

Yale University

## EliScholar – A Digital Platform for Scholarly Publishing at Yale

---

Public Health Theses

School of Public Health

---

January 2024

# Ecological Factors Influencing The Evolution Of Jamestown Canyon Virus In The Northern United States

Ellie Bourgikos

[ellie.bourgikos@yale.edu](mailto:ellie.bourgikos@yale.edu)

Follow this and additional works at: <https://elischolar.library.yale.edu/ysphtdl>

---

### Recommended Citation

Bourgikos, Ellie, "Ecological Factors Influencing The Evolution Of Jamestown Canyon Virus In The Northern United States" (2024). *Public Health Theses*. 2377.

<https://elischolar.library.yale.edu/ysphtdl/2377>

This Open Access Thesis is brought to you for free and open access by the School of Public Health at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Public Health Theses by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact [elischolar@yale.edu](mailto:elischolar@yale.edu).

**Ecological Factors Influencing the Evolution of  
Jamestown Canyon Virus in the Northern United States**

**Ellie Bourgikos**

Department of Epidemiology of Microbial Disease  
Yale School of Public Health

A thesis submitted to the Yale School of Public Health in partial fulfillment of the requirements  
for the Degree of Master of Public Health

Primary Advisor: Nathan Grubaugh, Ph.D., M.S.  
Secondary Advisory: Philip Armstrong, Ph.D.

**ABSTRACT**

Jamestown Canyon virus (JCV) is a re-emerging arthropod-borne virus in the midwest and northeast United States, infecting mosquitoes and white-tailed deer in a seasonal transmission cycle. Its spread may contribute to thousands of asymptomatic human infections, with 194 total cases of neuroinvasive disease diagnosed. However, very little is known about the virus's evolutionary history, transmission patterns, and lineage distribution. For this project, we sequenced 689 JCV samples collected from mosquito pools in the northern U.S., increasing the availability of this species' genomes. Using maximum likelihood estimation on surveillance data, we found that JCV is maintained through the overwintering of single-brood mosquitoes, with peaks in percent infection rate and vector index corresponding with their emergence in the early summer. To corroborate these results, we coupled the surveillance data with phylogenetic analysis of our sequenced genomes. JCV's genomic diversity can be divided into 2 lineages, A and B, which are made up of 6 and 2 sub-lineages respectively. These lineages demonstrate significant geographic clustering, suggesting that ecological factors, such as white-tailed deer distribution, may restrict the virus's movement within and between U.S. states. After comparing lineage designations across the genome, 0.73% of sequences exhibited segment reassortment. It is therefore likely that JCV primarily evolves through genetic drift, with shift occurring at a rate of 0.18 events per year. Incorporating both genomic and surveillance data, this project significantly expands our understanding of JCV, untangling a complex viral transmission cycle to identify future areas for public health intervention.

## ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge the generosity of my collaborators, without whom this project would not have been possible. Phil Armstrong, who also served as my advisor, was a wealth of knowledge in the research and writing process, giving excellent feedback throughout this project's development. He and his colleagues at the Connecticut Agricultural Experiment Station, especially Doug Brackney, Angela Brasfield, John Anderson, John Shepard, Michael Micensik, and Theodore Andreadis were invaluable experts who laid the foundation of our understanding of JCV today. I would also like to thank the incredible team at the New York State Department of Health's Wadsworth Center, Alexander Ciota, Kiet Ngo, Alan Dupuis, Joseph Maffei, and Rachel Lange for their collaboration and assistance with primer assay development. Guang Xu and Stephen Rich of University of Massachusetts Amherst were also great collaborators who generously shared their JCV samples for analysis.

I am also extremely lucky to be mentored and supported by a group of amazing researchers at the Grubaugh lab. I would especially like to thank my amazing advisor Nathan Grubaugh, who has provided so much support and guidance throughout my internship and thesis project. He has been an inspiration of mine since I was an undergraduate and I could not be more proud to work with him on this research. I cannot thank him enough for the time and energy he has put into this project. My mentor Chantal Vogels is also an incredible force behind this project's research and development. She was essential to the completion of sequencing and other wet-laboratory work and served as a constant support throughout the research process. I am extremely lucky to have her as a mentor and inspiration. Together, she and Nate are a constant reminder of the standards of scientific research to which I aspire.

Within the Grubaugh Lab, I was incredibly fortunate to work with so many other wonderful researchers. Nicole Feriancek was critical to this project's wet-laboratory work and provided considerable help with the project's conceptualization and troubleshooting. I am also beyond grateful for all the amazing help from Verity Hill, whose knowledge of bioinformatics and phylodynamics was incredibly helpful from conception to completion.

**TABLE OF CONTENTS**

<b>List of Figures</b>	5
<b>List of Tables</b>	6
<b>Introduction</b>	7
<b>Review of Studies Relevant to the Problem</b>	8
<b>Research Design</b>	11
<b>Presentation and Analysis of Findings</b>	14
<b>Conclusions</b>	27
<b>References</b>	30
<b>Supplemental Materials</b>	33

## LIST OF FIGURES

**Figure 1.** Seasonality of JCV infection in univoltine and multivoltine Connecticut mosquitoes from 1997-2022.

**Figure 2.** Detection of JCV in univoltine and multivoltine mosquitoes in the northern United States, 1966-2023.

**Figure 3.** Detection of JCV lineages in the northern United States, 1966-2023.

**Figure 4.** Total viral diversity of JCV observed in concatenated genomes, 1997-2023.

**Figure 5.** Spatial and temporal distribution of JCV lineages across Connecticut, New York, and North Dakota, 1997-2023.

**Figure 6.** Identified JCV reassortment events in the northern United States, 1966-2023.

**LIST OF TABLES**

**Table 1.** Identified reassortant JCV samples across genomic data from Connecticut, New York, Massachusetts, and North Dakota, from 1966-2023.

## INTRODUCTION

Arboviruses, a group of pathogens spread through blood-feeding arthropods, represent an urgent threat to public health around the world.<sup>1,3</sup> Especially as climate change threatens to expand the ranges of arthropod vectors,<sup>3</sup> increasing the emergence of new arbovirus variants, there is an urgent need for more research in this field. One such class of pathogens is the orthobunyavirus genus, which encompasses over 170 viral species, 30 of which infect humans.<sup>3</sup> It is estimated that orthobunyaviruses contribute to hundreds of thousands of asymptomatic infections in the United States, causing sporadic but severe disease in humans.<sup>3</sup> Despite this, many of these viruses are historically understudied and poorly understood.<sup>3</sup>

Jamestown Canyon virus (JCV) is a re-emerging orthobunyavirus<sup>2</sup> garnering increased awareness in the field of public health. While its name may not be widely recognized, it is one of the most prevalent mosquito-borne viruses in North America<sup>3</sup> and an underrecognized cause of neuroinvasive disease in the United States.<sup>4</sup> Since 2011, there have been 282 diagnosed cases of JCV, distributed across 25 states.<sup>5,6</sup> These cases have resulted in 194 hospitalizations due to neuroinvasive disease and 7 deaths.<sup>5,6</sup> JCV seroprevalence has been measured as high as 20-40% in some communities,<sup>7</sup> suggesting that many infections are asymptomatic<sup>3</sup> and cases may be underdiagnosed.<sup>8</sup> Much of the recent rise in cases can be attributed to enhanced surveillance<sup>8</sup> and changes to testing protocols.<sup>2</sup> Despite its widespread prevalence and impacts on human health, JCV remains an understudied arbovirus. In order to control the spread of JCV and mitigate its risk to humans, we must develop further understanding of its transmission patterns, existing genetic diversity, and evolution.

This thesis project utilizes methods of genomic epidemiology to approach the many unanswered questions concerning JCV's distribution, transmission, and evolution. Using a two-pronged design, we can explore the spread of JCV from multiple epidemiological perspectives. As part of the first aim, I used Connecticut mosquito surveillance data to explore seasonal patterns of JCV across the 26 mosquito species known to carry the virus. This data was also used to calculate JCV infection rates and vector indexes for each of the species over time. In the second aim, I used genomic sequencing data to further inform the spatiotemporal patterns of JCV spread in the United States. This included assessing the phylogeny of each JCV segment based on mosquito species, identifying reassortment events, analyzing the virus' evolutionary rate, and evaluating the spread of different viral lineages over space and time. These analyses will be the largest phylodynamic study of JCV to date, utilizing genomic and surveillance data to explore the transmission and evolution of the virus across four U.S. states, from 1997 to present. In using these data, we can answer several questions on JCV's epidemiology, including: 1) What ecological factors drive the transmission and maintenance of JCV? 2) What determines distinctions between JCV lineages? 3) What are the virus' evolutionary patterns and how often does genetic reassortment occur?

I hypothesize that Jamestown Canyon virus is primarily maintained through overwintering in snowmelt mosquito species, which emerge and introduce the virus in the early summer of each

year; lineages of JCV are restricted by ecological factors beyond the mosquito, possibly including distributions of white-tailed deer.

## REVIEW OF STUDIES RELEVANT TO THE PROBLEM

JCV was discovered in 1961 in Jamestown, Colorado in a *Culiseta inornata* mosquito.<sup>2,3,4,9</sup> Since then, the virus has been collected in 25 states,<sup>5</sup> with virus activity most concentrated in the northern midwest and northeastern United States.<sup>2,7,8,10</sup> The highest recorded cases of human JCV infection are in Wisconsin, Minnesota, Michigan, Maine, Massachusetts, Connecticut, New Jersey, New Hampshire, and New York.<sup>2,5</sup> However, uneven surveillance across the country may complicate analysis of the virus' maintenance and spread.<sup>2</sup> In order to increase our understanding of JCV in the United States, it must be analyzed from several unique scientific perspectives, ranging from ecology to virology to epidemiology.

### Genome and Evolution

A member of the California serogroup of Orthobunyaviruses (genus *Peribunyaviridae*), this negative-sense RNA species<sup>9</sup> is a unique virus subject to unique evolutionary pressures. Its genome is separated into three segments: small (S), medium (M) and large (L). The highly conserved S segment (989 b) is responsible for the construction of the nucleocapsid protein, in addition to a non-structural S protein.<sup>3</sup> The M segment (4510 b) produces glycoproteins G<sub>N</sub>, G<sub>C</sub>, and another non-structural protein.<sup>3</sup> The L segment (6960 b) codes for the RNA dependent RNA polymerase.<sup>3</sup> As a segmented virus, JCV evolves through two primary evolutionary mechanisms. The first, genetic drift, describes the accumulation of random mutations throughout the procession of replication.<sup>3</sup> Research indicates that JCV replicates slower than other California serogroup viruses<sup>10</sup> and experiences a slow rate of evolutionary change.<sup>4</sup> However, JCV is also capable of undergoing reassortment.<sup>3,12</sup> In cases of mixed infection, different viruses may exchange whole segments with one another,<sup>9</sup> resulting in rapid genetic shift. This primarily occurs between closely related viruses and may be very difficult to detect when segments are highly similar.<sup>11</sup> Research suggests that JCV evolves primarily through genetic drift, with rare reassortant events<sup>9</sup> that are very difficult to detect in the absence of whole genome sequencing.<sup>13</sup> Previous phylogenetic analyses of JCV in the northeastern United States have identified two distinct lineages of the virus, A and B.<sup>9</sup> These lineages, along with other groups of JCV in the west and midwest, have likely evolved independently of one another, although they may co-circulate in some regions.<sup>9</sup>

### Ecology and Transmission Cycle

JCV is maintained through an understudied seasonal cycle<sup>2</sup> in which the virus spreads through mosquito and white-tailed deer populations.<sup>14</sup> Like other orthobunyaviruses, JCV likely began as an insect-only species, capable of circulating without the infection of mammalian hosts.<sup>1</sup> It has been detected in 26 different mosquito species<sup>2</sup> across the *Aedes*, *Culex*, *Culiseta*, and *Coquillettidia* genera.<sup>7</sup> The spread of JCV through these species varies throughout the year. Detection of the virus typically spans from May to September,<sup>8</sup> peaking in late June.<sup>3</sup> Early

season spread is predominated by univoltine mosquitoes, such as boreal or snowpool *Aedes* species, which produce a single large brood of offspring upon emergence.<sup>3</sup> These species have been shown to experience transovarial transmission,<sup>3</sup> causing JCV to overwinter with the mosquitoes and re-emerge annually.<sup>2,3,7</sup> Vertical transmission of JCV has been documented in univoltine *Aedes provocans* and *Aedes stimulans* species, in addition to the multivoltine *Aedes triseriatus*.<sup>2</sup> Late season spread of the virus appears to be driven primarily by multivoltine mosquitoes, such as *Anopheles* species.<sup>3</sup> These mosquitoes are capable of producing multiple broods of offspring throughout the season, contributing to the continued spread of JCV through the late summer months.<sup>3</sup> Experimentally, JCV does not impact a mosquito's lifespan, even when introduced to species that have not been identified as viral hosts, such as *Aedes aegypti* and *Aedes albopictus*.<sup>7</sup> This information is particularly concerning given the expanding range of these species throughout the United States, as they may take up and circulate the virus, primarily feeding on humans.<sup>7</sup> White-tailed deer have been identified as the primary animal reservoir for JCV, although other mammals may also participate in the enzootic cycle.<sup>2</sup> Other deer species, such as mule and black-tailed deer, could also play a role in JCV transmission where these species occur.<sup>3</sup> Seropositivity in deer has been detected as early as May and June,<sup>2</sup> corresponding with the emergence of univoltine mosquitoes. Patterns of JCV maintenance within white-tailed deer populations, though, remains largely unknown.<sup>2</sup>

