain.

4.6 Host-State Reconstruction Reflects Ecological Associations, Not Direct Transmission Chains

Discrete phylogeographic reconstruction of host traits identified strongly supported host-state transitions (posterior inclusion probability (PP) ≥ 0.70 and Bayes factor (BF) ≥ 10) involving mosquitoes, birds, horses, humans, and several wildlife hosts (Figures 1.1 and 3.7b). Supporting transitions from mosquitoes to avian, equine, and human hosts are broadly consistent with the established ecology of EEEV, reflecting the central role of mosquitoes in maintaining enzootic cycles and causing spillover into incidental hosts (Zubair et al., 2024). Other supported transitions included horse-to-deer, bird-to-alpaca, horse-to-mosquito, and bird-to-horse. While the patterns align with known vector-host dynamics, these transitions are biologically improbable and demonstrate that the model does not independently validate direct transmission, highlighting

the limitations of inferring host-to-host pathways from an unevenly sampled dataset (De Maio et al., 2015; Lemey et al., 2009).

Reconstructed host-state transitions describe statistical changes in the sampled host trait along phylogenetic branches and must not be interpreted as direct, host-to-host transmission events (Frost et al., 2015). For example, the inferred horse-to-deer or bird-to-alpaca transitions do not demonstrate direct transmission between these hosts. Instead, these detections suggest that sampled animals independently acquired closely related viral lineages from local mosquito vectors or an avian reservoir population. The reconstructed horse-to-mosquito transition should not be interpreted as evidence that a mosquito acquired EEEV from a horse, because horses generally do not develop sufficient viremia for subsequent mosquito transmission (American Association of Equine Practitioners, 2023; Weaver & Barrett, 2004). Instead, the transition more likely reflects uncertainty in ancestral-state reconstruction, uneven host sampling, or the absence of intermediate mosquito and avian genomes.

Interpreting the host-state analysis is further constrained by uneven sampling and the high genomic conservation of EEEV genomes. Mosquitoes, horses, and birds dominated the dataset, whereas several wildlife host categories were represented by only a handful of genomes. These sampling imbalances increase the likelihood that closely related sequences from incidental hosts are placed adjacent to one another on the phylogenetic tree, causing the model to infer direct host-state changes even when the true intermediate transmission steps were unobserved (De Maio et al., 2015; Frost et al., 2015). Furthermore, high sequence identity and short phylogenetic branches limit the genetic variation available to resolve the precise order and direction of host-state transitions (Frost et al., 2015; Lemey et al., 2009).

Despite these limitations, host-state reconstruction provides a framework for evaluating broad patterns of host association across the evolutionary history of EEEV, which would not be possible from sequence data alone without incorporating ecological traits. Consequently, the reconstructed host-state transitions denote statistical trait associations among sampled sequences rather than observed transmission chains. Overall, while the host-state analysis is consistent with repeated spillover from enzootic mosquito-avian cycles into humans and animals, the direction of individual transitions and the specific intermediate hosts involved remain unresolved, meaning that discrete trait modelling is best used to evaluate broad patterns of host association rather than determine the exact pathways driving individual spillover events.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

This research substantially expanded the genomic representation of EEEV in Canada and provided the first broad assessment of Canadian EEEV diversity within a continental evolutionary framework. Forty-one novel Canadian genomes were generated from equine ($n = 28$), avian ($n = 7$), mosquito ($n = 4$), white-tailed deer ($n = 1$), and wolf ($n = 1$) surveillance samples spanning multiple geographic locations, collection years, and viral loads. Before this research, only one historical Canadian genome was publicly available, severely limiting assessment of Canadian viral diversity and evolutionary relationships. While important temporal and geographic sampling gaps remain, these newly generated genomes provide substantially improved context for interpreting EEEV diversity, evolutionary relationships, and circulation in Canada. Continued collaboration among veterinary diagnostic laboratories, public health agencies, and mosquito surveillance programs will be important for expanding EEEV surveillance across human, animal, and mosquito populations in Canada.

Comparative genomic analyses showed that North American EEEV remained highly conserved with high pairwise nucleotide identity, low genome-wide entropy, and widespread evidence of purifying selection. Four codon sites showed evidence of diversifying selection, although their low frequency and unknown functional effects require further exploration. Despite high overall genomic conservation, the Canadian sequences revealed previously undocumented haplotype diversity. Many haplotypes were observed among Canadian-only samples in the current dataset, but this apparent geographic restriction does not establish that they are endemic to Canada, as uneven sampling across locations, years, and host species may have caused more broadly circulating variants to appear locally restricted.

Phylogenetic, phylogeographic, and haplotype analyses placed the Canadian genomes across multiple evolutionary groups rather than a single, independently structured lineage. While these findings confirm strong overall evolutionary constraint across the continent, they demonstrate that Canadian viruses do not represent a distinct or isolated population. Instead, repeated incursions from the broader North American population provide the strongest explanation for the overall distribution of Canadian EEEV. However, several closely related Canadian clusters included genomes from different provinces, hosts, and transmission seasons, leaving open the possibility that some introductions were followed by localized transmission or short-term persistence. High genomic conservation, sparse temporal sampling, and long unsampled intervals in the time-scaled phylogeny prevented definitive discrimination between local persistence and repeated introduction of closely related lineages. Canadian EEEV should therefore be viewed as an integrated member of a connected North American population rather than a single isolated lineage.

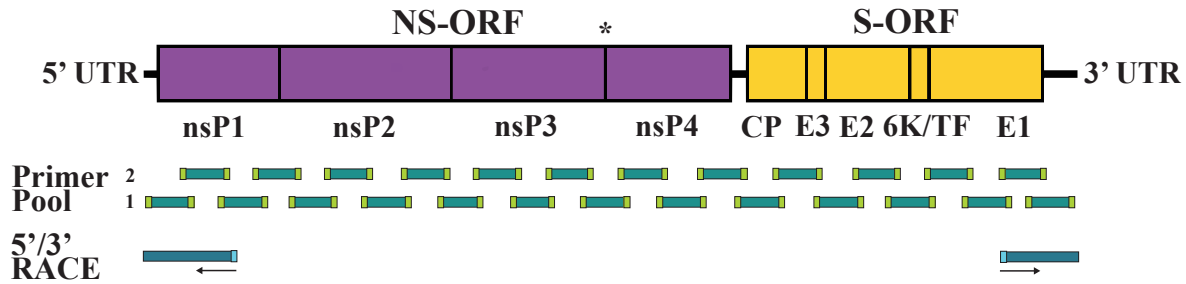
Future work should establish longitudinal genomic surveillance focused on the enzootic mosquito-bird transmission cycle. The Canadian dataset was composed mainly of diagnostic and spillover detections, particularly from horses, rather than systematic sampling of mosquito vectors and avian reservoir hosts. Repeated sampling within the same regions across successive transmission seasons would improve the ability to determine whether closely related viruses represent continued local circulation, overwintering, or repeated introductions from unsampled source populations. Geographic, temporal, and host representation should also be improved, particularly in Canadian provinces and neighbouring U.S. regions represented by few or no genomes, given that a small number of U.S. states contributed disproportionately to the continental dataset. Consistent annual sampling near the Canada-U.S. border, together with more

complete sample metadata, would improve the interpretation of geographic transitions and unusual detections. Periodic evaluation of the tiling-amplicon primers against newly available genomes will also be necessary to maintain sequencing efficiency as additional EEEV diversity is identified.

Finally, genomic surveillance would benefit from integrating genomic data with ecological and epidemiological data, including mosquito abundance, avian migration, vector distributions, wetland habitats, climatic conditions, and outbreak records. Combining these data may help distinguish ecological suitability from sampling intensity and identify the factors associated with EEEV introduction and persistence in Canada.

In conclusion, this research demonstrated the value of integrating genomics into Canadian arbovirus surveillance to improve the interpretation of viral diversity and movement, strengthen risk assessment and outbreak preparedness, and support informed measures to protect human and animal health.

APPENDICES



Supplementary Figure S1.1. Schematic representation of EEEV PrimalScheme and 5'/3' RACE. Linear viral genome map (top) showing coding and untranslated regions. Overlapping multiplex amplicon pools generated by PrimalScheme (middle) provide full-length coding sequence coverage. RACE gene-specific amplicons (bottom) overlap with the CDS to capture terminal 5'/3' UTR regions, with arrows indicating the direction of amplification.

Supplementary Table S1.1. Significant viral spatial transmission pathways estimated by Bayesian Stochastic Search Variable Selection (BSSVS).

Source	Destination	Posterior Inclusion Probability (PP)	Bayes Factor (BF) (Cutoff = 10)	Conditional Rate Mean
FL	MA	1.00	7.75x10 ⁶	2.22
FL	NY	1.00	7.75x10 ⁶	2.14
FL	QC	1.00	7.75x10 ⁶	1.04
MA	NY	1.00	7.75x10 ⁶	3.48
MA	TN	1.00	7.75x10 ⁶	3.25
MA	VA	1.00	7.75x10 ⁶	3.44
MA	NJ	1.00	90132.93	2.46
NY	AL	1.00	32280.95	0.76
MA	CT	1.00	6473.30	1.88
FL	GA	0.99	3072.99	0.47
NS	PEI	0.99	3018.24	0.99
FL	CT	0.99	1871.15	0.95
NY	VT	0.98	1200.65	0.56
QC	ON	0.97	956.72	2.96
NJ	DE	0.93	367.63	0.93
CT	MA	0.92	319.44	1.13
FL	MD	0.88	182.03	0.42
QC	NS	0.86	165.25	1.13
TN	MA	0.86	162.52	1.41
MS	Mexico	0.80	106.39	1.12
ON	SK	0.74	73.77	0.86
MS	MI	0.72	68.19	1.02
MD	MS	0.62	42.48	0.96
NJ	ON	0.62	42.03	1.16

QC	NB	0.60	39.69	0.74
NJ	NH	0.49	25.35	0.92
FL	VA	0.48	24.05	0.64
MA	NH	0.43	19.28	0.50
NJ	LA	0.41	18.32	1.05
MD	LA	0.34	13.69	0.99
NY	ON	0.33	12.69	0.78
NY	VA	0.31	11.52	0.88
ON	NB	0.28	10.19	0.77

Supplementary Table S1.2. Significant viral host-state transition pathways estimated by Bayesian Stochastic Search Variable Selection (BSSVS).

Origin Host	Destination Host	Posterior Inclusion Probability (PP)	Bayes Factor (BF) (Cutoff = 10)	Bayes Factor Support
Mosquito	Bird	1.00	1.87×10^6	Decisive (BF>100)
Mosquito	Horse	1.00	1.87×10^6	Decisive (BF>100)
Mosquito	Human	1.00	1.87×10^6	Decisive (BF>100)
Horse	Deer	0.99	951.07	Decisive (BF>100)
Bird	Alpaca	0.93	86.31	Very Strong (30<BF≤100)
Horse	Mosquito	0.89	50.87	Very Strong (30<BF≤100)
Bird	Horse	0.79	20.83	Strong (10≤BF≤30)

Supplementary Table S1.3. Geographic regions used in haplotype analysis.

Geographic Region	Geographic Location
Canada	QC, ON, NB, NS, PEI, SK
Northeastern U.S.	NY, MA, CT, NJ, NH, VT, MD, DE, VA, RI
Southeastern U.S.	FL, GA, TN, AL
Midwestern/Interior U.S.	MI
Gulf Coast U.S.	MS, LA, TX
Mexico and Caribbean	Mexico, Jamaica, Dominican Republic

Supplementary Table S1.4. Metadata for EEEV sequences analyzed in this study.