### Human Disease

The first case of JCV infection in humans was reported in 1980.<sup>4</sup> The CDC introduced screening protocols for suspected cases of JCV in the United States in 2001,<sup>7</sup> with infection becoming a reportable condition in 2004.<sup>4</sup> JCV is diagnosed in humans using serological testing, requiring screening antibody-capture ELISA and JCV-specific plaque reduction neutralization tests to differentiate it from other related arboviruses.<sup>4</sup> A change in testing procedures in 2013 resulted in a significant increase in JCV diagnoses.<sup>3</sup> However, cases, especially those experiencing no or mild symptoms, are still likely underdiagnosed.<sup>7</sup> Early symptoms can include fever, malaise, and headache,<sup>7</sup> in addition to some respiratory symptoms including sore throat and cough.<sup>8</sup> This may progress to neuroinvasive disease, such as meningitis and encephalitis,<sup>9,15</sup> with approximately 48% of reported cases resulting in hospitalization.<sup>2</sup> Patients can experience prolonged sequelae,<sup>4</sup> although mortality is rare and occurs in about 2% of reported cases.<sup>7</sup> Most cases of neuroinvasive JCV occur in adults.<sup>3</sup> The median age of infection was approximately 48 years old, although the virus has been shown to infect a range of age groups.<sup>3</sup> Seasonal peaks in diagnosis of human cases correlate with JCV detection in mosquitoes.<sup>9</sup>

### JCV in Connecticut

Only two human cases of JCV have been reported in Connecticut.<sup>5</sup> However, the state's mosquito surveillance has contributed to the most substantive study of JCV in the country. The Connecticut Agricultural Experiment Station (CAES) surveilles 108 mosquito collection sites across the state which are monitored on a rotational basis, per CDC epidemiological week (epi-week), from early June to late October.<sup>3</sup> This program has curated an extensive set of mosquito surveillance data, which can be used to understand the circulation patterns of JCV in Connecticut. The virus has been consistently detected in mosquito pools across the state since 1997, emerging seasonally in the early summer and preceding other arboviruses by at least one

month.<sup>3</sup> Infection rates are highest at the end of June, before tapering off throughout the month of July.<sup>3</sup> Previous ecological research found that *Aedes canadensis*, *Aedes cantator*, *Aedes aurifer*, *Aedes abserratus*, *Coquillettidia perturbans*, and *Anopheles punctipennis* are critical species in maintaining JCV in Connecticut, with seasonality following the previously discussed univoltine and multivoltine pattern.<sup>3,9</sup> JCV in Connecticut has also been studied from a genomic perspective, leading to the discovery of two distinct viral lineages, A and B.<sup>3,4,9,16</sup>

The distribution of these lineages in Connecticut demonstrates geographic stratification, with lineage A spreading predominantly in the west and lineage B in eastern areas of the state.<sup>3,9</sup> Reassortment has not been observed between the two lineages.<sup>3</sup> While there is some overlap between their distribution, the similarity of samples in the same area across years of isolation suggests that there is limited movement of the virus between eastern and western Connecticut.<sup>9</sup>

While there were no human cases in Connecticut in 2023, JCV positive mosquitoes were trapped in Fairfield, New Haven, Hartford, Middlesex, New London, and Windham counties.<sup>5</sup>

#### JCV in New York

There have been 4 total human cases of JCV reported in the state of New York.<sup>5</sup> Surveillance by the New York State Department of Health (NYSDOH) has contributed to significant understanding of the viral spread in mosquito populations as well. JCV has been detected in New York mosquito pools every year since 2001.<sup>16</sup> Researchers have identified *Aedes communis* and *Aedes provocans* as critically important to viral spread throughout the state, with higher recorded infection rates than other collected species.<sup>3</sup> However, JCV mosquito infection rates are highly variable from year to year and do not have significant correlation with observed human cases.<sup>16</sup> *Aedes canadensis* likely presents significant risk to humans, due to its biting preference and designation as a bridge vector for other viruses.<sup>16</sup> New York is one of the few states in the United States in which vertical transmission of JCV has been directly observed (*Aedes provocans*, *Aedes stimulans*).<sup>2</sup> Researchers have also studied JCV seropositivity in deer to further understand the virus' spread. In an examination of 2007-2015 hunter-harvest white-tailed deer in New York, about 55% of individuals were JCV positive.<sup>16</sup>

As in Connecticut, JCV genomes in New York demonstrated strong geographical relationships with one another, with little observed temporal signal.<sup>16</sup> Virtually all sequenced samples fell into lineage A.<sup>16</sup> Only 1 sequence from the state was clustered in lineage B.<sup>16</sup> However, it is important to note that the RT-PCR primers and probe used in New York do not detect lineage B.<sup>16</sup> While NYSDOH also uses cell culturing to identify any arbovirus-positive samples,<sup>16</sup> it is possible that lineage B is under-identified in the state. Segment reassortment has been observed within lineage A sequences.<sup>16</sup>

New York recorded 1 human case of JCV in 2023, with JCV positive mosquitoes in Sussex, Ulster, Orange, Suffolk, Chautauqua, Cattaraugus, Oswego, Onondaga, and Madison counties.<sup>5</sup>

### JCV in Massachusetts

13 total human cases of JCV have been reported in Massachusetts.<sup>5</sup> Surveillance by the Massachusetts Department of Public Health (MDPH) has found an approximate JCV positivity rate of 0.6% in mosquito pools.<sup>4</sup> Researchers have suggested viremia in these pools is very low.<sup>4</sup> Critically, there have only been 3 historic isolations of the virus, all in western Massachusetts.<sup>4</sup> The state's primary mosquito vector is currently unknown, although *Aedes canadensis*, *Aedes intrudens*, and *Aedes abserratus* have all been heavily implicated in viral spread.<sup>4</sup> The *Aedes communis* group also participates in the life cycle, with JCV being identified in *Aedes intrudens* mosquitoes.<sup>4</sup>

Sequencing of JCV from mosquito pools in Massachusetts has been fairly limited. Only one sequence, from 2016, is publicly available.<sup>4</sup> Phylogenetic analysis of this sequence along with data from Connecticut affirmed designations of lineage A and B.<sup>4</sup> The 2016 sequence fell within lineage B,<sup>4</sup> specifically the B2 "sister clade."<sup>9</sup>

There were no human cases or isolations of the virus from mosquitoes in 2023.<sup>5</sup>

### JCV in North Dakota

There have been no human cases of JCV reported in the state of North Dakota,<sup>5</sup> although JCV positive mosquitoes have been identified in Williams and Williston counties.<sup>17</sup> As in other states, several different mosquitoes are likely vectors of JCV. The majority of viral isolations were in *Aedes vexans* pools, although these had very low infection rates.<sup>17</sup> JCV was also found in *Culiseta inornata*, *Aedes dorsalis*, *Aedes flavescens*, *Aedes melanimon*, and *Aedes trivittatus*.<sup>17</sup> Mosquito positivity in North Dakota begins in June and continues to mid-September, with researchers suggesting that the virus is "stably maintained" in the state.<sup>17</sup>

Genomic sequencing has been performed on JCV-positive mosquito pools from 2003-2006. The vast majority of these samples clustered into the same clade, which was related to sequences in Connecticut's lineage A.<sup>17</sup> As in other states, there was little evidence of temporal clustering.<sup>17</sup>

There were no human cases or isolations of the virus from mosquitoes in North Dakota in 2023.<sup>5</sup>

## **RESEARCH DESIGN**

This project employs an experimental research design to sequence and analyze available samples of JCV across the United States. Because of the limited availability of whole JCV genomes, sequencing was critical to generate the amount of genomic data necessary for analysis. The segmented nature of the viral genome, in addition to the volume of positive samples, required that new PCR primers be developed. Prior to the beginning of my thesis work, as a part of an internship with the Grubaugh Lab, and in collaboration with the NYSDOH's Wadsworth Center, I developed a novel amplicon-based sequencing method for JCV using PrimalScheme,<sup>18</sup> a tool for multiplex PCR primer development. I created two separate primer

sets which could be mixed together during sequencing to cover the genomic diversity across both A and B lineages. These primers additionally demonstrated the ability to generate full coverage consensus sequences of all three JCV segments. Based on this, the new sequencing method, JCVSeq, was validated over the summer, using cell-passaged JCV positive samples from NYSDOH. It was additionally tested on JCV-positive mosquito pool samples and determined to be an effective and efficient sequencing method to use for the remaining pool positive samples.

JCV samples were sent to the Grubaugh Lab from several different public health agencies. These included 664 from CAES, 120 from the NYSDOH, 6 from the Michigan Department of Health and Human Services (MDHHS), and 6 from the University of Massachusetts Amherst. The CAES samples included 71 mosquito pools from the state of North Dakota.<sup>17</sup> We performed qPCR to check the concentration of each of the samples and performed sequencing on all 785 unique samples using the Yale Center for Genome Analysis's Illumina NovaSeq. All consensus sequences were generated using a modified version of the Grubaugh Lab's DengueSeq bioinformatics pipeline.<sup>18</sup> This modified pipeline used two different JCV reference genomes, one for each lineage (A and B), to align sequencing reads in an iterative loop.<sup>18</sup> For the purposes of phylogenetics, only sequences with greater than 70% coverage across each of the three segments will be included in analysis. Following sequencing, 689 sequences entered the analysis process. These include 500 sequences from the state of Connecticut, 123 from New York, 5 from Massachusetts, and 58 from North Dakota. Each of these sequences also had available metadata including information regarding its date of collection, mosquito species, and geographic location to at least county-level resolution. This data set could then be coupled with mosquito surveillance data to complete the rest of the analysis process.

### Aim One: Surveillance Data

As part of my first aim, I conducted statistical analyses of CAES's mosquito surveillance data to better understand the overall prevalence and seasonal patterns of JCV. This data, spanning 1997 to 2022, includes both quantitative information regarding arboviral positivity in trapped mosquitoes throughout the state and metadata reporting trapping location, date, and species. In order to filter and organize this data for analysis, I used RStudio,<sup>20</sup> specifically utilizing the pacman,<sup>21</sup> readr,<sup>22</sup> readxl,<sup>23</sup> tidyverse,<sup>24</sup> janitor,<sup>25</sup> lubridate,<sup>26</sup> PooledInfRate,<sup>27</sup> ggpubr,<sup>28</sup> and patchwork<sup>30</sup> packages. Once the data was cleaned, I subset it by species to only include the 26 mosquito species known to carry JCV in Connecticut. These are *Aedes abserratus*, *Aedes aurifer*, *Aedes canadensis*, *Aedes cinereus*, *Aedes communis*, *Aedes excrucians*, *Aedes provocans*, *Aedes sticticus*, *Aedes stimulans*, *Aedes thibaulti*, *Aedes cantator*, *Aedes sollicitans*, *Aedes taeniorhynchus*, *Aedes triseriatus*, *Aedes trivittatus*, *Aedes vexans*, *Anopheles punctipennis*, *Anopheles quadrimaculatus*, *Anopheles walkeri*, *Psorophora ferox*, *Coquillettidia perturbans*, *Culex erraticus*, *Culex salinarius*, *Culex restuans*, *Culiseta melanura*, and *Culiseta morsitans*.<sup>2,3,9</sup> Of particular interest were *Aedes canadensis*, *Aedes cantator*, *Aedes aurifer*, *Aedes abserratus*, *Coquillettidia perturbans*, and *Anopheles punctipennis*, which have been identified as important species in maintaining JCV transmission.<sup>3,9</sup> This subset of surveillance data was then used to sum the number of mosquitoes per trap per night, creating a new abundance variable, grouped by the date, site, trap type, and species. I also created a new

positivity variable, assigning pools that tested positive for JCV a value of 1 and all other pools 0. Using this data frame, the pooled infectivity rate (pIR), or the approximate proportion of mosquitoes infected<sup>27</sup> with JCV, could be calculated. The PooledInfRate package utilizes maximum likelihood estimation to calculate the probability of JCV positivity per week as a function of mosquitoes,<sup>27,30</sup> facilitating the calculation of the vector index (VI). Here, VI describes JCV's pathogen level in a particular species or group of mosquitoes.<sup>27</sup> For all further analysis of the surveillance data, I calculated these values by species, generation-genus (Multivoltine Aedes, Multivoltine Anopheles, Multivoltine Culex, Multivoltine Culiseta, Multivoltine Psorophora, Univoltine Coquillettidia, Univoltine Aedes), and generation (Multivoltine, Univoltine) groups. By analyzing this data over long (year) and short (week) time scales, I was able to further understand the dynamics of JCV in mosquito populations in Connecticut.

#### Aim Two: Phylogenetic Analysis

As part of my second aim, I utilized bioinformatics and phylogenetics methods to analyze JCV sequencing data in order to understand the total genomic diversity of the virus in the northern United States. A critical component of this process was the creation of three Nextstrain pages corresponding to each segment of the JCV genome (S, M, and L). Nextstrain is an open source software that uses bioinformatics to construct maximum likelihood phylogenetic trees and visualize the evolutionary relationships between pathogen sequences.<sup>31</sup> Nextstrain is also capable of incorporating available metadata into its analysis, for the purpose of both tree coloring and the mapping of infections across collection sites.<sup>31</sup> Due to poor temporal resolution of available JCV samples, the clock generated by Nextstrain could not be used for analysis. All phylogenetic analysis was therefore performed on divergence trees, which visualize the genomic distance between samples. Each of the segment pages could be individually analyzed based on mosquito groups (species, generation-genus, genus) and location (on the state, county, and city level). All additional tree visualization was performed using FigTree, which allows for the colorization of nodes and branches through the importation of metadata.<sup>32</sup> The original 1961 JCV sequence was used as the tree's root across all three segments.

I also utilized the Nextstrain pages in tandem with one another to assess incidence of reassortment in my dataset. This was possible by using Nextstrain Tanglegrams, a visualization tool which aligns trees and connects matching samples in order to identify incongruent phylogenetic relationships.<sup>31</sup> I generated Tanglegrams for each possible combination of JCV segments (small vs. medium, medium vs. large, small vs. large). This allowed for the recognition of reassortant JCV samples. These samples were then removed from the final dataset, prior to the calculation of evolutionary rates.

The final component of my phylogenetic analysis required the designation of JCV lineages to distinguish between clusters and further analyze their spatial distribution. Segment trees were processed into RStudio using the ggtree package,<sup>33</sup> before being run through fastBAPS, or fast hierarchical Bayesian analysis of population structure.<sup>34</sup> This program uses a Dirichlet process mixture model to cluster phylogenetic data and identify possible clades.<sup>34</sup> The results of this model can be found in Supplementary Figure 3. As a comparison for these clades, I also ran the auspice files for each segment's Nextstrain build through Autolin. This method assigns

phylogenetic nodes a “genotype representation index” and uses greedy maximization to assign lineages based on this value.<sup>35</sup> The program then sorts its lineage outputs into levels, becoming more specific with each subsequent designation.<sup>35</sup> The results of this program, on lineage level 2, can be found in Supplementary Figure 4. I compared the results of both analyses in order to create a final lineage designation system for JCV. For the most part, the Autolin level 1 results aligned with those created in fastBAPS, with level 2 increasing the specificity of clusters. I chose to primarily follow the Autolin designation system, using the fastBAPS results to resolve any inconsistencies or missing lineage assignments. I then applied this new lineage classification system to my Nextstrain builds, allowing for phylogenetic and spatial visualization.

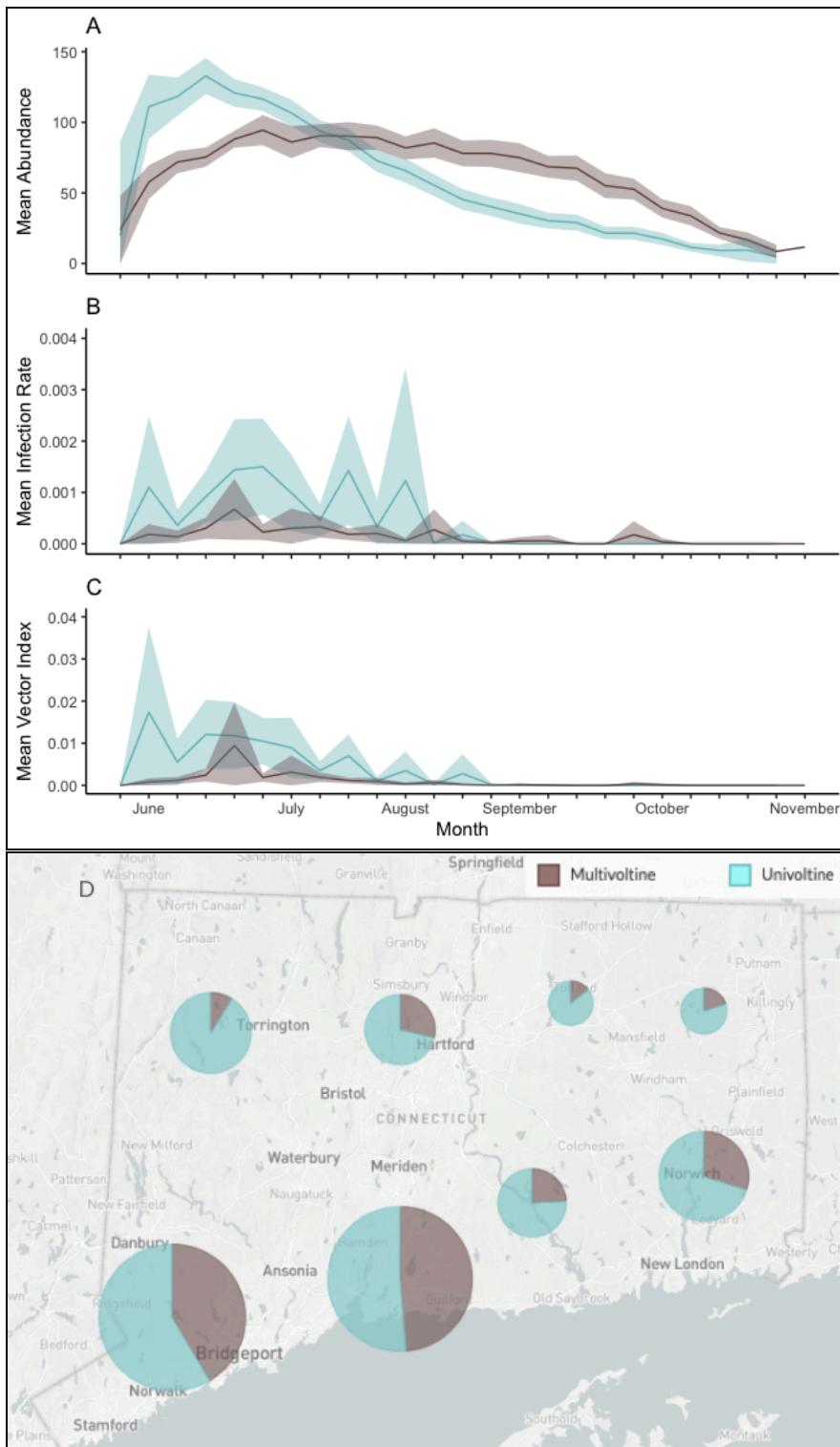
Through subsequent analysis, I found that the designations generally agreed across segments. This allowed for the concatenation of the JCV genome to summarize the virus’ phylogeny and assess its evolution and spread. Concatenation describes the process of combining segments into a single genome. In order to concatenate the genomes in my JCV dataset, I matched the sample names of each segment’s .fasta file, stringing the sequences together with small first, followed by medium and large. Prior to alignment and tree building, I removed the identified reassortant samples, in addition to all west-coast strains of JCV and samples for which lineage could not be assigned in Autolin across the first three designation levels. This final concatenated data set consisted of 672 sequences and was analyzed as a divergence phylogenetic tree. I was able to call lineages to these sequences by cross-referencing the sample’s assignment over each segment. If at least  $\frac{1}{3}$  of the segments had matching lineage assignments, this designation was used for the concatenated sequence. If the sequence failed to meet this threshold, its lineage was called as “none.” The concatenated tree could also be used for the calculation of JCV’s evolutionary rate. To do this, I used Clockor2, an application performing root-to-tip regression to determine global and local clocks for phylogenetic trees.<sup>36</sup> For the purposes of my analysis, I set the application to search for 3 local clocks across the concatenated tree, using the Bayesian Information Criterion<sup>36</sup> as the comparison metric. Because of the weak temporal signal, the heuristic root for the concatenated tree was not used for visualization purposes. Instead, I chose to visualize this tree with a midpoint root, which better demonstrates the lineages’ evolutionary history in the absence of a historic outgroup.

## **PRESENTATION AND ANALYSIS OF FINDINGS**

### JCV Dynamics in Connecticut

Prior to studying the virus itself, it is critical to understand how JCV is distributed throughout mosquito species in Connecticut. When considering the risk of infection, seasonal spread of the virus in univoltine and multivoltine mosquito populations must be considered.

Mosquitoes in Connecticut experience distinct seasonal patterns of emergence. Univoltine species emerge from vernal pools in late May to early June, rapidly expanding in population size (Fig. 1A). This results in a population spike around the middle of June, and a decline in



**Figure 1.** Seasonality of JCV infection in univoltine and multivoltine Connecticut mosquitoes from 1997-2022. (A) Mean abundance of univoltine and multivoltine mosquitoes in Connecticut per trap per epi-week, averaged from 1997-2022. (B) Mean pooled infection rate of univoltine and multivoltine mosquitoes in Connecticut per generation group per epi-week, averaged from 1997-2022. (C) Mean vector index of univoltine and multivoltine mosquitoes in Connecticut per generation group per epi-week, averaged from

1997-2022. (D) Map showing geographical distribution of JCV-infected univoltine and multivoltine mosquitoes in Connecticut by county from 1966-2022.

univoltine groups throughout the rest of the summer (Fig. 1A). In contrast, the multivoltine mosquitoes experience a slower emergence, with a more sustained population size throughout the summer months (Fig. 1A). There is not a distinct peak in their population, which instead persists throughout the late summer and into the fall (Fig. 1A).

When considering JCV positivity in these univoltine and multivoltine populations, a similar pattern can be observed. Mean percent infection rates of univoltine mosquitoes are highest from June to August (Fig. 1B). Critically, these infection rates experience an early peak corresponding with emergence (Fig. 1B), suggesting that univoltine mosquitoes are already infected when emerging from vernal pools. Across all summer months, infection rates of univoltine mosquitoes exceed those of their multi-brood counterparts (Fig. 1B). Multivoltine mosquitoes do not experience a peak in their infection rates until the third week of June (Fig. 1B). However, their populations maintain the virus for prolonged periods of time, with peaks observed in both August and the end of September (Fig. 1B). The estimated vector index for univoltine and multivoltine mosquito groups follows a similar pattern. The univoltine VI peaks with rapid emergence in early June and is sustained throughout the summer months (Fig. 1C). In contrast, the VI for multivoltine mosquitoes reaches its highest level at the end of June and beginning of July (Fig. 1C). In general, their vector index is much lower than that of the univoltine group (Fig. 1C). It is important to note that there is some variation in the peaks of infections and vector index from year to year (Supp Fig. 2). However, the seasonal patterns of these peaks still hold.

When examining the distribution of univoltine and multivoltine mosquito populations in Connecticut from a spatial perspective, there is no clear pattern (Fig. 1D). Counties further south may have a greater proportion of multivoltine species collected in comparison to the north (Fig. 1D). Still, this does not appear to be a significant difference. Collection of mosquitoes was even across generation groups from year to year. (Supp. Fig. 1).

Based on this mosquito surveillance data, it appears that JCV is primarily maintained through univoltine species, which rapidly emerge infected with the virus in the early summer and pose an immediate risk to mammalian and human populations.

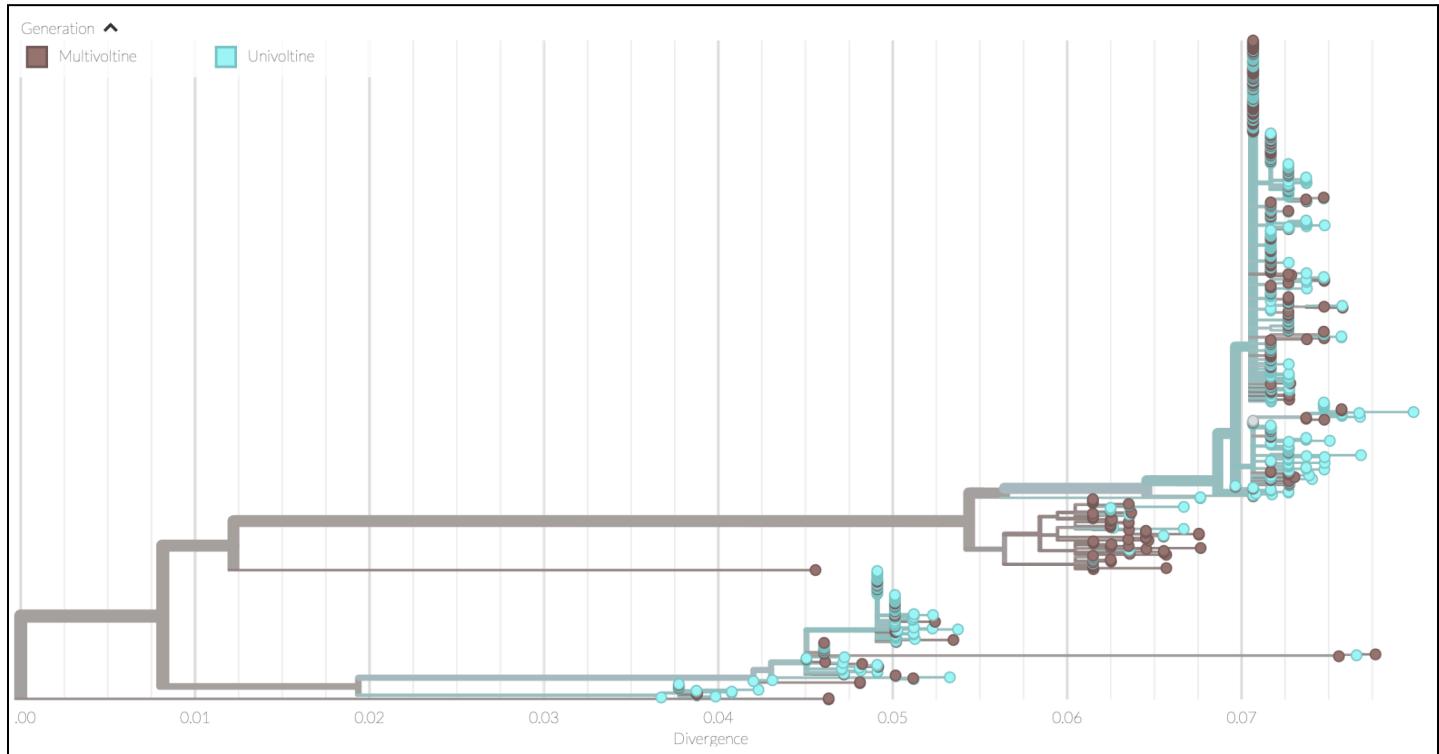
#### Maintenance of JCV in the Northeast

These theories of viral maintenance can also be explored from a genomic perspective, as mosquito metadata and sequences are combined to infer evolutionary relationships between samples. This genomic data may also elucidate maintenance patterns in states beyond Connecticut, as it includes sequences from New York, Massachusetts, and North Dakota.