Accession Number	Location	Collection Year	Host	Accession Number	Location	Collection Year	Host
AY705240.1	FL	1991	Mosquito	KX000153.1	NJ	2003	Bird
AY705241.1	GA	1997	Human	KX000154.1	NY	2011	Mosquito
AY722102.1	MD	1940	Horse	KX000155.1	NY	2003	Horse
EF151502.1	FL	1993	Mosquito	KX000156.1	NY	2007	Mosquito

EF568607.1	NJ	1960	Mosquito	KX000157.1	VT	2013	Mosquito
HM210093.1	FL	1992	Bird	KX000158.1	NY	2011	Horse
HM210094.1	FL	1992	Mosquito	KX000159.1	NY	2008	Mosquito
HM210095.1	FL	2008	Mosquito	KX000160.1	NY	2014	Mosquito
HM210096.1	FL	1994	Mosquito	KX000161.1	NY	2011	Mosquito
HM210097.1	FL	2005	Mosquito	KX000162.1	NY	2009	Mosquito
HM210098.1	FL	2003	Horse	KX000163.1	NJ	2003	Bird
JX262386.1	VT	2011	Bird	KX000164.1	NY	2009	Human
KJ469555.1	GA	2003	Bird	KX000165.1	NY	2005	Bird
KJ469556.1	NH	2005	Mosquito	KX000166.1	NY	2011	Mosquito
KJ469557.1	Jamaica	1962	Horse	KX000167.1	NY	2013	Mosquito
KJ469559.1	Mexico	2009	Horse	KX000168.1	NY	2009	Mosquito
KJ469560.1	GA	2001	Bird	KX000169.1	FL	2002	Bird
KJ469561.1	FL	2005	Horse	KX000170.1	NY	2005	Mosquito
KJ469563.1	CT	1996	Mosquito	KX000171.1	NY	2007	Mosquito
KJ469564.1	MD	1988	Mosquito	KX000172.1	NY	2010	Mosquito
KJ469566.1	NJ	1935	Horse	KX000173.1	NY	2008	Mosquito
KJ469567.1	NY	1974	Mosquito	KX000174.1	NY	2011	Mosquito
KJ469568.1	FL	2005	Mosquito	KX000175.1	NY	2006	Mosquito
KJ469570.1	CT	2001	Mosquito	KX000176.1	NY	2006	Horse
KJ469572.1	FL	2001	Bird	KX000177.1	NY	2009	Mosquito
KJ469573.1	MI	1982	Mosquito	KX000178.1	NY	2013	Mosquito
KJ469575.1	CT	1996	Mosquito	KX000179.1	NY	2009	Horse
KJ469577.1	FL	1994	Bird	KX000180.1	NY	2008	Mosquito
KJ469579.1	GA	1982	Dog	KX000181.1	NY	2014	Mosquito
KJ469582.1	CT	2003	Mosquito	KX000182.1	NY	2004	Bird
KJ469583.1	DE	1934	Horse	KX000183.1	NY	2009	Mosquito
KJ469584.1	CT	2004	Mosquito	KX000184.1	NY	2014	Mosquito
KJ469585.1	FL	2002	Horse	KX000185.1	NY	2007	Mosquito
KJ469587.1	MS	1983	Human	KX000186.1	NY	2006	Mosquito
KJ469591.1	FL	2001	Bird	KX000187.1	NY	2006	Mosquito
KJ469594.1	FL	2003	Mosquito	KX000188.1	NY	2006	Mosquito
KJ469595.1	ON	1961	Horse	KX000189.1	NY	2010	Mosquito
KJ469597.1	FL	2001	Bird	KX000190.1	NY	2006	Mosquito
KJ469599.1	FL	1960	Bird	KX000191.1	NY	2010	Mosquito
KJ469600.1	NJ	1933	Horse	KX000192.1	NY	2004	Mosquito
KJ469603.1	LA	1951	Human	KX000193.1	NY	2003	Mosquito
KJ469605.1	FL	1992	Bird	KX000194.1	NY	2013	Mosquito
KJ469606.1	CT	2004	Mosquito	KX000195.1	NY	2011	Horse
KJ469607.1	MD	1990	Mosquito	KX000196.1	NY	2007	Mosquito
KJ469608.1	CT	1998	Mosquito	KX000197.1	NY	2013	Mosquito
KJ469609.1	GA	1953	Horse	KX000198.1	NY	2006	Mosquito
KJ469610.1	RI	1990	Mosquito	KX000199.1	NY	2005	Horse
KJ469611.1	MA	1977	Mosquito	KX000200.1	NY	2011	Mosquito
KJ469612.1	CT	1998	Mosquito	KX000201.1	NY	2012	Mosquito
KJ469616.1	NY	1994	Mosquito	KX000202.1	NY	2009	Mosquito
KJ469617.1	MA	1979	Mosquito	KX000203.1	NY	2004	Horse
KJ469621.1	TX	2003	Bird	KX000204.1	FL	2007	Horse
KJ469624.1	MA	1984	Mosquito	KX000205.1	NY	2010	Mosquito
KJ469627.1	FL	2003	Mosquito	KX000206.1	NY	2012	Horse
KJ469628.1	CT	2001	Mosquito	KX000207.1	NY	2006	Mosquito
KJ469630.1	NY	1990	Bird	KX000208.1	NY	2010	Mosquito
KJ469633.1	CT	2000	Mosquito	KX000209.1	NY	2011	Mosquito
KJ469634.1	MS	1989	Horse	KX000210.1	NY	2010	Horse
KJ469635.1	NY	1969	Bird	KX000211.1	NY	2009	Mosquito
KJ469636.1	DR	1978	Horse	KX000212.1	NY	2014	Mosquito
KJ469639.1	MD	1983	Mosquito	KX000213.1	NY	2008	Mosquito
KJ469642.1	CT	2003	Mosquito	KX000214.1	NY	2004	Mosquito
KJ469643.1	NY	2005	Mosquito	KX000215.1	NY	2007	Mosquito
KJ469644.1	VA	2001	Mosquito	KX000216.1	NY	2006	Mosquito
KJ469646.1	VA	2004	Mosquito	KX000217.1	NY	2009	Mosquito
KJ469647.1	VA	2008	Mosquito	KX000218.1	NY	1973	Horse
KJ469649.1	VA	2006	Mosquito	KX000219.1	NY	2010	Mosquito
KJ469650.1	VA	2009	Mosquito	KX000220.1	NY	2011	Mosquito
KJ469651.1	VA	2003	Mosquito	KX000221.1	NY	2011	Mosquito
KJ659366.1	VA	2007	Mosquito	KX000222.1	NY	2005	Horse
KJ659367.1	VA	2005	Mosquito	KX000223.1	NY	2014	Mosquito
KP282670.1	MA	2005	Human	KX000224.1	NY	2011	Horse
KR260717.2	CT	2011	Mosquito	KX000225.1	NJ	2003	Bird
KT153580.1	VT	2012	Mosquito	KX000226.1	NY	2009	Horse

KU840291.1	FL	1991	Bird	KX000230.1	NY	2012	Horse
KU840292.1	FL	2013	Horse	KX000231.1	NY	2010	Mosquito
KU840293.1	FL	2005	Horse	KX029230.1	NJ	2014	Mosquito
KU840294.1	FL	2010	Horse	KX029231.1	TN	2006	Mosquito
KU840295.1	FL	2003	Bird	KX029232.1	MA	2014	Mosquito
KU840296.1	FL	2004	Horse	KX029233.1	MA	2012	Mosquito
KU840297.1	FL	2008	Horse	KX029234.1	TN	2008	Mosquito
KU840298.1	FL	2013	Bird	KX029235.1	MA	2013	Mosquito
KU840299.1	FL	2008	Bird	KX029236.1	MA	2012	Mosquito
KU840300.1	FL	2010	Mosquito	KX029237.1	TN	2006	Mosquito
KU840301.1	FL	2014	Horse	KX029238.1	TN	2012	Mosquito
KU840302.1	FL	2004	Horse	KX029239.1	VA	2013	Mosquito
KU840303.1	FL	2005	Mosquito	KX029240.1	TN	2010	Mosquito
KU840304.1	FL	2008	Horse	KX029241.1	MA	2007	Mosquito
KU840305.1	FL	2003	Bird	KX029242.1	MA	2013	Mosquito
KU840306.1	FL	2003	Horse	KX029243.1	TN	2009	Mosquito
KU840307.1	FL	2005	Horse	KX029244.1	MA	2012	Mosquito
KU840308.1	FL	2005	Mosquito	KX029245.1	NY	2013	Mosquito
KU840309.1	FL	2001	Horse	KX029246.1	TN	2004	Mosquito
KU840310.1	FL	2001	Bird	KX029247.1	VA	2006	Mosquito
KU840311.1	FL	2008	Mosquito	KX029248.1	MA	2012	Mosquito
KU840312.1	FL	1992	Mosquito	KX029249.1	TN	2004	Mosquito
KU840313.1	FL	2010	Mosquito	KX029250.1	MA	2012	Mosquito
KU840314.1	FL	2003	Horse	KX029251.1	TN	2005	Mosquito
KU840315.1	FL	2008	Horse	KX029252.1	TN	2007	Mosquito
KU840316.1	FL	2001	Bird	KX029253.1	TN	2006	Mosquito
KU840317.1	FL	2001	Bird	KX029254.1	VA	2011	Mosquito
KU840318.1	FL	1992	Bird	KX029255.1	MA	2011	Mosquito
KU840319.1	FL	2003	Bird	KX029256.1	TN	2013	Mosquito
KU840320.1	FL	2003	Bird	KX029257.1	MA	2011	Mosquito
KU840321.1	FL	2009	Horse	KX029258.1	VA	2012	Mosquito
KU840322.1	FL	2002	Horse	KX029259.1	MA	2004	Mosquito
KU840323.1	FL	2014	Horse	KX029260.1	TN	2014	Mosquito
KU840324.1	FL	2010	Horse	KX029261.1	MA	2005	Mosquito
KU840325.1	FL	2005	Horse	KX029262.1	MA	2006	Mosquito
KU840326.1	FL	2004	Horse	KX029263.1	MA	2008	Mosquito
KU840327.1	FL	2010	Horse	KX029264.1	MA	2010	Mosquito
KU840328.1	FL	2005	Bird	KX029265.1	NJ	2004	Mosquito
KU840329.1	FL	2003	Horse	KX029266.1	NY	2006	Mosquito
KU840330.1	FL	2003	Horse	KX029267.1	TN	2011	Mosquito
KU840331.1	FL	2010	Horse	KX029268.1	NY	2009	Mosquito
KU840332.1	FL	2003	Bird	KX029269.1	MA	2013	Mosquito
KU840333.1	FL	2008	Mosquito	KX029270.1	NY	2012	Mosquito
KU840334.1	FL	2010	Horse	KX029271.1	TN	2010	Mosquito
KU840335.1	FL	2002	Mosquito	KX029272.1	TN	2004	Mosquito
KU840336.1	FL	2008	Horse	KX029273.1	NJ	2009	Mosquito
KU840337.1	FL	2001	Horse	KX029274.1	MA	2014	Mosquito
KU840338.1	FL	2014	Horse	KX029275.1	MA	2011	Mosquito
KU840339.1	FL	2013	Bird	KX029276.1	MA	2006	Mosquito
KU840340.1	FL	2003	Bird	KX029277.1	MA	2010	Mosquito
KU840341.1	FL	2013	Bird	KX029278.1	MA	2008	Mosquito
KU840342.1	FL	2001	Bird	KX029279.1	TN	2013	Mosquito
KU840343.1	FL	2002	Bird	KX029280.1	TN	2005	Mosquito
KU840344.1	FL	2008	Alpaca	KX029281.1	MA	2009	Mosquito
KU840345.1	FL	2007	Horse	KX029282.1	TN	2009	Mosquito
KU840346.1	FL	2001	Horse	KX029283.1	MA	2012	Mosquito
KU840347.1	FL	2009	Horse	KX029284.1	MA	2010	Mosquito
KU840348.1	FL	2005	Horse	KX029285.1	VA	2004	Mosquito
KU840349.1	FL	2002	Bird	KX029286.1	MA	2009	Mosquito
KU840350.1	FL	2005	Mosquito	KX029287.1	NY	2012	Mosquito
KU840351.1	FL	2001	Bird	KX029288.1	MA	2012	Mosquito
KU840352.1	FL	2013	Bird	KX029289.1	NY	2006	Mosquito
KU840353.1	FL	1991	Bird	KX029290.1	MA	2010	Mosquito
KU840354.1	FL	1994	Bird	KX029291.1	MA	2012	Mosquito
KU840355.1	FL	2014	Horse	KX029292.1	MA	2006	Mosquito
KU840356.1	FL	2003	Mosquito	KX029293.1	NY	2013	Mosquito
KU840357.1	FL	2010	Horse	KX029294.1	MA	2009	Mosquito