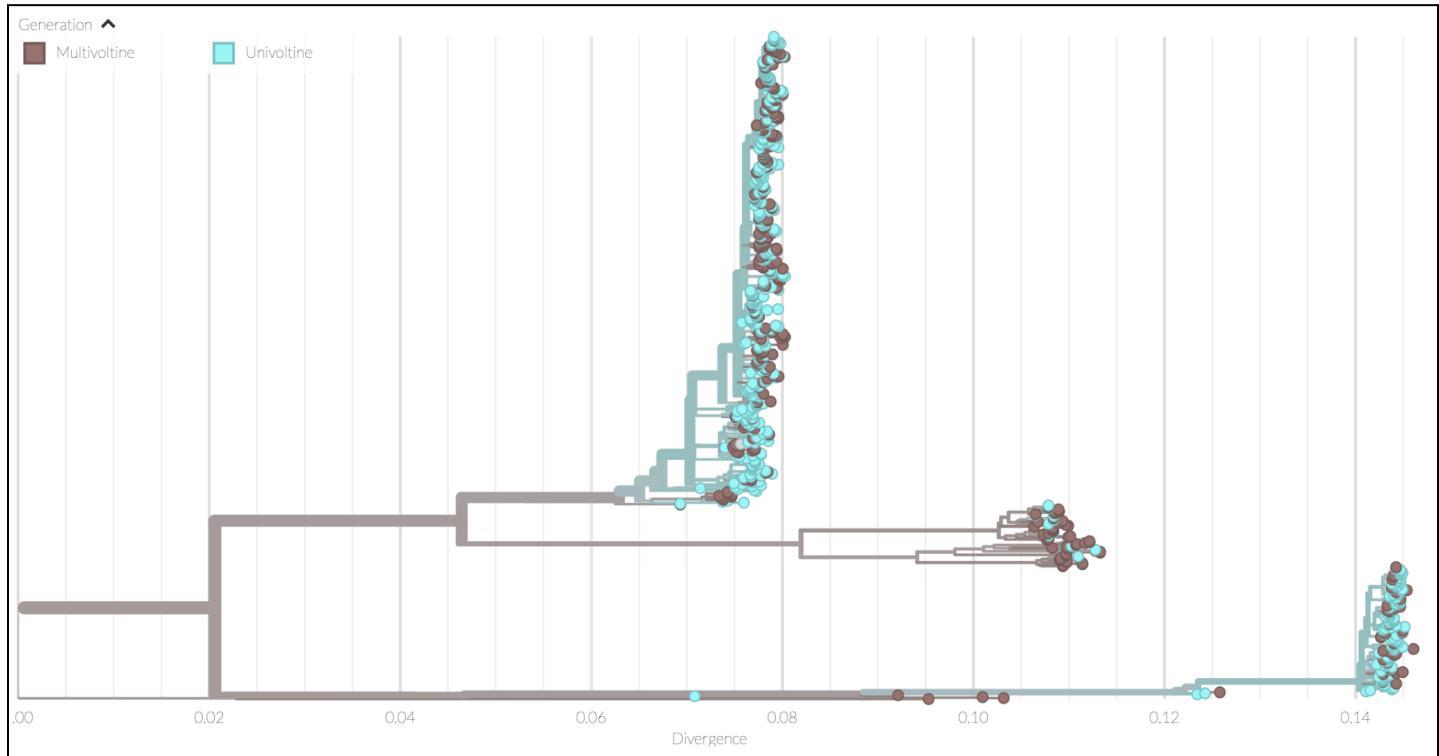
As JCV was first discovered in a multivoltine species, as were historical samples from the western United States, all segment trees have an early predominance of multivoltine mosquitoes in their phylogenetic backbone (Fig. 2A-C). This changes, though, when analyzing more modern sequences. The two most distinct groups in each segment, observed at the top

and the bottom of each tree, show univoltine predominance (Fig. 2A-C). While there are some multivoltine clades (Fig. 2A-C), a majority of the evolutionary history appears to be in univoltine mosquitoes. This univoltine backbone suggests that such species are primarily responsible for the maintenance of JCV from year to year. Vertical transmission in univoltine species may also

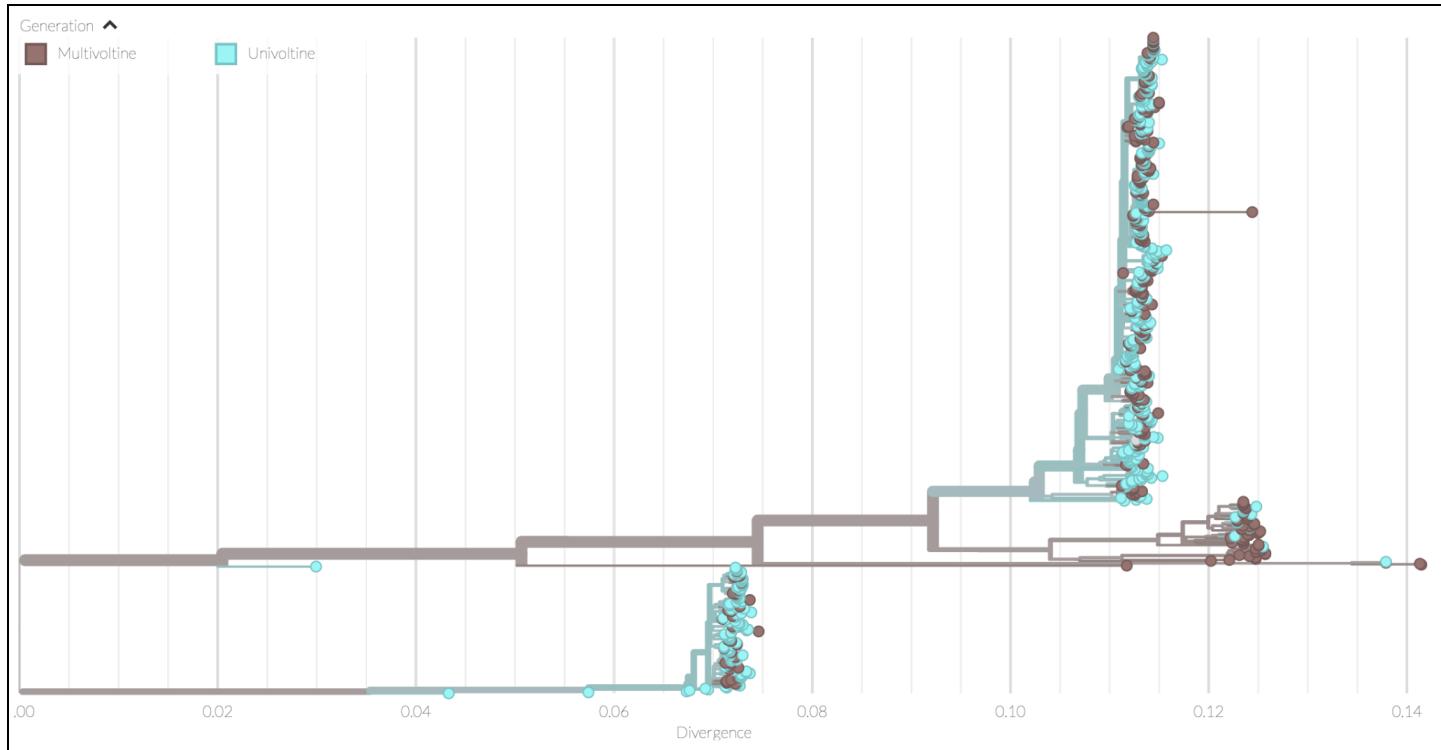
#### A. Small segment



#### B. Medium Segment



## C. Large Segment



**Figure 2.** Detection of JCV in univoltine and multivoltine mosquitoes in the northern United States, 1966-2023. (A) Maximum likelihood tree of the JCV small segment, colored by the generation group of the sample's mosquito pool. (B) Maximum likelihood tree of the JCV medium segment, colored by the generation group of the sample's mosquito pool. (C) Maximum likelihood tree of the JCV small segment, colored by the generation group of the sample's mosquito pool. Branches colored according to coloring in Nextstrain.

explain the high level of genetic relatedness between viruses from different years.

The cluster in the center of each segment tree, however, is strongly rooted in multivoltine mosquito species (Fig. 2A-C). Virtually all of these samples were collected in North Dakota. *Aedes vexans*, a multivoltine mosquito, is a critical vector in this state, which may account for the observed cluster.

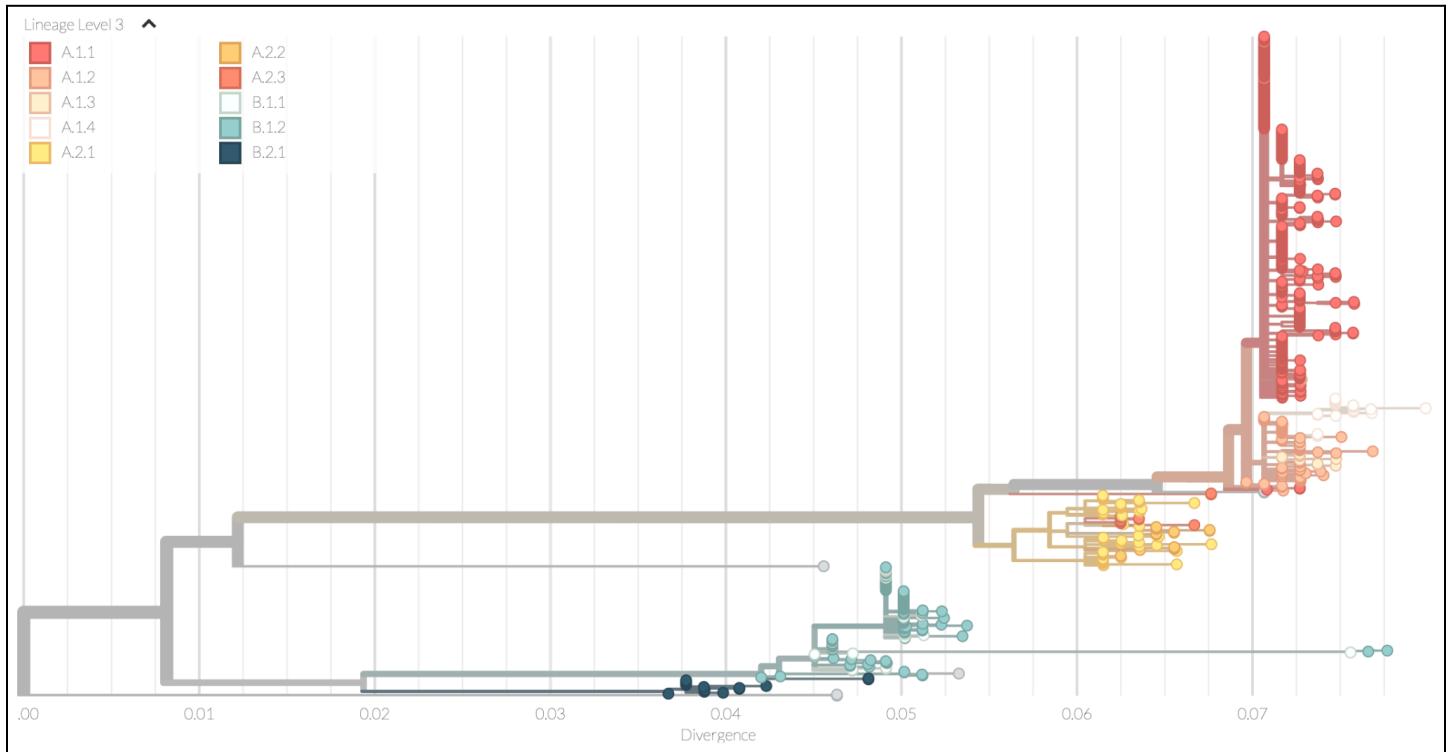
This suggests that JCV maintenance may vary on a state or regional basis, depending on the ecology and environment in which it circulates. In the northeast, univoltine species maintain the virus from year to year, emerging in the early summer to start transmission cycles into multivoltine mosquitoes and mammals. In contrast, multivoltine species seem to be responsible for JCV circulation in North Dakota. This may also hold true for other western states. These results suggest that JCV maintenance may differ based on geographical region, potentially influenced by other ecological and environmental factors.

#### Assessing JCV Genetic Diversity

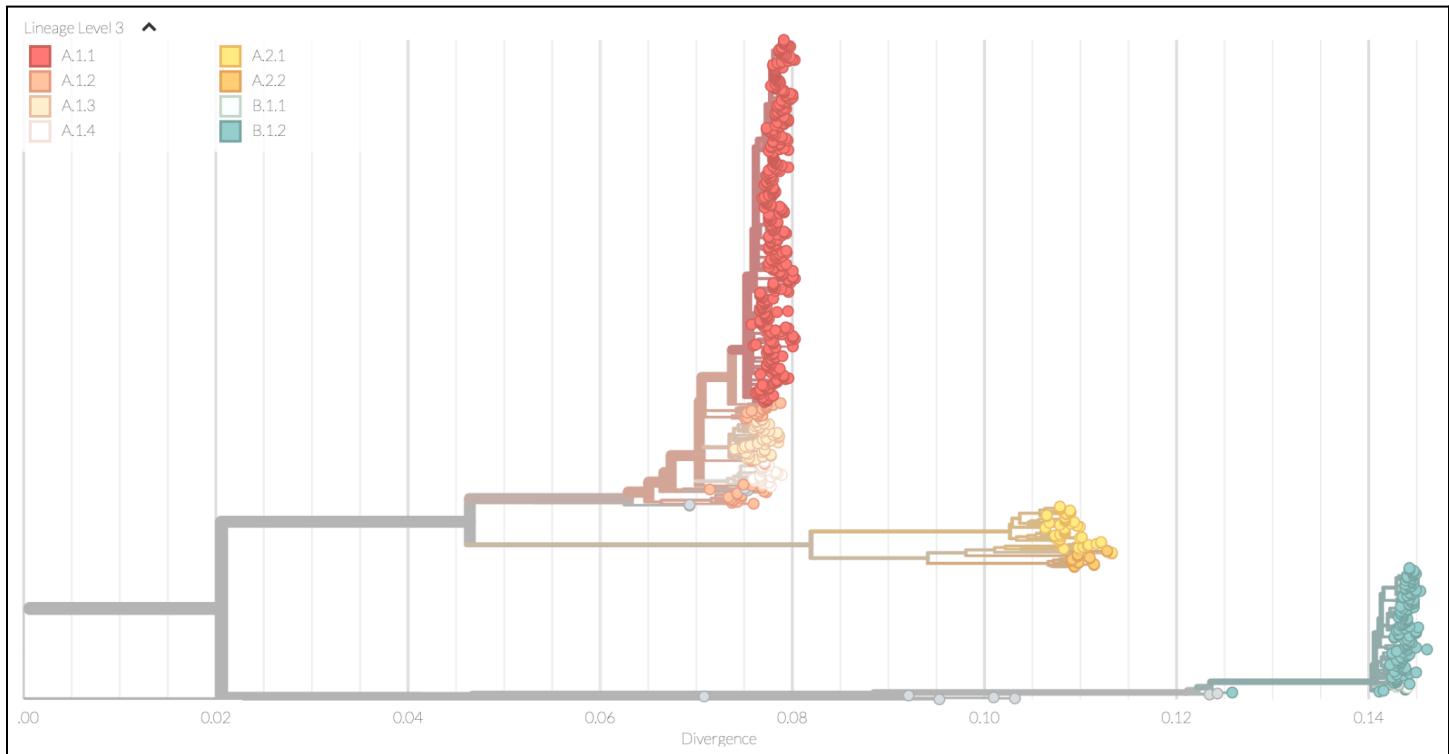
Creating a lineage designation system for JCV allows us to further describe the existing viral diversity in the United States. Such analysis facilitates spatial and temporal analysis of JCV across Connecticut, New York, and North Dakota.