KU840358.1	FL	2003	Bird	KX029295.1	TN	2012	Mosquito
KU840359.1	FL	2009	Horse	KX029296.1	NJ	2012	Mosquito
KU840360.1	FL	2010	Horse	KX029297.1	MA	2006	Mosquito
KU840361.1	FL	2010	Horse	KX029298.1	NJ	2011	Mosquito
KU840362.1	FL	2005	Horse	KX029299.1	TN	2008	Mosquito
KU840363.1	FL	1994	Mosquito	KX029300.1	MA	2007	Mosquito
KU840364.1	FL	2005	Horse	KX029301.1	VA	2011	Mosquito
KU840365.1	FL	1992	Bird	KX029302.1	NY	2014	Mosquito
KU840366.1	FL	2001	Bird	KX029303.1	NY	2005	Mosquito
KU840367.1	FL	2002	Horse	KX029304.1	MA	2011	Mosquito
KU840368.1	FL	2012	Horse	KX029305.1	MA	2006	Mosquito
KU840369.1	FL	1986	Horse	KX029306.1	NY	2012	Mosquito
KU840370.1	FL	2002	Mosquito	KX029307.1	TN	2009	Mosquito
KU840371.1	FL	2003	Horse	KX029308.1	MA	2010	Mosquito
KU840372.1	MD	2008	Bird	KX029309.1	NJ	2004	Mosquito
KU840373.1	FL	2001	Bird	KX029310.1	MA	2006	Mosquito
KU840374.1	FL	2013	Deer	KX029311.1	TN	2010	Mosquito
KU840375.1	FL	2001	Bird	KX029312.1	VA	2006	Mosquito
KU840376.1	FL	2012	Horse	KX029313.1	MA	2010	Mosquito
KU840377.1	FL	2004	Bird	KX029314.1	MA	2007	Mosquito
KU840378.1	FL	1994	Bird	KX029315.1	MA	2005	Mosquito
KU840379.1	FL	2002	Bird	KX029316.1	NJ	2012	Mosquito
KX000047.1	NY	1975	Horse	KX029317.1	MA	2011	Mosquito
KX000048.1	NY	2010	Mosquito	KX029318.1	MA	2005	Mosquito
KX000049.1	NY	2011	Horse	KX029319.1	TN	2013	Mosquito
KX000050.1	VT	2012	Mosquito	MT782294.1	AL	2019	Human
KX000051.1	NY	2003	Bird	MT782295.1	AL	2019	Human
KX000052.1	NY	2005	Mosquito	OQ184109.1	Mexico	2022	Horse
KX000053.1	NY	1974	Mosquito	OQ511733.1	MA	2018	Bird
KX000054.1	NY	2010	Mosquito	OQ511734.1	CT	2019	Mosquito
KX000055.1	NY	2006	Mosquito	OQ511735.1	MA	2018	Mosquito
KX000056.1	NY	2010	Mosquito	OQ511736.1	MA	2018	Horse
KX000057.1	NY	2006	Mosquito	OQ511737.1	NY	2015	Mosquito
KX000058.1	NY	2011	Mosquito	OQ511738.1	MA	2019	Mosquito
KX000059.1	NY	2009	Mosquito	OQ511739.1	MA	2019	Mosquito
KX000060.1	NY	2014	Mosquito	OQ511740.1	MA	2019	Mosquito
KX000061.1	NY	2004	Mosquito	OQ511741.1	MA	2019	Mosquito
KX000062.1	NY	2011	Mosquito	OQ511742.1	MA	2019	Mosquito
KX000063.1	NY	2009	Mosquito	OQ511743.1	MA	2019	Mosquito
KX000064.1	AL	2009	Bird	OQ511744.1	NY	2019	Horse
KX000065.1	NY	1974	Mosquito	OQ511745.1	CT	2019	Mosquito
KX000066.1	NY	1974	Bird	OQ511746.1	CT	2019	Mosquito
KX000067.1	NY	2009	Mosquito	OQ511747.1	CT	2019	Mosquito
KX000068.1	NY	2006	Mosquito	OQ511748.1	CT	2019	Mosquito
KX000069.1	NY	1974	Mosquito	OQ511749.1	CT	2019	Mosquito
KX000070.1	NY	2008	Mosquito	OQ511750.1	CT	2019	Mosquito
KX000071.1	NY	2013	Mosquito	OQ511751.1	CT	2019	Mosquito
KX000072.1	NY	2014	Mosquito	OQ511752.1	CT	2019	Mosquito
KX000073.1	NY	2013	Mosquito	OQ511753.1	CT	2016	Mosquito
KX000074.1	NY	2009	Mosquito	OQ511754.1	CT	2017	Mosquito
KX000075.1	NY	2008	Mosquito	OQ511755.1	CT	2019	Mosquito
KX000076.1	NY	2014	Mosquito	OQ511756.1	NY	2016	Horse
KX000077.1	NY	2005	Mosquito	OQ511757.1	CT	2019	Mosquito
KX000078.1	NY	1971	Bird	OQ511758.1	CT	2019	Mosquito
KX000079.1	NJ	2003	Bird	OQ511759.1	CT	2019	Mosquito
KX000080.1	NY	2010	Mosquito	OQ511760.1	NY	2016	Mosquito
KX000081.1	NY	2006	Mosquito	OQ511761.1	NY	2017	Mosquito
KX000082.1	NY	2009	Deer	OQ511762.1	NY	2017	Mosquito
KX000083.1	NY	2013	Mosquito	OQ511763.1	NY	2017	Mosquito
KX000084.1	NY	2013	Mosquito	OQ511764.1	NY	2019	Horse
KX000085.1	NY	2006	Mosquito	OQ511765.1	NY	2018	Mosquito
KX000086.1	FL	2002	Bird	OQ511766.1	NY	2018	Mosquito
KX000087.1	NY	2007	Mosquito	OQ511767.1	NY	2018	Mosquito
KX000088.1	NY	2006	Mosquito	OQ511768.1	NY	2015	Horse
KX000089.1	VA	1933	Horse	OQ511769.1	NY	2015	Horse
KX000090.1	NY	2010	Horse	OQ511770.1	NY	2019	Mosquito
KX000091.1	NY	2004	Mosquito	OQ511771.1	NY	2015	Mosquito
KX000092.1	NY	2014	Mosquito	OQ511772.1	NY	2015	Mosquito
KX000093.1	NY	2014	Mosquito	OQ511773.1	NY	2015	Mosquito
KX000094.1	NY	2011	Mosquito	OQ511774.1	NY	2015	Mosquito

KX000095.1	NY	2004	Mosquito	OQ511775.1	MA	2019	Mosquito
KX000096.1	NY	2011	Mosquito	OQ511776.1	NY	2015	Mosquito
KX000097.1	NY	2006	Mosquito	OQ511777.1	NY	2015	Horse
KX000098.1	NY	2005	Mosquito	OQ511778.1	CT	2019	Horse
KX000099.1	NY	1969	Bird	OQ511779.1	CT	2019	Mosquito
KX000100.1	NY	2004	Mosquito	OQ511780.1	CT	2019	Mosquito
KX000101.1	FL	2002	Bird	OQ511781.1	CT	2019	Mosquito
KX000102.1	VA	2003	Bird	OQ511782.1	CT	2019	Mosquito
KX000103.1	NY	2013	Mosquito	OQ511783.1	CT	2019	Mosquito
KX000104.1	NY	2006	Mosquito	OQ511784.1	NY	2018	Mosquito
KX000105.1	NY	2008	Horse	OQ511785.1	MA	2018	Mosquito
KX000106.1	NY	2007	Mosquito	OQ511786.1	MA	2019	Mosquito
KX000107.1	NY	2009	Mosquito	OQ511787.1	VA	2019	Mosquito
KX000108.1	NY	2013	Horse	OQ511788.1	MA	2019	Mosquito
KX000109.1	NY	2014	Mosquito	OQ511789.1	CT	2017	Mosquito
KX000110.1	NY	2011	Mosquito	OQ511790.1	CT	2018	Mosquito
KX000111.1	NY	2014	Mosquito	OQ511791.1	CT	2018	Mosquito
KX000112.1	NY	2006	Mosquito	OQ511792.1	CT	2018	Mosquito
KX000113.1	NY	2013	Mosquito	OQ511793.1	CT	2017	Mosquito
KX000114.1	NY	1971	Mosquito	OQ511794.1	CT	2018	Mosquito
KX000115.1	NY	2007	Mosquito	OQ511795.1	MA	2019	Mosquito
KX000116.1	NY	2014	Mosquito	OQ511796.1	NY	2017	Mosquito
KX000117.1	NY	2008	Mosquito	OQ511797.1	CT	2019	Mosquito
KX000118.1	NY	2013	Mosquito	OQ511798.1	MA	2019	Mosquito
KX000119.1	NY	2003	Mosquito	OQ511799.1	CT	2019	Mosquito
KX000120.1	NY	1971	Horse	OQ511800.1	CT	2019	Mosquito
KX000121.1	NY	2014	Horse	OQ511801.1	CT	2019	Mosquito
KX000122.1	NY	2003	Horse	OQ511802.1	CT	2019	Mosquito
KX000123.1	NY	2014	Mosquito	OQ511803.1	CT	2019	Mosquito
KX000124.1	NY	2004	Horse	OQ511804.1	CT	2019	Mosquito
KX000125.1	NY	2008	Mosquito	OQ511805.1	CT	2019	Mosquito
KX000126.1	NY	2014	Horse	OQ511806.1	CT	2019	Mosquito
KX000127.1	NY	2010	Mosquito	OQ511807.1	CT	2019	Mosquito
KX000128.1	NY	2009	Mosquito	OQ511808.1	CT	2019	Mosquito
KX000129.1	NY	2011	Mosquito	OQ511809.1	NY	2019	Mosquito
KX000130.1	NY	2004	Mosquito	OQ511810.1	NY	2019	Mosquito
KX000131.1	NY	2008	Mosquito	OQ511811.1	NY	2016	Horse
KX000132.1	NY	2006	Mosquito	OQ511812.1	MA	2019	Mosquito
KX000133.1	NY	2014	Mosquito	OQ709265.1	MA	1938	Human
KX000134.1	NY	2011	Mosquito	OQ709266.1	MA	2004	Human
KX000135.1	NY	2014	Mosquito	OQ709267.1	MA	2004	Human
KX000136.1	NY	2005	Horse	OQ709268.1	MA	2004	Human
KX000137.1	NY	2014	Horse	OQ709269.1	MA	2004	Human
KX000138.1	NY	2011	Mosquito	OQ709270.1	MA	2004	Human
KX000139.1	NY	2013	Mosquito	OQ709271.1	MA	2005	Human
KX000140.1	NY	2010	Horse	OQ709272.1	MA	2012	Human
KX000141.1	NY	2004	Bird	OQ709273.1	MA	2012	Human
KX000142.1	NY	2013	Mosquito	OQ709274.1	MA	2012	Human
KX000143.1	NY	2014	Horse	OQ709275.1	MA	2012	Human
KX000144.1	VA	2003	Bird	OQ709276.1	MA	2019	Human
KX000145.1	NY	2003	Bird	OQ709277.1	MA	2019	Human
KX000146.1	VT	2012	Mosquito	OQ709278.1	MA	2019	Human
KX000147.1	NY	2006	Mosquito	OQ709279.1	MA	2019	Human
KX000148.1	NY	2013	Mosquito	OQ709280.1	MA	2020	Human
KX000149.1	NY	2014	Mosquito	OR042426.1	Mexico	2022	Horse
KX000150.1	NY	2006	Mosquito	OR988088.1	AL	2023	Human
KX000151.1	NY	2009	Mosquito	U01034.1	FL	1982	Mosquito
KX000152.1	NY	2004	Mosquito				

LITERATURE CITED

- Adeyinka, O. S., Barrera, M. D., Metibemu, D. S., Boghdeh, N., Anderson, C. A., Baha, H., Crown, O., Falode, J. A., Bleach, J. L., Bliss, A. R., Hampton, T. P., Ojobor, J.-F. C., Alem, F., Narayanan, A., & Ogungbe, I. V. (2025). nsP2 Protease Inhibitor Blocks the Replication of New World Alphaviruses and Offer Protection in Mice. *ACS Infectious Diseases*, 11(1), 181–196. <https://doi.org/10.1021/acsinfecdis.4c00701>
- Aguilar, P. V., Weaver, S. C., & Basler, C. F. (2007). Capsid Protein of Eastern Equine Encephalitis Virus Inhibits Host Cell Gene Expression. *Journal of Virology*, 81(8), 3866–3876. <https://doi.org/10.1128/JVI.02075-06>
- American Association of Equine Practitioners. (2023). *AAEP Infectious Disease Guidelines: Eastern Equine Encephalitis*. American Association of Equine Practitioners. https://aaep.org/wp-content/uploads/2024/02/EEE_DZ_Guidelines_Final_2023.pdf
- Andrews, E. S., Xu, G., & Rich, S. M. (2014). Microbial communities within field-collected *Culiseta melanura* and *Coquillettidia perturbans*. *Medical and Veterinary Entomology*, 28(2), 125–132. <https://doi.org/10.1111/mve.12017>
- Armstrong, P. M., & Andreadis, T. G. (2010). Eastern Equine Encephalitis Virus in Mosquitoes and Their Role as Bridge Vectors. *Emerging Infectious Diseases*, 16(12), 1869–1874. <https://doi.org/10.3201/eid1612.100640>
- Armstrong, P. M., & Andreadis, T. G. (2022). Ecology and Epidemiology of Eastern Equine Encephalitis Virus in the Northeastern United States: An Historical Perspective. *Journal of Medical Entomology*, 59(1), 1–13. <https://doi.org/10.1093/jme/tjab077>
- 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. *Journal of Virology*, 84(2), 1014–1025. <https://doi.org/10.1128/JVI.01586-09>
- Atkins, G. J. (2013). The Pathogenesis of Alphaviruses. *ISRN Virology*, 2013, 1–22. <https://doi.org/10.5402/2013/861912>
- Baele, G., Suchard, M. A., Rambaut, A., & Lemey, P. (2016). Emerging Concepts of Data Integration in Pathogen Phylodynamics. *Systematic Biology*, syw054. <https://doi.org/10.1093/sysbio/syw054>

- Bakhiyi, B., Irace-Cima, A., Ludwig, A., Rakotoarinia, M. R., Therrien, C., Dusfour, I., & Adam-Poupart, A. (2024). Public health contributions of entomological surveillance of West Nile virus (WNV) and other mosquito-borne arboviruses in a context of climate change. *Canada Communicable Disease Report*, 50(9), 394–304.
<https://doi.org/10.14745/ccdr.v50i09a02>
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews*, 35(3), 235–241. <https://doi.org/10.1128/br.35.3.235-241.1971>
- Batzer, D. P., Mehmood, A., Mead, D. G., & Champagne, D. E. (2020). Phenology of *Coquillettidia perturbans* and *Culiseta melanura* (Diptera: Culicidae) in East-Central Georgia, USA: Implications for the Ecology of Eastern Equine Encephalitis Virus. *Journal of Entomological Science*, 55(2), 156. <https://doi.org/10.18474/0749-8004-55.2.156>
- Beerenwinkel, N., Günthard, H. F., Roth, V., & Metzner, K. J. (2012). Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Frontiers in Microbiology*, 3. <https://doi.org/10.3389/fmicb.2012.00329>
- Belák, S., Thorén, P., LeBlanc, N., & Viljoen, G. (2009). Advances in viral disease diagnostic and molecular epidemiological technologies. *Expert Review of Molecular Diagnostics*, 9(4), 367–381. <https://doi.org/10.1586/erm.09.19>
- Bergren, N. A., Auguste, A. J., Forrester, N. L., Negi, S. S., Braun, W. A., & Weaver, S. C. (2014). Western Equine Encephalitis Virus: Evolutionary Analysis of a Declining Alphavirus Based on Complete Genome Sequences. *Journal of Virology*, 88(16), 9260–9267. <https://doi.org/10.1128/JVI.01463-14>
- Bialosuknia, S. M., Tan, Y., Zink, S. D., Koetzner, C. A., Maffei, J. G., Halpin, R. A., Mueller, E. A., Novotny, M., Shilts, M., Fedorova, N. B., Amedeo, P., Das, S. R., Pickett, B., Kramer, L. D., & Ciota, A. T. (2019). Evolutionary dynamics and molecular epidemiology of West Nile virus in New York State: 1999–2015. *Virus Evolution*, 5(2), vez020.
<https://doi.org/10.1093/ve/vez020>
- Biek, R., Pybus, O. G., Lloyd-Smith, J. O., & Didelot, X. (2015). Measurably evolving pathogens in the genomic era. *Trends in Ecology & Evolution*, 30(6), 306–313.
<https://doi.org/10.1016/j.tree.2015.03.009>

- Bielejec, F., Rambaut, A., Suchard, M. A., & Lemey, P. (2011). SPREAD: Spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics*, 27(20), 2910–2912.
<https://doi.org/10.1093/bioinformatics/btr481>
- Bingham, A. M., Burkett-Cadena, N. D., Hassan, H. K., McClure, C. J. W., & Unnasch, T. R. (2014). Field Investigations of Winter Transmission of Eastern Equine Encephalitis Virus in Florida. *The American Society of Tropical Medicine and Hygiene*, 91(4), 685–693.
<https://doi.org/10.4269/ajtmh.14-0081>
- Bingham, A. M., Graham, S. P., Burkett-Cadena, N. D., White, G. S., Hassan, H. K., & Unnasch, T. R. (2012). Detection of Eastern Equine Encephalomyelitis Virus RNA in North American Snakes. *The American Journal of Tropical Medicine and Hygiene*, 87(6), 1140–1144. <https://doi.org/10.4269/ajtmh.2012.12-0257>
- Blosser, E. M., Lord, C. C., Stenn, T., Acevedo, C., Hassan, H. K., Reeves, L. E., Unnasch, T. R., & Burkett-Cadena, N. D. (2017). Environmental Drivers of Seasonal Patterns of Host Utilization by *Culiseta melanura* (Diptera: Culicidae) in Florida. *Journal of Medical Entomology*, 54(5), 1365–1374. <https://doi.org/10.1093/jme/tjx140>
- Bonenfant, Q., Noé, L., & Touzet, H. (2023). Porechop_ABI: Discovering unknown adapters in Oxford Nanopore Technology sequencing reads for downstream trimming. *Bioinformatics Advances*, 3(1), vbac085. <https://doi.org/10.1093/bioadv/vbac085>
- Bonsall, D., Ansari, M. A., Ip, C., Trebes, A., Brown, A., Klenerman, P., Buck, D., STOP-HCV Consortium, Piazza, P., Barnes, E., & Bowden, R. (2015). ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *F1000Research*, 4, 1062.
<https://doi.org/10.12688/f1000research.7111.1>
- Boonham, N., Kreuze, J., Winter, S., Van Der Vlugt, R., Bergervoet, J., Tomlinson, J., & Mumford, R. (2014). Methods in virus diagnostics: From ELISA to next generation sequencing. *Virus Research*, 186, 20–31. <https://doi.org/10.1016/j.virusres.2013.12.007>
- Bourgikos, E., Dellicour, S., Lemey, P., Feriancek, N. M., Bransfield, A. B., Breban, M. I., Misencik, M. J., Petruff, T. A., Shepard, J. J., Andreadis, T. G., Anderson, J. F., Ngo, K. A., Maffei, J. G., Dupuis, A. P., Rich, S. M., Xu, G., Sakolsky, G., Price, K. J., Metzger, M. L., ... Grubaugh, N. D. (2026). Evolutionary history of Jamestown Canyon virus

<https://doi.org/10.1016/j.cub.2026.03.076>

Brüssow, H., & Figuerola, J. (2025). The Spread of the Mosquito-Transmitted West Nile Virus in North America and Europe. *Microbial Biotechnology*, 18(3), e70120.
<https://doi.org/10.1111/1751-7915.70120>

Calisher, C. H. (1994). Medically important arboviruses of the United States and Canada. *Clinical Microbiology Reviews*, 7(1), 89–116. <https://doi.org/10.1128/CMR.7.1.89>

Campbell, L. P., Guralnick, R. P., Giordano, B. V., Sallam, M. F., Bauer, A. M., Tavares, Y., Allen, J. M., Efstathion, C., Bartlett, S., Wishard, R., Xue, R.-D., Allen, B., Tressler, M., Qualls, W., & Burkett-Cadena, N. D. (2022). Spatiotemporal Modeling of Zoonotic Arbovirus Transmission in Northeastern Florida Using Sentinel Chicken Surveillance and Earth Observation Data. *Remote Sensing*, 14(14), 3388.
<https://doi.org/10.3390/rs14143388>

Carey, B. D., Bakovic, A., Callahan, V., Narayanan, A., & Kehn-Hall, K. (2019). New World alphavirus protein interactomes from a therapeutic perspective. *Antiviral Research*, 163, 125–139. <https://doi.org/10.1016/j.antiviral.2019.01.015>

- Carman, P. S., Artsob, H., Emery, S., Maxie, M. G., Pooley, D., Barker, I. K., & Mahdy, M. S. (1995). *Eastern equine encephalitis in a horse from southwestern Ontario*. 36, 170–172.
- Cassedy, A., Parle-McDermott, A., & O’Kennedy, R. (2021). Virus Detection: A Review of the Current and Emerging Molecular and Immunological Methods. *Frontiers in Molecular Biosciences*, 8, 637559. <https://doi.org/10.3389/fmolb.2021.637559>
- Chénier, S., Côté, G., Vanderstock, J., Macieira, S., Laperle, A., & Hélie, P. (2010). An eastern equine encephalomyelitis (EEE) outbreak in Quebec in the fall of 2008. *Can Vet J*, 51, 1011–1015.
- Chrzastek, K., Lee, D., Smith, D., Sharma, P., Suarez, D. L., Pantin-Jackwood, M., & Kapczynski, D. R. (2017). Use of Sequence-Independent, Single-Primer-Amplification (SISPA) for rapid detection, identification, and characterization of avian RNA viruses. *Virology*, 509, 159–166. <https://doi.org/10.1016/j.virol.2017.06.019>
- Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & De Hoon, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>
- Cohen, S. B., Lewoczko, K., Huddleston, D. B., Moody, E., Mukherjee, S., Dunn, J. R., Jones, T. F., Wilson, R., & Moncayo, A. C. (2009). Host Feeding Patterns of Potential Vectors of Eastern Equine Encephalitis Virus at an Epizootic Focus in Tennessee. *The American Journal of Tropical Medicine and Hygiene*, 81(3), 452–456. <https://doi.org/10.4269/ajtmh.2009.81.452>
- Conway, M. J., Colpitts, T. M., & Fikrig, E. (2014). Role of the Vector in Arbovirus Transmission. *Annual Review of Virology*, 1(1), 71–88. <https://doi.org/10.1146/annurev-virology-031413-085513>
- Corrin, T., Ackford, R., Mascarenhas, M., Greig, J., & Waddell, L. A. (2021). Eastern Equine Encephalitis Virus: A Scoping Review of the Global Evidence. *Vector-Borne and Zoonotic Diseases*, 21(5), 305–320. <https://doi.org/10.1089/vbz.2020.2671>
- De Coster, W., & Rademakers, R. (2023). NanoPack2: Population-scale evaluation of long-read sequencing data. *Bioinformatics*, 39(5), btad311. <https://doi.org/10.1093/bioinformatics/btad311>

- De La Calle-Prieto, F., Arsuaga, M., Rodríguez-Sevilla, G., Paiz, N. S., & Díaz-Menéndez, M. (2024). The current status of arboviruses with major epidemiological significance in Europe. *Enfermedades Infecciosas y Microbiología Clínica (English Ed.)*, 42(9), 516–526. <https://doi.org/10.1016/j.eimce.2024.09.008>
- De Maio, N., Wu, C.-H., O'Reilly, K. M., & Wilson, D. (2015). New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation. *PLOS Genetics*, 11(8), e1005421. <https://doi.org/10.1371/journal.pgen.1005421>
- De Souza, W. M., & Weaver, S. C. (2024). Effects of climate change and human activities on vector-borne diseases. *Nature Reviews Microbiology*, 22(8), 476–491. <https://doi.org/10.1038/s41579-024-01026-0>
- Dellicour, S., Lequime, S., Vrancken, B., Gill, M. S., Bastide, P., Gangavarapu, K., Matteson, N. L., Tan, Y., Du Plessis, L., Fisher, A. A., Nelson, M. I., Gilbert, M., Suchard, M. A., Andersen, K. G., Grubaugh, N. D., Pybus, O. G., & Lemey, P. (2020). Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework. *Nature Communications*, 11(1), 5620. <https://doi.org/10.1038/s41467-020-19122-z>
- Dellicour, S., Rose, R., & Pybus, O. G. (2016). Explaining the geographic spread of emerging epidemics: A framework for comparing viral phylogenies and environmental landscape data. *BMC Bioinformatics*, 17(1), 82. <https://doi.org/10.1186/s12859-016-0924-x>
- Deng, X., Achari, A., Federman, S., Yu, G., Somasekar, S., Bártolo, I., Yagi, S., Mbala-Kingebeni, P., Kapetshi, J., Ahuka-Mundeki, S., Muyembe-Tamfum, J.-J., Ahmed, A. A., Ganesh, V., Tamhankar, M., Patterson, J. L., Ndembu, N., Mbanya, D., Kaptue, L., McArthur, C., ... Chiu, C. Y. (2020). Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nature Microbiology*, 5(3), 443–454. <https://doi.org/10.1038/s41564-019-0637-9>
- Deperasińska, I., Schulz, P., & Siwicki, A. K. (2018). Salmonid alphavirus (SAV). *Journal of Veterinary Research*, 62(1), 1–6. <https://doi.org/10.2478/jvetres-2018-0001>
- Desjardins, F., Vincent, C., & Major, M. (2000). *VETERINARY ADVISORY NO. 21*.
- Ding, G., Fu, Y., Li, B., Chen, J., Wang, J., Yin, B., Sha, W., & Liu, G. (2020). Development of a multiplex RT-PCR for the detection of major diarrhoeal viruses in pig herds in China. *Transboundary and Emerging Diseases*, 67(2), 678–685. <https://doi.org/10.1111/tbed.13385>