The designation system I created reaffirms the major A and B lineages proposed by CAES. Within these lineages is further genetic diversity. These are A.1.1, A.1.2, A.1.3, A.1.4, A.2.1,

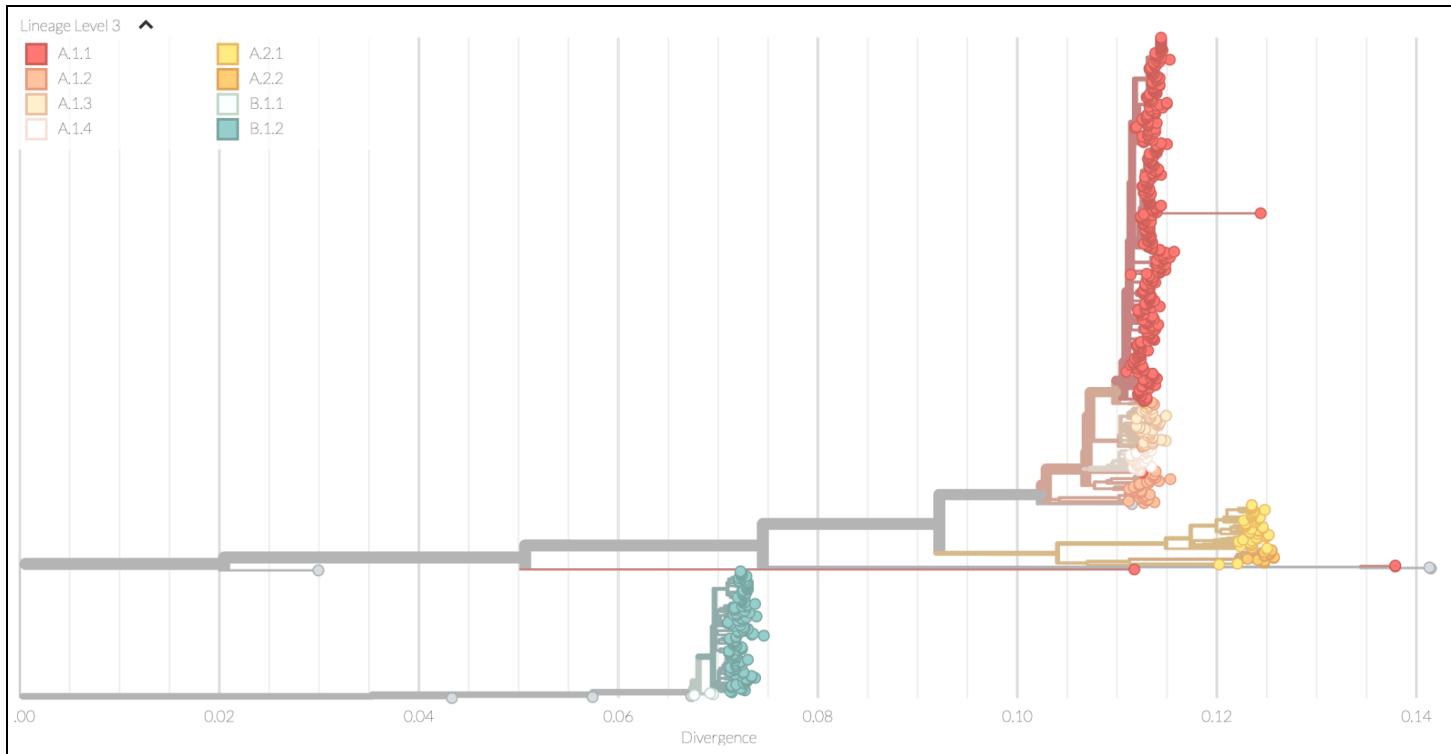
#### A. Small segment



#### B. Medium segment



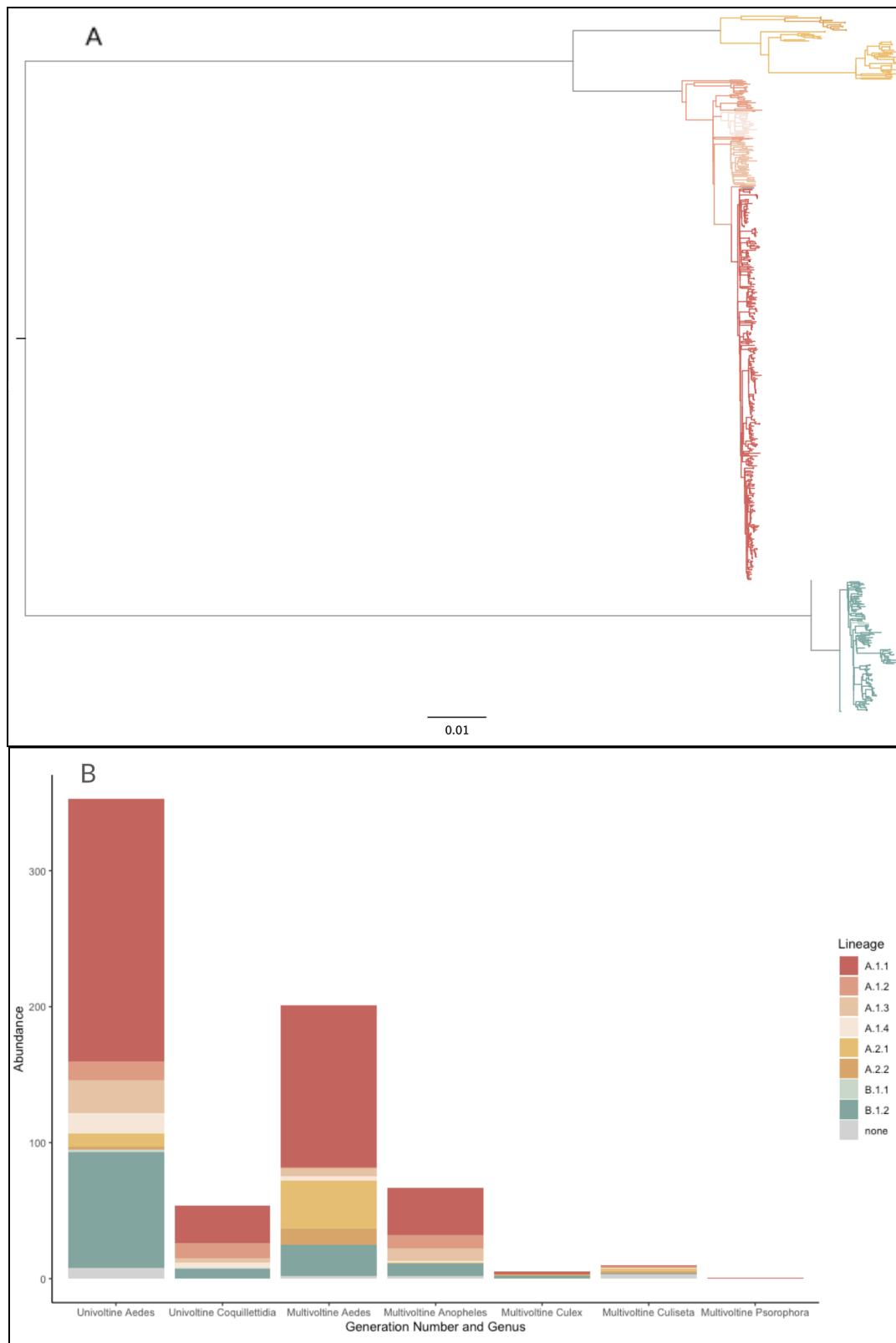
### C. Large Segment



**Figure 3.** Detection of JCV lineages in the northern United States, 1966-2023. (A) Maximum likelihood tree of the JCV small segment, colored by lineage. (B) Maximum likelihood tree of the JCV medium segment, colored by lineage. (C) Maximum likelihood tree of the JCV large segment, colored by lineage.

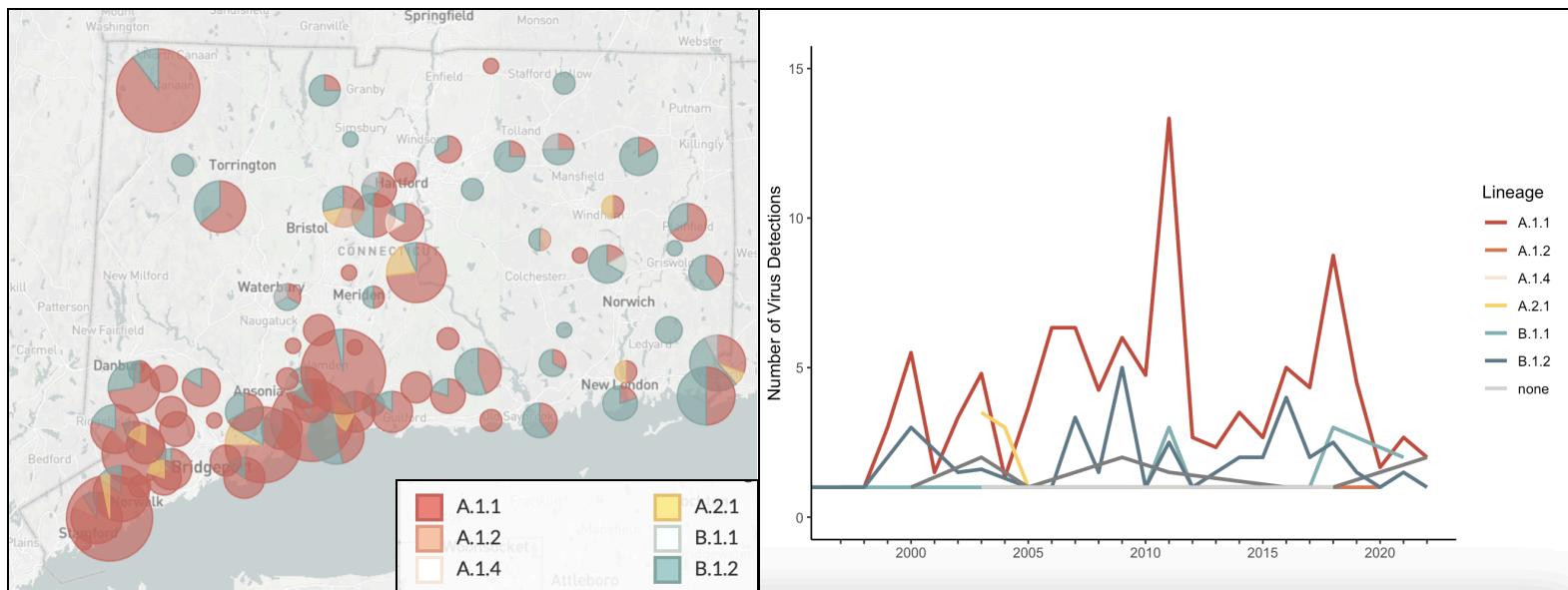
A.2.2, B.1.1, B.1.2 (Fig. 3A-C). The vast majority of samples (379) fall into the A.1.1 lineage, followed by B.1.2 (127), A.2.1 (48), A.1.3 (43), A.1.2 (37), A.1.4 (24), A.2.2 (16), and B.1.1 (3). These designations are broadly conserved across all three JCV segments, with limited disagreement (Fig. 3A-C). The phylogeny, and therefore lineages, of the small segment are the most divergent of the three (Fig. 3A). Here, the B lineage is divided into 3 groups as opposed to 2, as the B.2.1 cluster represents the previously observed B2 “sister clade.”<sup>9</sup> The A.2 cluster is also divided into 3 sub-lineages, as opposed to 2 (Fig. 3A). These slight differences in topology do not impact the final lineage calling system. However, they do suggest that there is increased diversity in the small segment of the JCV genome, as opposed to medium and large.