- Domingo, E., García-Crespo, C., Lobo-Vega, R., & Perales, C. (2021). Mutation Rates, Mutation Frequencies, and Proofreading-Repair Activities in RNA Virus Genetics. *Viruses*, 13(9), 1882. <https://doi.org/10.3390/v13091882>
- Downs, J. A., Hyzer, G., Marion, E., Smith, Z. J., Kelen, P. V., & Unnasch, T. R. (2015). Mapping eastern equine encephalitis virus risk for white-tailed deer in Michigan. *Applied Geography*, 64, 66–73. <https://doi.org/10.1016/j.apgeog.2015.09.006>
- Downs, J., Chakraborty, S., Beeman, S. P., Loraamm, R., Miley, K., & Unnasch, T. R. (2024). Effects of land use/land cover, bioclimatic, and topographic variables on the seasonal occurrence of eastern equine encephalitis virus in Florida. *Journal of Land Use Science*, 19(1), 24–35. <https://doi.org/10.1080/1747423X.2024.2321398>
- Dronina, J., Samukaite-Bubniene, U., & Ramanavicius, A. (2021). Advances and insights in the diagnosis of viral infections. *Journal of Nanobiotechnology*, 19(1), 348. <https://doi.org/10.1186/s12951-021-01081-2>
- Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8), 1969–1973. <https://doi.org/10.1093/molbev/mss075>
- Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., & Baele, G. (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evolution*, 6(2), veaa061. <https://doi.org/10.1093/ve/veaa061>
- Elias, S. P., Keenan, P., Kenney, J. L., Morris, S. R., Covino, K. M., Robinson, S., Foss, K. A., Rand, P. W., Lubelczyk, C., Lacombe, E. H., Mutebi, J.-P., Evers, D., & Smith, R. P. (2017). Seasonal Patterns in Eastern Equine Encephalitis Virus Antibody in Songbirds in Southern Maine. *Vector-Borne and Zoonotic Diseases*, 17(5), 325–330. <https://doi.org/10.1089/vbz.2016.2029>
- Estep, L. K., McClure, C. J. W., Burkett-Cadena, N. D., Hassan, H. K., Hicks, T. L., Unnasch, T. R., & Hill, G. E. (2011). A Multi-Year Study of Mosquito Feeding Patterns on Avian Hosts in a Southeastern Focus of Eastern Equine Encephalitis Virus. *The American Society of Tropical Medicine and Hygiene*, 84(5), 718–726. <https://doi.org/10.4269/ajtmh.2011.10-0586>
- Fagre, A. C., Soto, R. A., Magleby, R., Cuaderna, M. K. Q., Sun, A., Cervantes, K., Crans, S. C., Panella, N. A., Kenney, J. L., Angelus, A., Burkhalter, K. L., Woell, D., Horiuchi, K.,

- Biggerstaff, B. J., Staples, J. E., Connelly, R., Martin, S. W., & Komar, N. (2024). Enhancing Eastern Equine Encephalitis Virus Surveillance in New Jersey: Optimized Collection of *Culiseta melanura*. *Journal of the American Mosquito Control Association*, 40(2), 92–101. <https://doi.org/10.2987/23-7148>
- Faria, N. R., Suchard, M. A., Rambaut, A., & Lemey, P. (2011). Toward a quantitative understanding of viral phylogeography. *Current Opinion in Virology*, 1(5), 423–429. <https://doi.org/10.1016/j.coviro.2011.10.003>
- Faye, O., Faye, O., Diallo, D., Diallo, M., Weidmann, M., & Sall, A. A. (2013). Quantitative real-time PCR detection of Zika virus and evaluation with field-caught Mosquitoes. *Virology Journal*, 10(1), 311. <https://doi.org/10.1186/1743-422X-10-311>
- Featherstone, L. A., Zhang, J. M., Vaughan, T. G., & Duchene, S. (2022). Epidemiological inference from pathogen genomes: A review of phylodynamic models and applications. *Virus Evolution*, 8(1), veac045. <https://doi.org/10.1093/ve/veac045>
- Franklinos, L. H. V., Jones, K. E., Redding, D. W., & Abubakar, I. (2019). The effect of global change on mosquito-borne disease. *The Lancet Infectious Diseases*, 19(9), e302–e312. [https://doi.org/10.1016/S1473-3099\(19\)30161-6](https://doi.org/10.1016/S1473-3099(19)30161-6)
- Freed, N. E., Vlková, M., Faisal, M. B., & Silander, O. K. (2020). Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biology Methods and Protocols*, 5(1), bpaa014. <https://doi.org/10.1093/biomethods/bpaa014>
- Fros, J. J., & Pijlman, G. P. (2016). Alphavirus Infection: Host Cell Shut-Off and Inhibition of Antiviral Responses. *Viruses*, 8(6), 166. <https://doi.org/10.3390/v8060166>
- Frost, S. D. W., Pybus, O. G., Gog, J. R., Viboud, C., Bonhoeffer, S., & Bedford, T. (2015). Eight challenges in phylodynamic inference. *Epidemics*, 10, 88–92. <https://doi.org/10.1016/j.epidem.2014.09.001>
- Gomez, D. E., Suarez, G., De Tonnerre, D. J., Mallicote, M. F., Hernadez, J. A., Sanchez, L. C., & Mackay, R. J. (2026). Eastern equine encephalitis—104 horses (1979-2017). *Journal of Veterinary Internal Medicine*, 40(3), aalag087. <https://doi.org/10.1093/jvimsj/aalag087>
- Gómez-Carballa, A., Bello, X., Pardo-Seco, J., Martín-Torres, F., & Salas, A. (2020). Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. *Genome Research*, 30(10), 1434–1448. <https://doi.org/10.1101/gr.266221.120>

- Gould, E., Pettersson, J., Higgs, S., Charrel, R., & De Lamballerie, X. (2017). Emerging arboviruses: Why today? *One Health*, 4, 1–13.
<https://doi.org/10.1016/j.onehlt.2017.06.001>
- Government of Nova Scotia. (2026, April). *Eastern Equine Encephalitis (EEE)*.
<https://novascotia.ca/dhw/cdpc/cdc/documents/eee.pdf>
- Government of Ontario. (2022, May 24). *Horse disease surveillance* | *ontario.ca*.
<http://www.ontario.ca/page/horse-disease-surveillance>
- Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., Tan, A. L., Paul, L. M., Brackney, D. E., Grewal, S., Gurfield, N., Van Rompay, K. K. A., Isern, S., Michael, S. F., Coffey, L. L., Loman, N. J., & Andersen, K. G. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biology*, 20(1), 8. <https://doi.org/10.1186/s13059-018-1618-7>
- Grubaugh, N. D., Ladner, J. T., Lemey, P., Pybus, O. G., Rambaut, A., Holmes, E. C., & Andersen, K. G. (2018). Tracking virus outbreaks in the twenty-first century. *Nature Microbiology*, 4(1), 10–19. <https://doi.org/10.1038/s41564-018-0296-2>
- Hao, Y., Fan, Q., Yu, F., Xu, F., Qin, H., Yuan, Y., Ma, W., Zhang, D., Peng, C., Li, N., & Xiao, P. (2025). Global Transmission and Evolution of Chikungunya Virus: Origins, Adaptive Mutations, and Intercontinental Spread of the Three Genotypes. *Transboundary and Emerging Diseases*, 2025(1), 3315650. <https://doi.org/10.1155/tbed/3315650>
- Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hasan, S. S., Dey, D., Singh, S., & Martin, M. (2021). The Structural Biology of Eastern Equine Encephalitis Virus, an Emerging Viral Threat. *Pathogens*, 10(8), 973.
<https://doi.org/10.3390/pathogens10080973>
- Heberlein-Larson, L. A., Tan, Y., Stark, L. M., Cannons, A. C., Shilts, M. H., Unnasch, T. R., & Das, S. R. (2019). Complex Epidemiological Dynamics of Eastern Equine Encephalitis

- Virus in Florida. *The American Journal of Tropical Medicine and Hygiene*, 100(5), 1266–1274. <https://doi.org/10.4269/ajtmh.18-0783>
- Heled, J., & Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, 8(1), 289. <https://doi.org/10.1186/1471-2148-8-289>
- Herman, C. M. (1962). The Role of Birds in the Epizootiology of Eastern Encephalitis. *The Auk*, 79(1), 99–103. <https://doi.org/10.2307/4082454>
- Hill, V., Dellicour, S., Giovanetti, M., & Grubaugh, N. D. (2026). Phylogenetic insights into the transmission dynamics of arthropod-borne viruses. *Nature Reviews Genetics*, 27(1), 47–61. <https://doi.org/10.1038/s41576-025-00854-x>
- Hill, V., Koch, R. T., Bialosuknia, S. M., Ngo, K., Zink, S. D., Koetzner, C. A., Maffei, J. G., Dupuis, A. P., Backenson, P. B., Oliver, J., Bransfield, A. B., Misencik, M. J., Petruff, T. A., Shepard, J. J., Warren, J. L., Gill, M. S., Baele, G., Vogels, C. B. F., Gallagher, G., ... Grubaugh, N. D. (2023). Dynamics of eastern equine encephalitis virus during the 2019 outbreak in the Northeast United States. *Current Biology*, 33(12), 2515-2527.e6. <https://doi.org/10.1016/j.cub.2023.05.047>
- Hill, V., Ruis, C., Bajaj, S., Pybus, O. G., & Kraemer, M. U. G. (2021). Progress and challenges in virus genomic epidemiology. *Trends in Parasitology*, 37(12), 1038–1049. <https://doi.org/10.1016/j.pt.2021.08.007>
- Ho, S. Y. W., & Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, 23(24), 5947–5965. <https://doi.org/10.1111/mec.12953>
- Hobson, K. A., Van Wilgenburg, S. L., Dunn, E. H., Hussell, D. J. T., Taylor, P. D., & Collister, D. M. (2015). Predicting origins of passerines migrating through Canadian migration monitoring stations using stable-hydrogen isotope analyses of feathers: A new tool for bird conservation. *Avian Conservation and Ecology*, 10(1), art3. <https://doi.org/10.5751/ACE-00719-100103>
- Holmes, E. C. (2003). Molecular Clocks and the Puzzle of RNA Virus Origins. *Journal of Virology*, 77(7), 3893–3897. <https://doi.org/10.1128/JVI.77.7.3893-3897.2003>
- Holmes, E. C. (2009). *The Evolution and Emergence of RNA Viruses*. Oxford University PressOxford. <https://doi.org/10.1093/oso/9780199211128.001.0001>

- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Hyde, J. L., Chen, R., Trobaugh, D. W., Diamond, M. S., Weaver, S. C., Klimstra, W. B., & Wilusz, J. (2015). The 5' and 3' ends of alphavirus RNAs – Non-coding is not non-functional. *Virus Research*, 206, 99–107. <https://doi.org/10.1016/j.virusres.2015.01.016>
- Jenkins, G. M., Rambaut, A., Pybus, O. G., & Holmes, E. C. (2002). Rates of Molecular Evolution in RNA Viruses: A Quantitative Phylogenetic Analysis. *Journal of Molecular Evolution*, 54(2), 156–165. <https://doi.org/10.1007/s00239-001-0064-3>
- Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N. J. C., & Wickett, N. J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences*, 4(7), 1600016. <https://doi.org/10.3732/apps.1600016>
- Jones, R. T., Ant, T. H., Cameron, M. M., & Logan, J. G. (2021). Novel control strategies for mosquito-borne diseases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1818), 20190802. <https://doi.org/10.1098/rstb.2019.0802>
- Jose, J., Snyder, J. E., & Kuhn, R. J. (2009). A structural and functional perspective of alphavirus replication and assembly. *Future Microbiology*, 4(7), 837–856. <https://doi.org/10.2217/fmb.09.59>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Kantor, R. S., & Jiang, M. (2024). Considerations and Opportunities for Probe Capture Enrichment Sequencing of Emerging Viruses from Wastewater. *Environmental Science & Technology*, 58(19), 8161–8168. <https://doi.org/10.1021/acs.est.4c02638>
- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kehn-Hall, K., & Bradfute, S. B. (2022). Understanding host responses to equine encephalitis virus infection: Implications for therapeutic development. *Expert Review of Anti-Infective Therapy*, 20(12), 1551–1566. <https://doi.org/10.1080/14787210.2022.2141224>

- Khan, S. U., Ogden, N. H., Fazil, A. A., Gachon, P. H., Dueymes, G. U., Greer, A. L., & Ng, V. (2020). Current and Projected Distributions of *Aedes aegypti* and *Ae. Albopictus* in Canada and the U.S. *Environmental Health Perspectives*, 128(5), 057007. <https://doi.org/10.1289/EHP5899>
- Khan, Z., Khan, M., Khan, S. B., Liaqat, I., Alasmari, S. M. N., Altwaim, S. A., Sun, Z., & Ali, A. (2025). Haplotype network analyses, genetic diversity, and population structure of *Hyalomma anatolicum* based on cytochrome c oxidase subunit I (COI) and the large subunit ribosomal RNA (16S rRNA). *Parasitology Research*, 124(11), 142. <https://doi.org/10.1007/s00436-025-08549-2>
- Kielian, M., Chanel-Vos, C., & Liao, M. (2010). Alphavirus Entry and Membrane Fusion. *Viruses*, 2(4), 796–825. <https://doi.org/10.3390/v2040796>
- Kim, D. Y., Atasheva, S., Frolova, E. I., & Frolov, I. (2013). Venezuelan Equine Encephalitis Virus nsP2 Protein Regulates Packaging of the Viral Genome into Infectious Virions. *Journal of Virology*, 87(8), 4202–4213. <https://doi.org/10.1128/JVI.03142-12>
- Kissling, E. E., Chamberlain, B. W., Bikes, B. K., & Eidson, M. E. (1954). *STUDIES ON THE NORTH AMERICAN ARTHROPOD-BORNE ENCEPHALITIDES III. EASTERN EQUINE ENCEPHALITIS VIRUS IN WILD BIRDS.*
- Komar, N., Dohm, D. J., Turell, M. J., & Spielman, A. (1999). Eastern equine encephalitis virus in birds: Relative competence of European starlings (*Sturnus vulgaris*). *The American Journal of Tropical Medicine and Hygiene*, 60(3), 387–391. <https://doi.org/10.4269/ajtmh.1999.60.387>
- Koonin, E. V., Krupovic, M., & Agol, V. I. (2021). The Baltimore Classification of Viruses 50 Years Later: How Does It Stand in the Light of Virus Evolution? *Microbiology and Molecular Biology Reviews*, 85(3), e00053-21. <https://doi.org/10.1128/MMBR.00053-21>
- Kosakovsky Pond, S. L., & Frost, S. D. W. (2005). Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Molecular Biology and Evolution*, 22(5), 1208–1222. <https://doi.org/10.1093/molbev/msi105>
- Kril, V., Hons, M., Amadori, C., Zimmerger, C., Couture, L., Bouery, Y., Burlaud-Gaillard, J., Karpov, A., Ptchelkine, D., Thienel, A. L., Kümmerer, B. M., Desfosses, A., Jones, R., Roingeard, P., Meertens, L., Amara, A., & Reguera, J. (2024). Alphavirus nsP3 organizes

- into tubular scaffolds essential for infection and the cytoplasmic granule architecture. *Nature Communications*, 15(1), 8106. <https://doi.org/10.1038/s41467-024-51952-z>
- Kumar, A., Saini, S., Anvikar, A., Mishra, N., & Misra, G. (2026). Evolving Landscape of Emerging Virus Diagnosis: Challenges and Innovations. *Molecular Biotechnology*, 68(2), 386–408. <https://doi.org/10.1007/s12033-025-01385-w>
- Lam, S. K., Pitrou, A., & Seibert, S. (2015). Numba: A LLVM-based Python JIT compiler. *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, 1–6. <https://doi.org/10.1145/2833157.2833162>
- Lam, T. T.-Y., Hon, C.-C., & Tang, J. W. (2010). Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. *Critical Reviews in Clinical Laboratory Sciences*, 47(1), 5–49. <https://doi.org/10.3109/10408361003633318>
- Lambert, A. J., Martin, D. A., & Lanciotti, R. S. (2003). Detection of North American Eastern and Western Equine Encephalitis Viruses by Nucleic Acid Amplification Assays. *Journal of Clinical Microbiology*, 41(1), 379–385. <https://doi.org/10.1128/JCM.41.1.379-385.2003>
- Layan, M., Müller, N. F., Dellicour, S., De Maio, N., Bourhy, H., Cauchemez, S., & Baele, G. (2023). Impact and mitigation of sampling bias to determine viral spread: Evaluating discrete phylogeography through CTMC modeling and structured coalescent model approximations. *Virus Evolution*, 9(1), vead010. <https://doi.org/10.1093/ve/vead010>
- Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110–1116. <https://doi.org/10.1111/2041-210X.12410>
- Leland, D. S., & Ginocchio, C. C. (2007). Role of Cell Culture for Virus Detection in the Age of Technology. *Clinical Microbiology Reviews*, 20(1), 49–78. <https://doi.org/10.1128/CMR.00002-06>
- Lemey, P., Rambaut, A., Drummond, A. J., & Suchard, M. A. (2009). Bayesian Phylogeography Finds Its Roots. *PLoS Computational Biology*, 5(9), e1000520. <https://doi.org/10.1371/journal.pcbi.1000520>
- Letson, G. W., Bailey, R. E., Pearson, J., & Tsai, T. F. (1993). Eastern Equine Encephalitis (EEE): A Description of the 1989 Outbreak, Recent Epidemiologic Trends, and the

- Association of Rainfall with EEE Occurrence. *The American Journal of Tropical Medicine and Hygiene*, 49(6), 677–685. <https://doi.org/10.4269/ajtmh.1993.49.677>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, L., Xu, B., Tian, D., Wang, A., Zhu, J., Li, C., Li, N., Zhao, W., Shi, L., Xue, Y., Zhang, Z., Bao, Y., Zhao, W., & Song, S. (2023). McAN: A novel computational algorithm and platform for constructing and visualizing haplotype networks. *Briefings in Bioinformatics*, 24(3), bbad174. <https://doi.org/10.1093/bib/bbad174>
- Lindsey, N. P., Martin, S. W., Staples, J. E., & Fischer, M. (2020). *Notes from the Field: Multistate Outbreak of Eastern Equine Encephalitis Virus—United States, 2019*. *MMWR. Morbidity and Mortality Weekly Report*, 69(2), 50–51. <https://doi.org/10.15585/mmwr.mm6902a4>
- Ludwig, A., Zheng, H., Vrbova, L., Drebot, M., Iranpour, M., & Lindsay, L. (2019). Increased risk of endemic mosquito-borne diseases in Canada due to climate change. *Canada Communicable Disease Report*, 45(4), 91–97. <https://doi.org/10.14745/ccdr.v45i04a03>
- Luo, D., Tan, Y. B., Law, M. C. Y., & Jin, J. (2025). *A Structural Perspective on the Alphavirus Life Cycle*.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), 10. <https://doi.org/10.14806/ej.17.1.200>
- McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>
- McLean, R. G., Crans, W. J., Caccamise, D. F., McNelly, J., Kirk, L. J., Mitchell, C. J., & Calisher, C. H. (1995). EXPERIMENTAL INFECTION OF WADING BIRDS WITH EASTERN EQUINE ENCEPHALITIS VIRUS. *Journal of Wildlife Diseases*, 31(4), 502–508. <https://doi.org/10.7589/0090-3558-31.4.502>
- Mencattelli, G., Ndione, M. H. D., Silverj, A., Diagne, M. M., Curini, V., Teodori, L., Di Domenico, M., Mbaye, R., Leone, A., Marcacci, M., Gaye, A., Ndiaye, E., Diallo, D., Ancora, M., Secondini, B., Di Lollo, V., Mangone, I., Bucciachio, A., Polci, A., ... Savini, G. (2023). Spatial and temporal dynamics of West Nile virus between Africa and Europe. *Nature Communications*, 14(1), 6440. <https://doi.org/10.1038/s41467-023-42185-7>

- Meshram, C. D., Agback, P., Shiliaev, N., Urakova, N., Mobley, J. A., Agback, T., Frolova, E. I., & Frolov, I. (2018). Multiple Host Factors Interact with the Hypervariable Domain of Chikungunya Virus nsP3 and Determine Viral Replication in Cell-Specific Mode. *Journal of Virology*, 92(16), e00838-18. <https://doi.org/10.1128/JVI.00838-18>
- Minh, B. Q., Nguyen, M. A. T., & Von Haeseler, A. (2013). Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*, 30(5), 1188–1195. <https://doi.org/10.1093/molbev/mst024>
- Mitri, M. E., Ludwig, A., Tataryn, J., Gasmi, S., Whitlock, M., Buck, P. A., & Bourgeois, A.-C. (2026). Hospitalizations associated with endemic and non-endemic mosquito-borne arboviruses in Canada, 2002–2023. *PLOS One*, 21(4), e0347106. <https://doi.org/10.1371/journal.pone.0347106>
- Molaei, G., & Andreadis, T. G. (2006). *Identification of Avian- and Mammalian-Derived Bloodmeals in Aedes vexans and Culiseta melanura (Diptera: Culicidae) and Its Implication for West Nile Virus Transmission in Connecticut, U.S.A.*
- Molaei, G., Andreadis, T. G., Armstrong, P. M., Thomas, M. C., Deschamps, T., Cuebas-Incle, E., Montgomery, W., Osborne, M., Smole, S., Matton, P., Andrews, W., Best, C., Cornine, F., Bidlack, E., & Texeira, T. (2013). Vector-Host Interactions and Epizootiology of Eastern Equine Encephalitis Virus in Massachusetts. *Vector-Borne and Zoonotic Diseases*, 13(5), 312–323. <https://doi.org/10.1089/vbz.2012.1099>
- Molaei, G., Armstrong, P. M., Abadam, C. F., Akaratovic, K. I., Kiser, J. P., & Andreadis, T. G. (2015). Vector-Host Interactions of Culiseta melanura in a Focus of Eastern Equine Encephalitis Virus Activity in Southeastern Virginia. *PLOS ONE*, 10(9), e0136743. <https://doi.org/10.1371/journal.pone.0136743>
- Molaei, G., Armstrong, P. M., Graham, A. C., Kramer, L. D., & Andreadis, T. G. (2015). Insights into the recent emergence and expansion of eastern equine encephalitis virus in a new focus in the Northern New England USA. *Parasites & Vectors*, 8(1), 516. <https://doi.org/10.1186/s13071-015-1145-2>
- Molaei, G., Oliver, J., Andreadis, T. G., Armstrong, P. M., & Howard, J. J. (2006). MOLECULAR IDENTIFICATION OF BLOOD-MEAL SOURCES IN CULISETA MELANURA AND CULISETA MORSITANS FROM AN ENDEMIC FOCUS OF EASTERN EQUINE ENCEPHALITIS VIRUS IN NEW YORK. *The American Journal*