In order to understand the total diversity of the genome, all three segments were concatenated together for final phylogenetic analysis, excluding reassortants or samples for which lineage could not consistently be determined. The resulting tree demonstrates that the A and B lineages are highly divergent from one another (Fig. 4A). This affirms past suggestions that the lineages have been evolving independently from one another for an extended period of time.<sup>9</sup> Estimation of whole-genome evolutionary rates using Clocklor2 found that the two lineages likely evolve at different rates (Supp. Fig. 4A-B). The program estimated that the A lineage evolved at an approximate rate of  $9.033 \times 10^{-6}$  nucleotide substitutions per nucleotide site per year, with the B

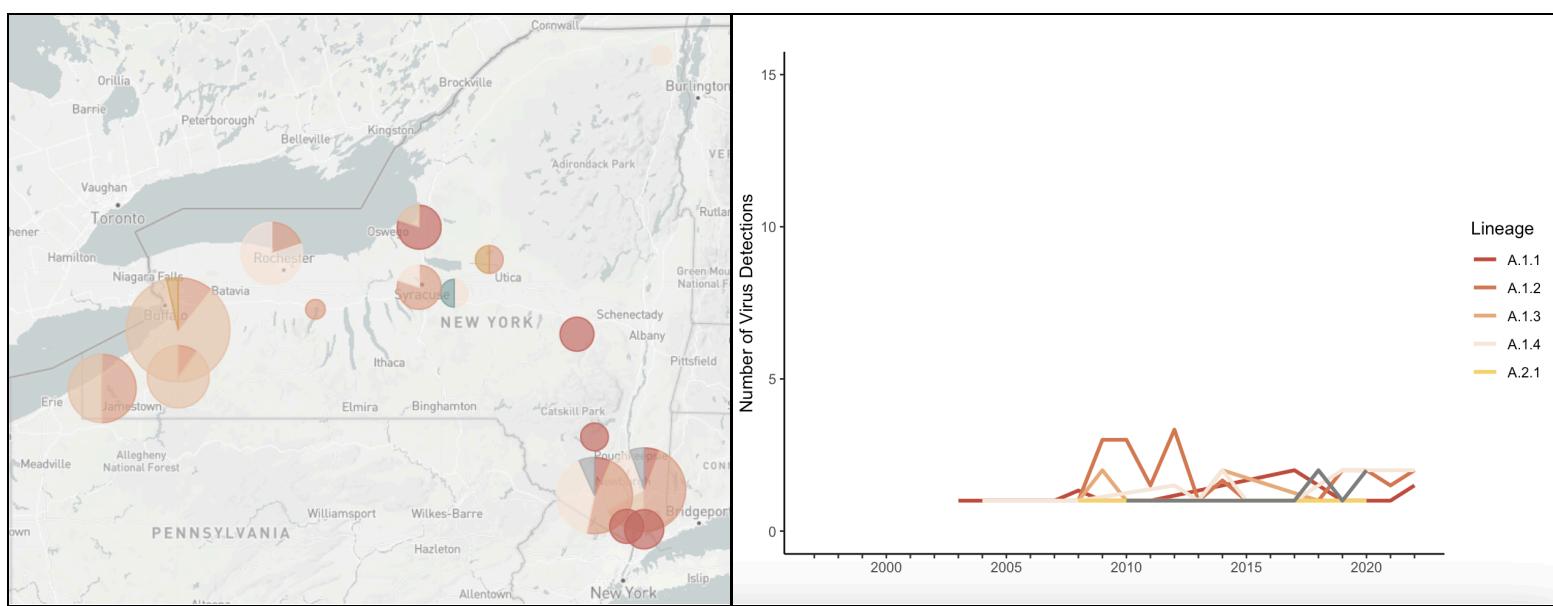


**Figure 4.** Total viral diversity of JCV observed in concatenated genomes, 1997-2023. (A) Maximum likelihood tree of JCV concatenated sequences, colored by lineage. (B) Bar plot of isolations of JCV lineages in generation-groups of mosquitoes.

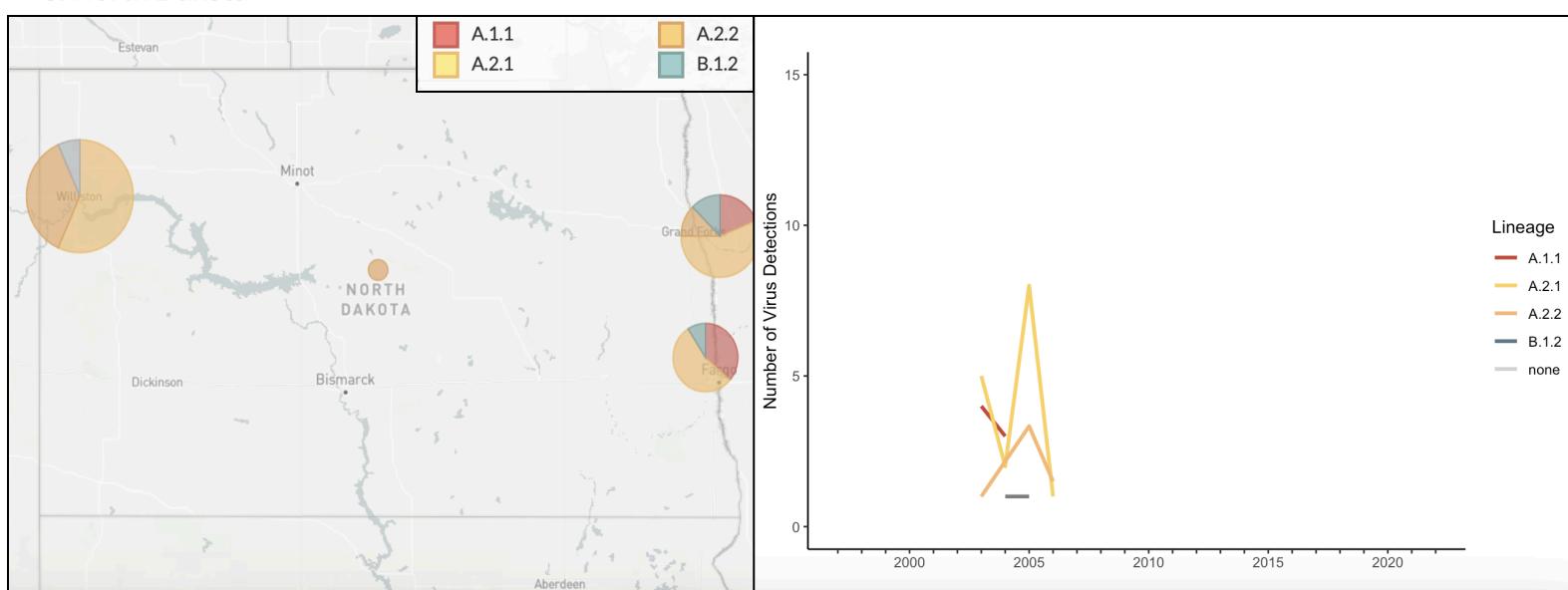
### A. Connecticut



### B. New York



### C. North Dakota



**Figure 5.** Spatial and temporal distribution of JCV lineages across Connecticut, New York, and North Dakota. (A) Left: Map of JCV isolations and number of annual virus detections in Connecticut, colored by lineage. Right: Isolation of JCV lineages in Connecticut per year, 1997-2022. (B) Left: Map of JCV isolations across cities in New York, colored by lineage. Right: Isolation of JCV lineages in New York per year, 2003-2022. (C) Left: Map of JCV isolations across cities in North Dakota, colored by lineage. Right: Isolation of JCV lineages in North Dakota per year, 2003-2006.

lineage at  $2.480 \times 10^{-5}$  (Supp. Table 2). This is in line with the predicted “slow” evolutionary rate.<sup>4,10</sup> However, these estimations are associated with very small  $R^2$  values, suggesting poor temporal signal in the data that should be further explored.

The distribution of JCV lineages can also be understood in reference to its mosquito vectors. There is no clear correlation between the generation-genus group of a mosquito and the lineages it carries (Fig. 4B). All groups appear capable of carrying both major lineages A and B (Fig. 4B). Most groups predominantly carry A.1.1 (Fig. 4B), as this is the most widely circulating lineage of the virus in the U.S. Lineage A.2 is most commonly isolated in *Aedes* mosquitoes, with some preference for multivoltine species (Fig. 4B). Notably, this lineage is geographically based in North Dakota (Fig. 5C), which may impact its distribution throughout mosquito groups. Because of this geographic clustering pattern, distribution of JCV should be considered spatially. The proportion of JCV lineage isolations per year can also be assessed to better understand long-term transmission patterns within states.

The predominant lineage isolated in Connecticut is A.1.1, followed by B.1.2 (Fig. 5A). Most samples collected in the state fall into one of these two groups, with A.1.1 isolated more frequently in every year except 2009 (Fig. 5A). In general, it seems these two lineages stay in relative proportion with one another from year to year (Fig. 5A). From a geographic perspective, though, there is an observable difference in the lineages collected across the state. As proposed by CAES,<sup>3,9</sup> there is a higher proportion of lineage A collected in western Connecticut, compared to the high prevalence of lineage B in the state’s eastern region (Fig. 5A). The primary ecological barrier separating these two regions is the Connecticut River,<sup>3,9</sup> which may reflect the separation of white-tailed deer populations on either side (Fig. 5A). Determination of other ecological factors influencing this stratified lineage distribution will require further analysis. Spread of other JCV lineages in the state is relatively limited. Lineage A.2.1, which is most common in North Dakota, was isolated in Connecticut from 2003-2005, before dropping from circulation entirely. It is unclear how this western strain appeared and disappeared in the northeast so quickly. Similarly, lineage B.1.1 emerged in Connecticut in 2007, but was not detected in the state after 2011 (Fig. 5A). The mechanisms which underlie this rapid cycle of emergence and extinction are unknown. In contrast, lineage A.1.2 first appeared in Connecticut in 2016, consistently detected and increasing in prevalence since (Fig. 5A). Future epidemiological analysis of JCV in the northeast should undoubtedly explore the possible drivers of these unique transmission patterns from a spatiotemporal perspective.

New York has a strikingly different lineage distribution than its neighboring state. While molecular detection methods likely skew the identification of lineage B in the state,<sup>16</sup> the circulating lineage A sequences are markedly different. Here, lineages A.1.3 and A.1.2 are the most commonly isolated, with frequent peaks in lineage A.1.4 (Fig. 5B). There was also limited isolation of lineage A.2.2 from 2005-2006 (Fig. 5B). Lineage A.1.1 is found primarily in the east,

the region of the state, closer to Connecticut (Fig. 5B). JCV in western New York is almost entirely composed of A.1.2, A.1.3, and A.1.4 lineages (Fig. 5B). These lineages are rarely found outside of New York, only recently spreading to Connecticut (Fig. 5A-B). Once again, this suggests some form of transmission restriction, with ecological factors preventing spread of lineages from region to region. It is unclear which aspects of JCV ecology are most important in New York state, especially in the absence of major geographical barriers.

Of the states included in this project, North Dakota has the most homogeneous distribution of JCV lineages. The vast majority of isolations belong to the A.2.1 and A.2.2 lineage (Fig. 5C). However, there is also observed circulation of A.1.1 and B.1.2, the dominant lineages from Connecticut, primarily in the eastern region of North Dakota (Fig. 5C). Critically, these sequences are restricted to a short window of time (2003-2006), limiting our understanding of long-term lineage trends in the state. It is very difficult to infer the mechanisms by which lineages co-circulate in highly separated regions. Data from neighboring states may be important to further inform the genetic diversity observed in the midwest and western U.S.

#### Measuring Reassortment of JCV

Given the unique evolutionary patterns of the sequences, it is essential to understand to what degree has segment reassortment impacted the evolution of JCV. Reassortment events were visually assessed by cross-referencing the segments by lineage in phylogenetic tanglegrams (Fig. 6A-D).

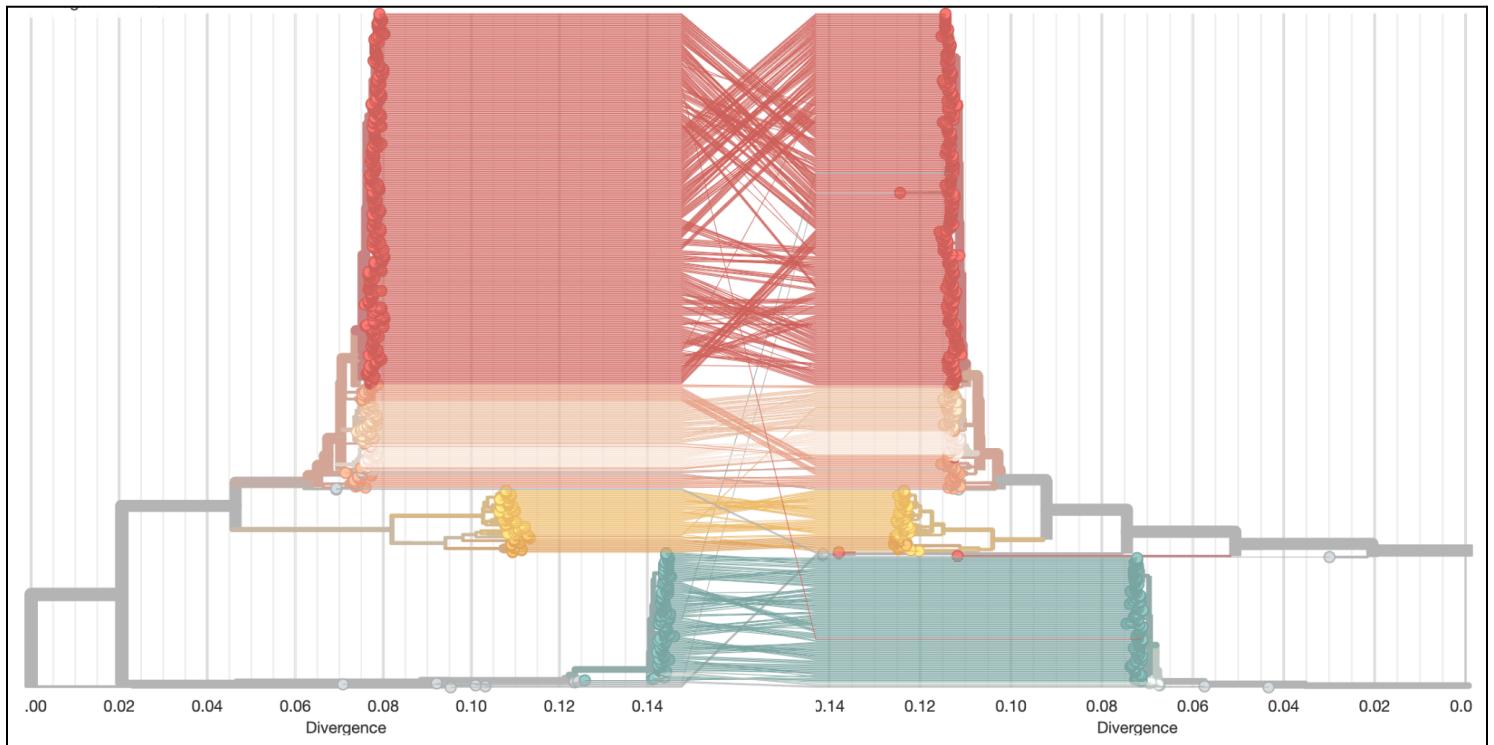
**Table 1.** Identified reassortant JCV samples across genomic data from Connecticut, New York, Massachusetts, and North Dakota, from 1966-2023.

Sample Name	Collection Date	State	Mosquito Species	Small Segment Lineage	Medium Segment Lineage	Large Segment Lineage
CT Yale-JC0251 2006	2006-07-11	Connecticut	<i>Aedes vexans</i>	A.1.1	B-related	A.1.1
NY JC13 2009	2009-06-01	New York	<i>Aedes stimulans</i>	A.1.3	A.2.2	A.1.3
CT Yale-JC0343 2009	2009-07-01	Connecticut	<i>Aedes aurifer</i>	A.1.1	A.1.1	B.1.2
CT Yale-JC0344 2009	2009-07-01	Connecticut	<i>Aedes cantator</i>	B.1.1	B.1.2	A.1.1
CT Yale-JC0529 2017	2017-06-07	Connecticut	<i>Aedes abserratus</i>	A.1.1	B-related	B-related

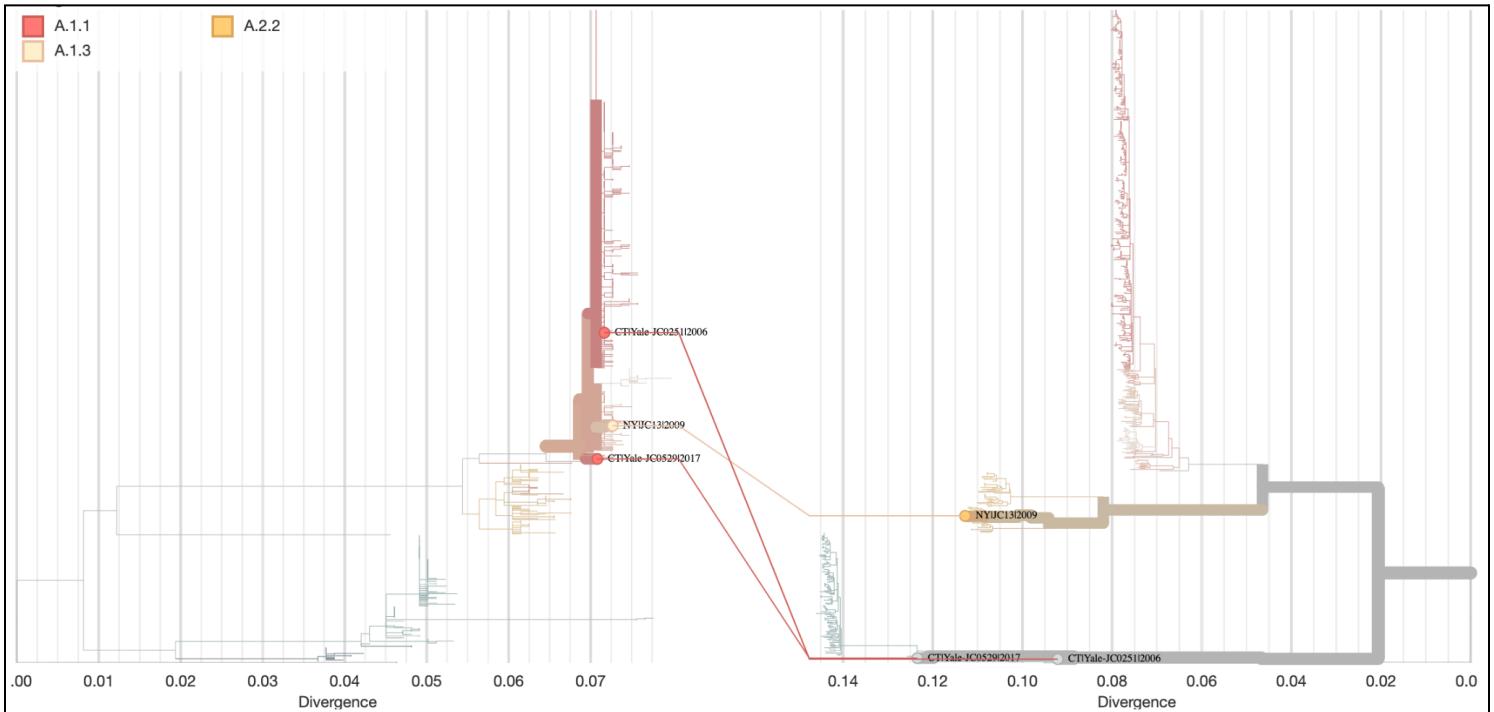
Reassortment events were observed across all three segments of JCV (Fig. 6B-D), leading to the identification of 5 total reassortant samples (Table 1). These account for 0.73% of the total genomic dataset, with an estimated 0.18 reassortment events occurring per year. 3 of 5 of the

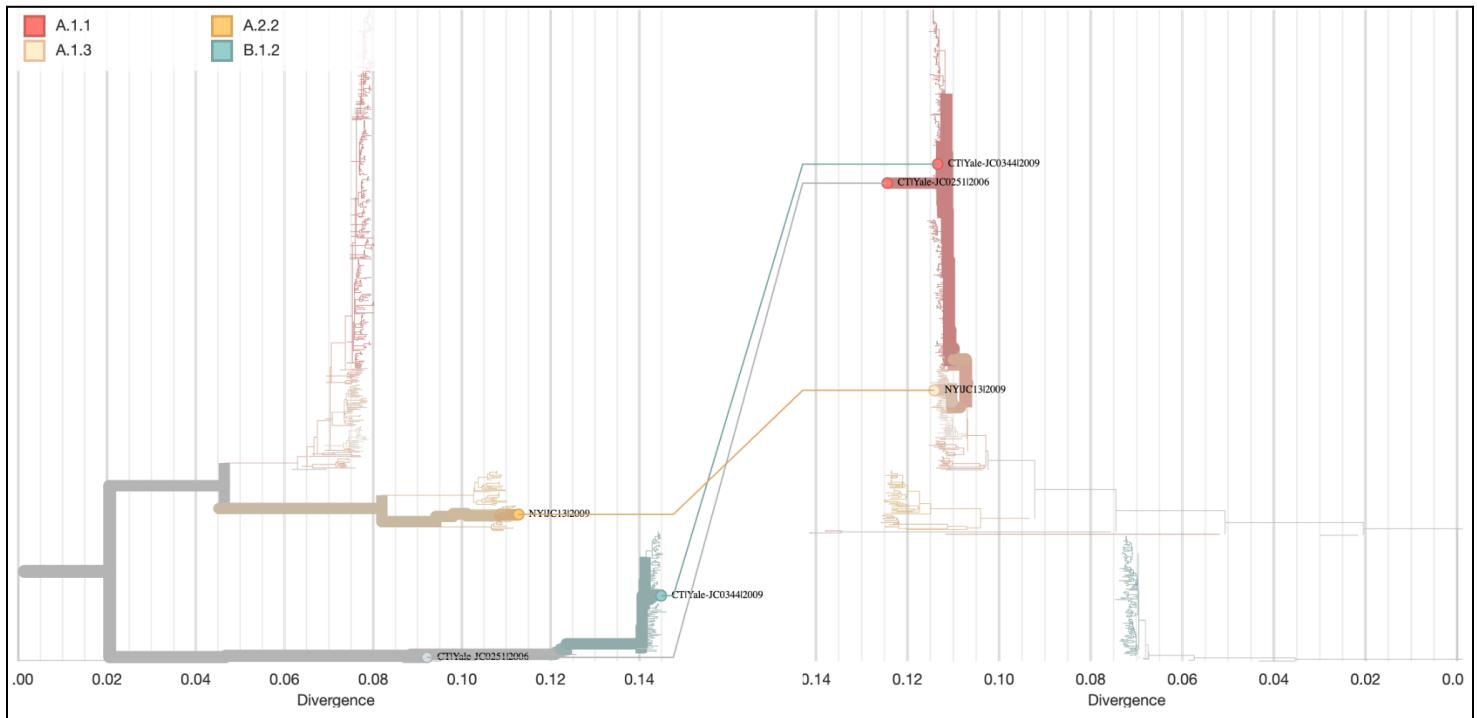
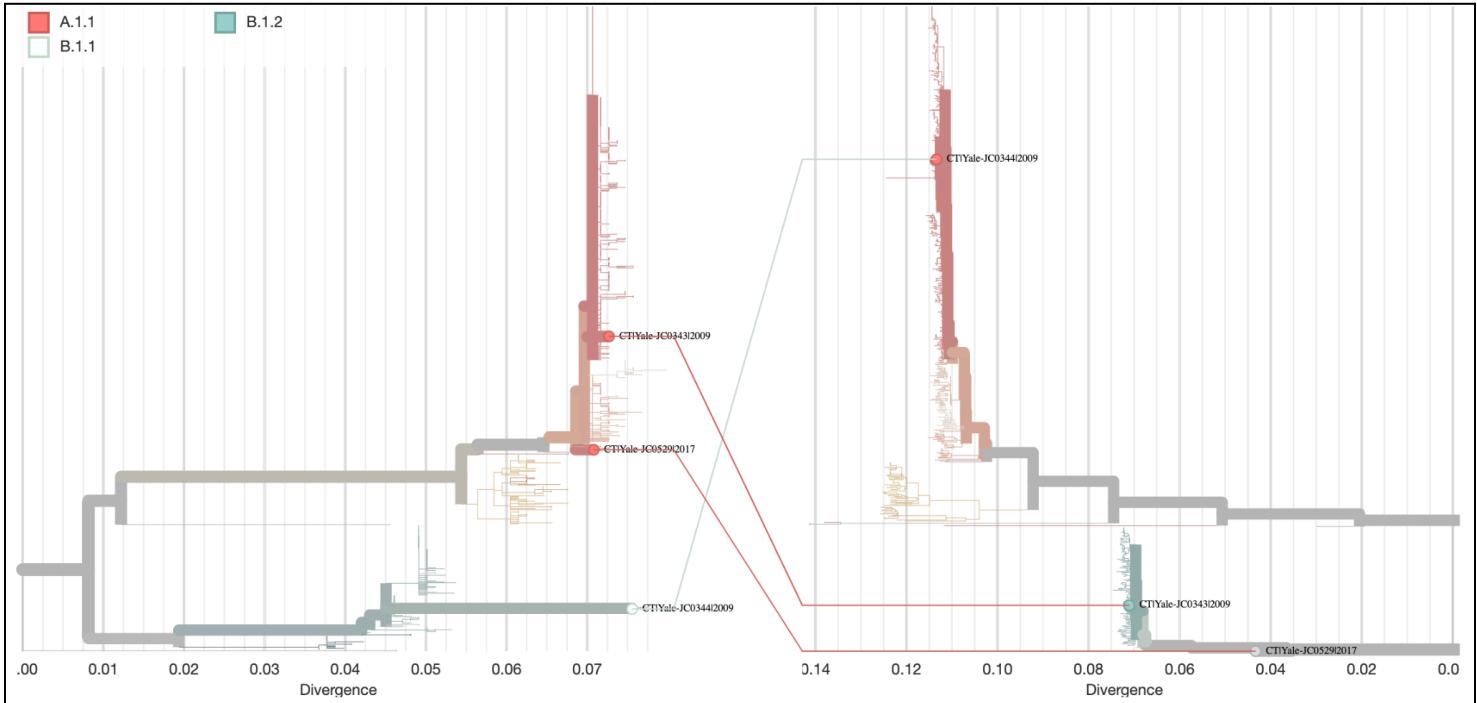
identified reassortments were in 2009 (Table 1), an observation for which there is no clear explanation. Reassortants were collected from univoltine and multivoltine mosquitoes alike, with no clear preference for either group (Table 1). This analysis also records reassortment

A. Full Tanglegram (Medium vs. Large Segment)



B. Small vs. Medium Segment



**C. Medium vs. Large Segment****D. Small vs. Large Segment**

**Figure 6.** Identified JCV reassortment events in the northern United States, 1966-2023. (A) Tanglegram of the medium (left) and large (right) segments of JCV, colored by lineage. (B) Tanglegram of the small (left) and medium (right) segments of JCV, colored by the generation group of the sample's mosquito pool, with reassortant samples highlighted. (C) Tanglegram of the medium (left) and

large (right) segments of JCV, colored by the generation group of the sample's mosquito pool, with reassortant samples highlighted. (D) Tanglegram of the small (left) and large (right) segments of JCV, colored by the generation group of the sample's mosquito pool, with reassortant samples highlighted

occurring between major lineages of JCV for the first time, as observed in CT|Yale-JC0251|2006, CT|Yale-JC0343|2009, CT|Yale-JC0344|2009, and CT|Yale-JC0529|2017 (Fig. 6B-D). This phenomenon demonstrates that highly divergent lineages of the virus are still capable of exchanging segments during coinfection, a phenomenon which was previously in question.

As suggested by previous studies, JCV evolves primarily through genetic drift, complicated by random but rare occurrences of genetic shift.<sup>3,9</sup> Further phylogenetic and statistical analysis is necessary to determine validated reassortment rates and distinguish reassortment rates for each segment separately.