- of Tropical Medicine and Hygiene*, 75(6), 1140–1147.
<https://doi.org/10.4269/ajtmh.2006.75.1140>
- Molaei, G., Thomas, M. C., Muller, T., Medlock, J., Shepard, J. J., Armstrong, P. M., & Andreadis, T. G. (2016). Dynamics of Vector-Host Interactions in Avian Communities in Four Eastern Equine Encephalitis Virus Foci in the Northeastern U.S. *PLOS Neglected Tropical Diseases*, 10(1), e0004347. <https://doi.org/10.1371/journal.pntd.0004347>
- Muhire, B. M., Varsani, A., & Martin, D. P. (2014). SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE*, 9(9), e108277. <https://doi.org/10.1371/journal.pone.0108277>
- Mukherjee, S., Moody, E. E., Lewokzco, K., Huddleston, D. B., Huang, J., Rowland, M. E., Wilson, R., Dunn, J. R., Jones, T. F., & Moncayo, A. C. (2012). Eastern Equine Encephalitis in Tennessee: 2002–2008. *Journal of Medical Entomology*, 49(3), 731–738. <https://doi.org/10.1603/ME11151>
- Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research*, 1(1), 2–9. [https://doi.org/10.1016/0027-5107\(64\)90047-8](https://doi.org/10.1016/0027-5107(64)90047-8)
- Mundis, S. J., Harrison, S., Pelley, D., Durand, S., & Ryan, S. J. (2022). Spatiotemporal Environmental Drivers of Eastern Equine Encephalitis Virus in Central Florida: Towards a Predictive Model for a Lethal Disease. *Journal of Medical Entomology*, 59(5), 1805–1816. <https://doi.org/10.1093/jme/tjac113>
- Mutebi, J.-P., Godsey, M., Smith, R. P., Renell, M. R., Smith, L., Robinson, S., Sears, S., & Lubelczyk, C. (2015). Prevalence of Eastern Equine Encephalitis Virus Antibodies Among White-Tailed Deer Populations in Maine. *Vector-Borne and Zoonotic Diseases*, 15(3), 210–214. <https://doi.org/10.1089/vbz.2014.1696>
- Mutebi, J.-P., Mathewson, A. A., Elias, S. P., Robinson, S., Graham, A. C., Casey, P., & Lubelczyk, C. B. (2022). Use of Cervid Serosurveys to Monitor Eastern Equine Encephalitis Virus Activity in Northern New England, United States, 2009–2017. *Journal of Medical Entomology*, 59(1), 49–55. <https://doi.org/10.1093/jme/tjab133>
- Mutebi, J.-P., Swope, B. N., Saxton-Shaw, K. D., Graham, A. C., Turmel, J. P., & Berl, E. (2012). Eastern Equine Encephalitis in Moose (*Alces americanus*) in Northeastern Vermont. *Journal of Wildlife Diseases*, 48(4), 1109–1112. <https://doi.org/10.7589/2012-03-076>

- Nasci, R., & Mitchell, C. (1996). Arbovirus titer variation in field-collected mosquitoes. *J Am Mosq Control Assoc.*
- Negi, V., Miller, A. S., & Kuhn, R. J. (2025). Advances in Viroporin Function and Structure: A Comparative Analysis of Alphavirus 6K with Well-Characterized Viroporins. *Viruses*, *17*(6), 868. <https://doi.org/10.3390/v17060868>
- Nelder, M. P., Russell, C., Moore, S., & Sider, D. (2014). *Eastern Equine Encephalitis Virus: History and Enhanced Surveillance in Ontario*. Ontario Agency for Health Protection and Promotion (Public Health Ontario).
- Oliver, J., Lukacik, G., Kokas, J., Campbell, S. R., Kramer, L. D., Sherwood, J. A., & Howard, J. J. (2018). Twenty years of surveillance for Eastern equine encephalitis virus in mosquitoes in New York State from 1993 to 2012. *Parasites & Vectors*, *11*(1), 362. <https://doi.org/10.1186/s13071-018-2950-1>
- Oliver, R. Y., Mahoney, P. J., Gurarie, E., Krikun, N., Weeks, B. C., Hebblewhite, M., Liston, G., & Boelman, N. (2020). Behavioral responses to spring snow conditions contribute to long-term shift in migration phenology in American robins. *Environmental Research Letters*, *15*(4), 045003. <https://doi.org/10.1088/1748-9326/ab71a0>
- Orton, R. J., Wright, C. F., Morelli, M. J., King, D. J., Paton, D. J., King, D. P., & Haydon, D. T. (2015). Distinguishing low frequency mutations from RT-PCR and sequence errors in viral deep sequencing data. *BMC Genomics*, *16*(1), 229. <https://doi.org/10.1186/s12864-015-1456-x>
- Parashar, B., Malviya, R., Sridhar, S. B., Wadhwa, T., Talath, S., & Shareef, J. (2025). Eastern equine encephalitis virus: Pathogenesis, immune response, and clinical manifestations. *Infectious Medicine*, *4*(1), 100167. <https://doi.org/10.1016/j.imj.2025.100167>
- Patterson, E. I., Khanipov, K., Swetnam, D. M., Walsdorf, S., Kautz, T. F., Thangamani, S., Fofanov, Y., & Forrester, N. L. (2020). Measuring Alphavirus Fidelity Using Non-Infectious Virus Particles. *Viruses*, *12*(5), 546. <https://doi.org/10.3390/v12050546>
- Payne, S. (2017). Introduction to RNA Viruses. In *Viruses* (pp. 97–105). Elsevier. <https://doi.org/10.1016/B978-0-12-803109-4.00010-6>
- Pedersen, K., Marks, D. R., Wang, E., Eastwood, G., Weaver, S. C., Goldstein, S. M., Sinnott, D. R., & DeLiberto, T. J. (2016). Widespread Detection of Antibodies to Eastern Equine Encephalitis, West Nile, St. Louis Encephalitis, and Turlock Viruses in Various Species of

- Wild Birds from Across the United States. *The American Society of Tropical Medicine and Hygiene*, 95(1), 206–211. <https://doi.org/10.4269/ajtmh.15-0840>
- Pennick, K. E., McKnight, C. A., Patterson, J. S., Latimer, K. S., Maes, R. K., Wise, A. G., & Kiupel, M. (2012). Diagnostic sensitivity and specificity of in situ hybridization and immunohistochemistry for *Eastern equine encephalitis virus* and *West Nile virus* in formalin-fixed, paraffin-embedded brain tissue of horses. *Journal of Veterinary Diagnostic Investigation*, 24(2), 333–338. <https://doi.org/10.1177/1040638711435230>
- Perumalsamy, N., Sharma, R., Subramanian, M., & Nagarajan, S. A. (2024). Hard Ticks as Vectors: The Emerging Threat of Tick-Borne Diseases in India. *Pathogens*, 13(7), 556. <https://doi.org/10.3390/pathogens13070556>
- Pisano, M. B., Torres, C., Ré, V. E., Fariás, A. A., Sánchez Seco, M. P., Tenorio, A., Campos, R., & Contigiani, M. S. (2014). Genetic and evolutionary characterization of Venezuelan equine encephalitis virus isolates from Argentina. *Infection, Genetics and Evolution*, 26, 72–79. <https://doi.org/10.1016/j.meegid.2014.05.011>
- Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. *Nature Reviews Microbiology*, 15(8), 502–510. <https://doi.org/10.1038/nrmicro.2017.45>
- Posada-Céspedes, S., Seifert, D., Topolsky, I., Jablonski, K. P., Metzner, K. J., & Beerenwinkel, N. (2021). V-pipe: A computational pipeline for assessing viral genetic diversity from high-throughput data. *Bioinformatics*, 37(12), 1673–1680. <https://doi.org/10.1093/bioinformatics/btab015>
- Pretorius, M., & Venter, M. (2017). Diagnosis of Viral Infections. In R. J. Green (Ed.), *Viral Infections in Children, Volume I* (pp. 151–182). Springer International Publishing. https://doi.org/10.1007/978-3-319-54033-7_6
- Public Health Agency of Canada. (2017). *West Nile virus and other mosquito-borne diseases surveillance in Canada: Annual edition 2016*. <https://www.canada.ca/content/dam/phac-aspc/documents/services/publications/diseases-conditions/west-nile-virus-other-mosquito-borne-disease-national-surveillance-report-2016-final-summary/west-nile-virus-other-mosquito-borne-disease-national-surveillance-report-2016-final-summary-eng.pdf>

- Public Health Agency of Canada. (2021). *Surveillance of West Nile virus* [Surveillance of West Nile virus]. <https://www.canada.ca/en/public-health/services/diseases/west-nile-virus/surveillance-west-nile-virus.html>
- Public Health Agency of Canada. (2023). *Mosquito-borne disease surveillance: Annual reports*. <https://health-infobase.canada.ca/zoonoses/archive/2023/mosquito/annual-report.html>
- Public Health Agency of Canada. (2024a). *West Nile virus and other mosquito-borne diseases surveillance in Canada: Annual edition 2022*. <https://www.canada.ca/content/dam/phac-aspc/documents/services/publications/diseases-conditions/west-nile-virus-surveillance/2022/annual-edition/annual-edition.pdf>
- Public Health Agency of Canada. (2024b, January 16). *Mosquito-borne disease surveillance: Annual reports* [Datasets;statistics;education and awareness]. <https://health-infobase.canada.ca/zoonoses/mosquito/annual-report.html>
- Public Health Agency of Canada. (2025). *Mosquito-borne disease surveillance: Seasonal update* [Datasets;statistics;education and awareness]. <https://health-infobase.canada.ca/zoonoses/mosquito/#a2>
- Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., Burton, D. R., Lewis-Ximenez, L. L., De Jesus, J. G., Giovanetti, M., Hill, S. C., Black, A., Bedford, T., Carroll, M. W., Nunes, M., ... Loman, N. J. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols*, 12(6), 1261–1276. <https://doi.org/10.1038/nprot.2017.066>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Rakotoarinia, M. R., Blanchet, F. G., Gravel, D., Lapen, D. R., Leighton, P. A., Ogden, N. H., & Ludwig, A. (2022). Effects of land use and weather on the presence and abundance of mosquito-borne disease vectors in a urban and agricultural landscape in Eastern Ontario, Canada. *PLOS ONE*, 17(3), e0262376. <https://doi.org/10.1371/journal.pone.0262376>
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, 67(5), 901–904. <https://doi.org/10.1093/sysbio/syy032>

- Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution*, 2(1), vew007. <https://doi.org/10.1093/ve/vew007>
- Ritsch, M., Cassman, N. A., Saghaei, S., & Marz, M. (2023). Navigating the Landscape: A Comprehensive Review of Current Virus Databases. *Viruses*, 15(9), 1834. <https://doi.org/10.3390/v15091834>
- Rocheleau, J.-P., Arsenault, J., Lindsay, L. R., DiBernardo, A., Kulkarni, M. A., Côté, N., & Michel, P. (2013). Eastern Equine Encephalitis Virus: High Seroprevalence in Horses from Southern Quebec, Canada, 2012. *Vector-Borne and Zoonotic Diseases*, 13(10), 712–718. <https://doi.org/10.1089/vbz.2012.1242>
- Rocheleau, J.-P., Arsenault, J., Ogden, N. H., Lindsay, L. R., Drebot, M., & Michel, P. (2017). Characterizing areas of potential human exposure to eastern equine encephalitis virus using serological and clinical data from horses. *Epidemiology and Infection*, 145(4), 667–677. <https://doi.org/10.1017/S0950268816002661>
- Rozen-Gagnon, K., Stapleford, K. A., Mongelli, V., Blanc, H., Failloux, A.-B., Saleh, M.-C., & Vignuzzi, M. (2014). Alphavirus Mutator Variants Present Host-Specific Defects and Attenuation in Mammalian and Insect Models. *PLoS Pathogens*, 10(1), e1003877. <https://doi.org/10.1371/journal.ppat.1003877>
- Sahu, S. P., Alstad, A. D., Pedersen, D. D., & Pearson, J. E. (1994). Diagnosis of Eastern Equine Encephalomyelitis Virus Infection in Horses by Immunoglobulin M and G Capture Enzyme-Linked Immunosorbent Assay. *Journal of Veterinary Diagnostic Investigation*, 6(1), 34–38. <https://doi.org/10.1177/104063879400600107>
- Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M., & Belshaw, R. (2010). Viral Mutation Rates. *Journal of Virology*, 84(19), 9733–9748. <https://doi.org/10.1128/JVI.00694-10>
- Schmitt, S. M., Cooley, T. M., Fitzgerald, S. D., Bolin, S. R., Lim, A., Schaefer, S. M., Kiupel, M., Maes, R. K., Hogle, S. A., & O'Brien, D. J. (2007). AN OUTBREAK OF EASTERN EQUINE ENCEPHALITIS VIRUS IN FREE-RANGING WHITE-TAILED DEER IN MICHIGAN. *Journal of Wildlife Diseases*, 43(4), 635–644. <https://doi.org/10.7589/0090-3558-43.4.635>
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2). <https://doi.org/10.1214/aos/1176344136>

- Sellers, R. F. (1989). Eastern Equine Encephalitis in Quebec and Connecticut, 1972: Introduction by Infected Mosquitoes on the Wind? *Can J Vet Res*, 53, 76–79.
- Shahhosseini, N., Wong, G., Frederick, C., & Kobinger, G. P. (2020). Mosquito Species Composition and Abundance in Quebec, Eastern Canada. *Journal of Medical Entomology*, 57(4), 1025–1031. <https://doi.org/10.1093/jme/tjaa020>
- Signore, A. V., Giacinti, J., Jones, M. E. B., Erdelyan, C. N. G., McLaughlin, A., Alkie, T. N., Cox, S., Lair, S., Jardine, C. M., Stevens, B., Bravo-Araya, M., Pople, N., Pybus, M. J., Hisanaga, T., Xu, W., Koziuk, J., Lung, O., Kruczkiewicz, P., Fisher, M., ... Berhane, Y. (2025). Spatiotemporal reconstruction of the North American A(H5N1) outbreak reveals successive lineage replacements by descendant reassortants. *Science AdvAnceS*.
- Sparagano, O., George, D., Giangaspero, A., & Špitalská, E. (2015). Arthropods and associated arthropod-borne diseases transmitted by migrating birds. The case of ticks and tick-borne pathogens. *Veterinary Parasitology*, 213(1–2), 61–66. <https://doi.org/10.1016/j.vetpar.2015.08.028>
- Stamm, D. D. (1968). ARBOVIRUS STUDIES IN BIRDS IN SOUTH ALABAMA, 1959–1960. *American Journal of Epidemiology*, 87(1), 127–137. <https://doi.org/10.1093/oxfordjournals.aje.a120793>
- Stanley, C. Q., McKinnon, E. A., Fraser, K. C., Macpherson, M. P., Casbourn, G., Friesen, L., Marra, P. P., Studds, C., Ryder, T. B., Diggs, N. E., & Stutchbury, B. J. M. (2015). Connectivity of wood thrush breeding, wintering, and migration sites based on range-wide tracking. *Conservation Biology*, 29(1), 164–174. <https://doi.org/10.1111/cobi.12352>
- Stephens, C. R. A., Forbes, M. R., Wilson, S., Ludwig, A., Lapen, D. R., & Mitchell, G. W. (2026). Mosquito feeding patterns in the context of West Nile and eastern equine encephalitis viruses in eastern Ontario, Canada. *Medical and Veterinary Entomology*, mve.70077. <https://doi.org/10.1111/mve.70077>
- Stobierski, M. G., Signs, K., Dinh, E., Cooley, T. M., Melotti, J., Schalow, M., Patterson, J. S., Bolin, S. R., & Walker, E. D. (2022). Eastern Equine Encephalomyelitis in Michigan: Historical Review of Equine, Human, and Wildlife Involvement, Epidemiology, Vector Associations, and Factors Contributing to Endemicity. *Journal of Medical Entomology*, 59(1), 27–40. <https://doi.org/10.1093/jme/tjab153>

- Suttle, C. A. (2007). Marine viruses—Major players in the global ecosystem. *Nature Reviews Microbiology*, 5(10), 801–812. <https://doi.org/10.1038/nrmicro1750>
- Tan, Y., Lam, T. T.-Y., Heberlein-Larson, L. A., Smole, S. C., Auguste, A. J., Hennigan, S., Halpin, R. A., Fedorova, N., Puri, V., Stockwell, T. B., Shilts, M. H., Andreadis, T., Armstrong, P. M., Tesh, R. B., Weaver, S. C., Unnasch, T. R., Ciota, A. T., Kramer, L. D., & Das, S. R. (2018). Large-Scale Complete-Genome Sequencing and Phylodynamic Analysis of Eastern Equine Encephalitis Virus Reveals Source-Sink Transmission Dynamics in the United States. *Journal of Virology*, 92(12), e00074-18. <https://doi.org/10.1128/JVI.00074-18>
- Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, 132(2), 619–633. <https://doi.org/10.1093/genetics/132.2.619>
- Tengelsen, L. A., Bowen, R. A., Royals, M. A., Campbell, G. L., Komar, N., & Craven, R. B. (2001). Response to and efficacy of vaccination against eastern equine encephalomyelitis virus in emus. *Journal of the American Veterinary Medical Association*, 218(9), 1469–1473. <https://doi.org/10.2460/javma.2001.218.1469>
- U.S. CDC. (2026a, June 3). *Historic Data (2003–2025)*. <https://www.cdc.gov/eastern-equine-encephalitis/data-maps/historic-data.html>
- U.S. CDC. (2026b, June 22). *Symptoms of Eastern Equine Encephalitis*. <https://www.cdc.gov/eastern-equine-encephalitis/symptoms-diagnosis-treatment/index.html>
- Vogels, C. B. F., Brackney, D. E., Dupuis, A. P., Robich, R. M., Fauver, J. R., Brito, A. F., Williams, S. C., Anderson, J. F., Lubelczyk, C. B., Lange, R. E., Prusinski, M. A., Kramer, L. D., Gangloff-Kaufmann, J. L., Goodman, L. B., Baele, G., Smith, R. P., Armstrong, P. M., Ciota, A. T., Dellicour, S., & Grubaugh, N. D. (2023). Phylogeographic reconstruction of the emergence and spread of Powassan virus in the northeastern United States. *Proceedings of the National Academy of Sciences*, 120(16), e2218012120. <https://doi.org/10.1073/pnas.2218012120>
- Volz, E. M., Koelle, K., & Bedford, T. (2013). Viral Phylodynamics. *PLoS Computational Biology*, 9(3), e1002947. <https://doi.org/10.1371/journal.pcbi.1002947>

- Wadas, I., & Domingues, I. (2025). Systematic Review of Phylogenetic Analysis Techniques for RNA Viruses Using Bioinformatics. *International Journal of Molecular Sciences*, 26(5), 2180. <https://doi.org/10.3390/ijms26052180>
- WDG Public Health. (2025). *A One Health Approach to Vector-Borne Disease: Tick & Mosquito Program Updates* (BH.01.APR0225.R09). Wellington-Dufferin-Guelph Public Health.
- Weaver, S. C., & Barrett, A. D. T. (2004). Transmission cycles, host range, evolution and emergence of arboviral disease. *Nature Reviews Microbiology*, 2(10), 789–801. <https://doi.org/10.1038/nrmicro1006>
- Weaver, S. C., Forrester, N. L., Liu, J., & Vasilakis, N. (2021). Population bottlenecks and founder effects: Implications for mosquito-borne arboviral emergence. *Nature Reviews Microbiology*, 19(3), 184–195. <https://doi.org/10.1038/s41579-020-00482-8>
- Weaver, S. C., Scott, T. W., & Rico-Hesse, R. (1991). Molecular evolution of eastern equine encephalomyelitis virus in North America. *Virology*, 182(2), 774–784. [https://doi.org/10.1016/0042-6822\(91\)90618-L](https://doi.org/10.1016/0042-6822(91)90618-L)
- Weaver, S. C., Winegar, R., Manger, I. D., & Forrester, N. L. (2012). Alphaviruses: Population genetics and determinants of emergence. *Antiviral Research*, 94(3), 242–257. <https://doi.org/10.1016/j.antiviral.2012.04.002>
- Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution*, 35(3), 773–777. <https://doi.org/10.1093/molbev/msx335>
- Williams, R. A. J., Sánchez-Llatas, C. J., Doménech, A., Madrid, R., Fandiño, S., Cea-Callejo, P., Gomez-Lucia, E., & Benítez, L. (2023). Emerging and Novel Viruses in Passerine Birds. *Microorganisms*, 11(9), 2355. <https://doi.org/10.3390/microorganisms11092355>
- Williamson, L. E., Gilliland, T., Yadav, P. K., Binshtein, E., Bombardi, R., Kose, N., Nargi, R. S., Sutton, R. E., Durie, C. L., Armstrong, E., Carnahan, R. H., Walker, L. M., Kim, A. S., Fox, J. M., Diamond, M. S., Ohi, M. D., Klimstra, W. B., & Crowe, J. E. (2020). Human Antibodies Protect against Aerosolized Eastern Equine Encephalitis Virus Infection. *Cell*, 183(7), 1884–1900.e23. <https://doi.org/10.1016/j.cell.2020.11.011>
- Wong, T. K. F., Ly-Trong, N., Ren, H., Demotte, P., Baños, H., Roger, A. J., Susko, E., Bielow, C., De Maio, N., Goldman, N., Hahn, M. W., Dos Reis, M., Vinh, L. S., Huttley, G.,

- Lanfear, R., & Minh, B. Q. (2026). IQ-TREE 3: Phylogenomic inference software using complex evolutionary models. *Molecular Biology and Evolution*, 43(5), msag117. <https://doi.org/10.1093/molbev/msag117>
- World Health Organization. (2024). *Vector-borne diseases*. <https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases>
- World Organisation for Animal Health. (2021). *Equine Encephalomyelitis (Eastern, Western, or Venezuelan)* (Manual of Diagnostic Tests and Vaccines for Terrestrial Animals Chapter 3.06.05). World Organisation for Animal Health. https://www.woah.org/fileadmin/Home/fr/Health_standards/tahm/3.06.05_EEE_WEE_VEE.pdf
- Wright, A. M., Lyons, K. M., Brandley, M. C., & Hillis, D. M. (2015). Which came first: The lizard or the egg? Robustness in phylogenetic reconstruction of ancestral states. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 324(6), 504–516. <https://doi.org/10.1002/jez.b.22642>
- Wylie, K. M., Wylie, T. N., Buller, R., Herter, B., Cannella, M. T., & Storch, G. A. (2018). Detection of Viruses in Clinical Samples by Use of Metagenomic Sequencing and Targeted Sequence Capture. *Journal of Clinical Microbiology*, 56(12), e01123-18. <https://doi.org/10.1128/JCM.01123-18>
- Wylie, T. N., Wylie, K. M., Herter, B. N., & Storch, G. A. (2015). Enhanced virome sequencing using targeted sequence capture. *Genome Research*, 25(12), 1910–1920. <https://doi.org/10.1101/gr.191049.115>
- Xu, Z., Peng, Y., Yang, M., Li, X., Wang, J., Zou, R., Liang, J., Fang, S., Liu, Y., & Yang, Y. (2022). Simultaneous detection of Zika, chikungunya, dengue, yellow fever, West Nile, and Japanese encephalitis viruses by a two-tube multiplex real-time RT-PCR assay. *Journal of Medical Virology*, 94(6), 2528–2536. <https://doi.org/10.1002/jmv.27658>
- Yang, Z., & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: A Markov Chain Monte Carlo Method. *Molecular Biology and Evolution*, 14(7), 717–724. <https://doi.org/10.1093/oxfordjournals.molbev.a025811>
- Young, D. S., Kramer, L. D., Maffei, J. G., Dusek, R. J., Backenson, P. B., Mores, C. N., Bernard, K. A., & Ebel, G. D. (2008). Molecular Epidemiology of Eastern Equine

- Encephalitis Virus, New York. *Emerging Infectious Diseases*, 14(03), 454–460.
<https://doi.org/10.3201/eid1403.070816>
- Zacks, M. A., & Paessler, S. (2010). Encephalitic alphaviruses. *Veterinary Microbiology*, 140(3–4), 281–286. <https://doi.org/10.1016/j.vetmic.2009.08.023>
- Zhang, Y., Gu, R., & Zhang, L. (2026). Alphavirus nsP4: An RNA-dependent RNA polymerase and an antiviral target. *Virology*, 621, 110948. <https://doi.org/10.1016/j.virol.2026.110948>
- Zhao, J., Dellicour, S., Yan, Z., Veit, M., Gill, M. S., He, W.-T., Zhai, X., Ji, X., Suchard, M. A., Lemey, P., & Su, S. (2023). Early Genomic Surveillance and Phylogeographic Analysis of Getah Virus, a Reemerging Arbovirus, in Livestock in China. *Journal of Virology*, 97(1), e01091-22. <https://doi.org/10.1128/jvi.01091-22>
- Zubair, A. S., McAlpine, L. S., & Gobeske, K. T. (2024). Virology, ecology, epidemiology, pathology, and treatment of eastern equine encephalitis. *Journal of the Neurological Sciences*, 457, 122886. <https://doi.org/10.1016/j.jns.2024.122886>