## CONCLUSIONS

Jamestown Canyon virus (JCV) is an understudied arbovirus that requires further examination from a public health perspective. By sequencing 689 JCV genomes across Connecticut, New York, North Dakota, and Massachusetts and analyzing it alongside comprehensive mosquito surveillance data, I was able to describe the virus's maintenance, evolution, and spread. This analysis suggests that JCV is primarily maintained in univoltine mosquito populations, which serve as "early season amplification vectors,"<sup>3</sup> emerging from vernal pools while infected with the virus. The results in an early peak in their infection rates and vector index, which may reflect increased risk for human infections. While multivoltine mosquitoes also participate in the circulation of JCV, they are less critical for its year-to-year transmission. Of the observed species, univoltine *Aedes* mosquitoes are critical to the maintenance and spread of JCV in the northeast. As previously suggested, *Aedes canadensis* is of particular interest, given its early emergence and human biting preference.<sup>16</sup> With this in mind, public health interventions should focus on control in the late spring and early summer to best mitigate human JCV infection risk.

JCV itself can be further described from a molecular perspective, characterizing the diversity and evolution of its segmented genome. Isolated viruses fall into 2 major lineages (A and B), which can then be divided into 8 total sublineages (A.1.1, A.1.2, A.1.3, A.1.4, B.1.1, B.1.2). The highly differentiated major lineages likely diverged and have been evolving independently from each other for a significant period of time. While there is little evidence of temporal clustering in this diversity, the lineages are geographically separated. East-west distinctions in lineage circulation were observed in both Connecticut and New York. However, there is also evidence of spread between states and, potentially, across long distances. Ecological factors beyond the scope of this project, such as white-tailed deer distribution, land use, and other variables, may be responsible for these observed phenomena. Despite geographic distribution of lineages, phylogenetic analysis identified 5 cross-lineage reassortment events, which represented 0.73%

of the total dataset. This suggests that reassortment is a rare evolutionary event for JCV, punctuating a regular pattern of genetic drift.

This project has several limitations. The first is in regards to data availability. Given the intensity of JCV infections in the midwest, sequences of positive mosquito pools from midwest states may be critical to understanding national circulation of the virus. The geographical gap in available genomic data for this project, spanning from North Dakota to New York, is a significant barrier. Ideally, sequences from states such as Minnesota, Michigan, and Pennsylvania should be incorporated to understand the total viral diversity, in addition to the relationship between midwest and northeastern JCV. Differences between mosquito surveillance programs across states may also have impacted this project's conclusions. Connecticut sequences are overrepresented in this dataset, a focus which may have biased our understanding of existing viral diversity and the process of lineage designation. Molecular detection methods also vary across states, which may affect the distribution of lineages observed in this project's data. Specifically, the use of RT-PCR methods that do not detect JCV lineage B may lead to the underrecognition of lineage B positive pools, potentially skewing the results observed in states such as New York.

The reassortment analysis of this project was also conducted through a purely visual methodology. While this allows for the identification of significant reassortment events, it is not supported by statistical assessments and does not allow for the specific calculation of reassortment rates. Beyond visual analysis of the tanglegram, there is no clear definition of what constitutes a "reassortment event" in the JCV genome. While cross-lineage reassortment is easy to identify, exchange between more closely related viruses may be more difficult to detect in the absence of further analysis. The sequences for which lineage could not be assigned, or which had several different lineage assignments across segments, suggest that reassortment is occurring more frequently than estimated using the tanglegram alone. Further analysis is required to make more significant conclusions about JCV's reassortment history.

Another significant limitation of this project was the lack of temporal resolution in the phylogenetic analysis. While I was able to roughly estimate the overall evolutionary rates of both major lineages, I was not able to incorporate time-scaling into my trees. When applied to maximum likelihood trees on Nextstrain, the time-scale was not able to accurately estimate time of JCV emergence or time to most recent common ancestor (TMRCA). This missing time variable limits this analysis, making it difficult to infer temporal relationships between samples and timing of JCV emergence in different states.

Many of this project's limitations, though, may be addressed and resolved with future analysis using more complex methods. This process would begin with resolving issues of time-scaling in the phylogenetic trees. BEAST, or Bayesian Evolutionary Analysis Sampling Trees, utilizes Markov Chain Monte Carlo (MCMC) to construct phylogenies and is, critically, capable of using a "relaxed clock" as a part of its time-scaling.<sup>37</sup> Rather than a single likely tree, the program infers evolutionary relationships between sequences by constructing multiple phylogenies and weighting them based on their corresponding posterior probabilities.<sup>37</sup> This method is the best

candidate for creating time-scaled phylogenetic trees for the segmented or concatenated data sets. BEAST can also integrate spatiotemporal data to yield reconstructions of a virus's evolution over space and time,<sup>37</sup> which I hope to incorporate into future work.

This project can also be expanded to answer other questions about JCV's evolutionary history and distribution in the United States. While the geographic area covered by our current dataset is not large, the resolution may be granular enough to complete phylodynamic analysis. Also utilizing BEAST, this methodology would allow for the incorporation of spatiotemporal data to estimate the emergence and spread<sup>39</sup> of JCV in the northeast. This could further elucidate the mechanisms by which the virus circulates, identifying areas of high risk<sup>39</sup> for the purpose of future control efforts. Regression analysis can also be used in the future to assess ecological determinants of spread for each lineage. As identified in this and other studies, a variety of ecological factors, including the distribution of white-tailed deer, may contribute to spatial isolation of the viral lineages. Deer distribution, in addition to other factors such as elevation, land use, and anthropogenic measures,<sup>14</sup> can be incorporated into regression models as predictors, with JCV lineage serving as the response variable.

It is also critical that further study of the JCV genome is completed to assess conserved amino acid mutations and their possible impacts on transmissibility and virulence. This can and should be assessed in tandem with more rigorous examination of segment reassortment. TreeKnit is a method developed to analyze influenza reassortment, generating maximum compatibility clades (MCCs) and ancestral reassortment graphs (ARGs) as a program output.<sup>38</sup> This method assesses topological disparities across phylogenetic trees with matching tips and assigns compatible clusters iteratively in a greedy process.<sup>38</sup> Critically, TreeKnit is capable of applying this method to more than 2 segments,<sup>38</sup> which would allow for tripartite analysis of the JCV genome. The use of this method in future JCV analysis will allow for a quantitative, statistically validated assessment of segment reassortment.

Only 25% of known orthobunyaviruses have been sequenced.<sup>1</sup> Their total impact on humans is still not completely understood.<sup>2,3</sup> As the threat of such re-emerging viruses rises, we must also increase our efforts to understand and address them. This project, and its proposed future directions, demonstrate the power of coupling genomic and ecological data in an epidemiological context. With this information, and a variety of analytical techniques, we have the ability to untangle complex viral transmission cycles. This has the potential to increase our understanding of pathogen spread and identify areas for future public health intervention.

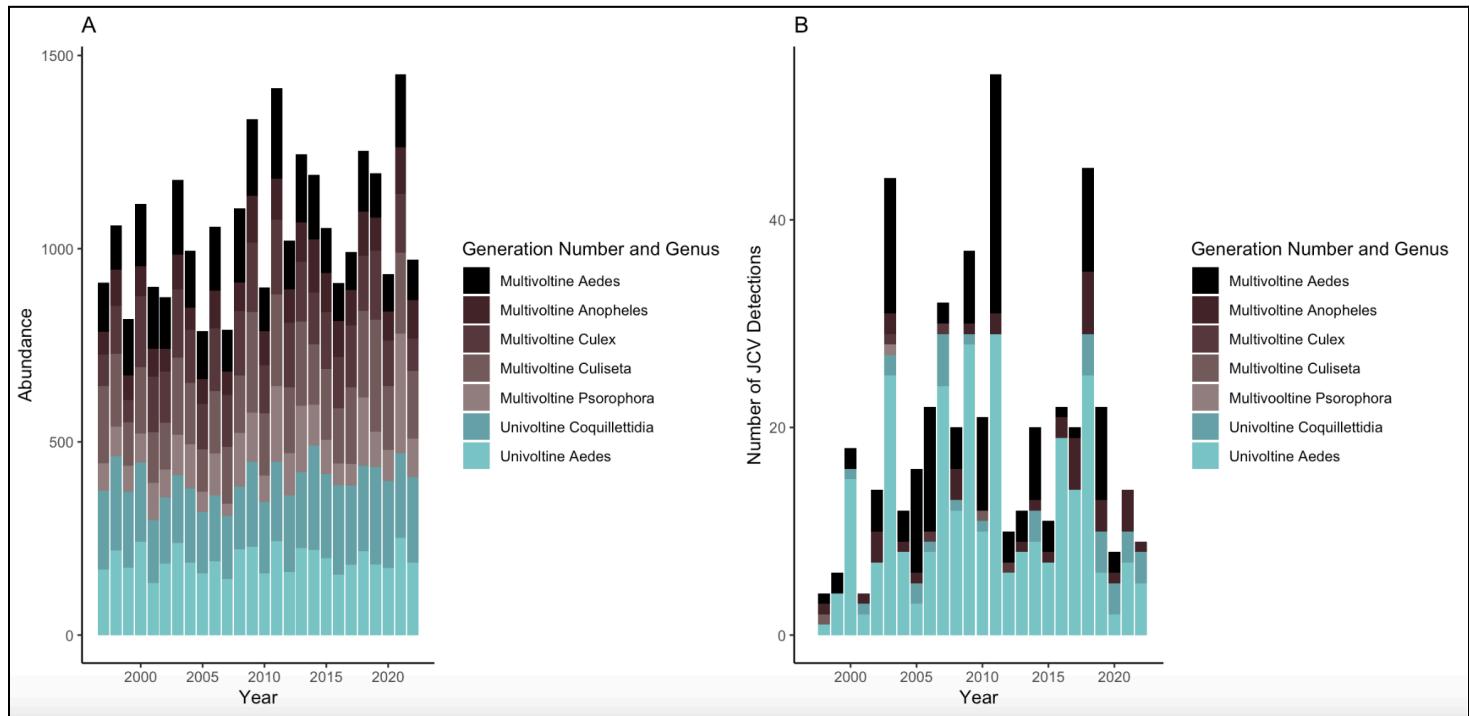
## REFERENCES

1. R. M. Elliot. (2014). Orthobunyaviruses: recent genetic and structural insights. *Nat. Rev. Microbiol.* 12, 673-685.
2. J. D. Poggi, C. Conery, A. Mathewson, D. Bolton, et al. (2023). Jamestown Canyon Virus (Bunyavirales: Peribunyaviridae) vector ecology in a focus of human transmission in New Hampshire, USA. *J. Med. Entomol.* 60(4), 778-788.
3. J. J. Shepard, P. M. Armstrong. (2023). Jamestown Canyon virus comes into view: understanding the threat from an underrecognized arbovirus. *J. Med. Entomol.*
4. C. M. Kinsella, M. L. Paras, S. Smole, S. Mehta, et al. (2020). Jamestown Canyon virus in Massachusetts: clinical case series and vector screening. *Emerg. Microbes Infect.* 9, 903-912.
5. Centers for Disease Control and Prevention. ArboNet. Jamestown Canyon Virus. Historic Data (2011-2022). [cited 2024 Mar 18].
6. Centers for Disease Control and Prevention. ArboNet. Jamestown Canyon Virus. Current Year Data (2023). [cited 2024 Mar 18].
7. M. Lau, H. L. C. Dutra, M. J. Jones, B. P. McNulty, et al. (2023). Jamestown Canyon Virus is transmissible by *Aedes aegypti* and is only moderately blocked by *Wolbachia* co-infection. *PLoS Negl. Trop. Dis.* 17(9).
8. C. M. Gill, J. D. Beckham, A. L. Piquet, K. L. Tyler, et al. (2019). Five emerging neuroinvasive arboviral diseases: Cache Valley, Eastern Equine Encephalitis, Jamestown Canyon, Powassan, and Usutu. *Semin. Neurol.* 39, 419-427.
9. P. M. Armstrong, T. G. Andreadis. (2007). Genetic relationships of Jamestown Canyon virus strains infecting mosquitoes collected in Connecticut. *Am. J. Trop. Med. Hyg.*, 77(6), 1157-1162.
10. A. B. Evans, C. W. Winkler, K. E. Peterson. (2019). Differences in neuropathogenesis of encephalitic California serogroup viruses. *Emerg. Infect. Dis.* 25(4), 728-738.
11. T. Briese, C. H. Calisher, & S. Higgs. (2013). Viruses of the family *Bunyaviridae*: Are all available isolates reassortants? *Virol. J.* 446, 207-216.
12. M. M. Ogg & J. L. Patterson. (2007). RNA binding domain of Jamestown Canyon virus S segment RNAs. *J. Virol.* 81(24)
13. H. R. Hughes, R. S. Lanciotti, C. D. Blair, A. J. Lambert. (2017). Full genomic characterization of California serogroup viruses, genus *Orthobunyavirus*, family *Peribunyaviridae* including phylogenetic relationships. *Virol.* 512, 201-210.
14. K. M. Hollis-Etter, R. A. Montgomery, D. R. Etter, C. L. Anchor, J. E. Chelvig, R. E. Warner, P. R. Grimstad, D. D. Lovin, M. & S. Godsey Jr. (2019). Environmental conditions for Jamestown Canyon virus correlated with population-level resource selection by white-tailed deer in a suburban landscape. *PLoS One.* 14(10).
15. C. A. Schneider, J. M. Leung, P. C. Valenzuela-Leon, N. A. Golviznina, E. A. Toso, D. Bosnakovski, M. Kyba, E. Calvo, & K. E. Peterson. (2024). Skin muscle is the initial site of viral replication for arboviral bunyavirus infection. *Nat. Commun.*, 15.
16. K. A. Ngo, J. G. Maffei, C. A. Koetzner, S. D. Zink, A. F. Payne, P. B. Backenson, J. L. White, A. P. Dupuis II, L. D. Kramer, & A. T. Ciota. (2023). Surveillance and genetic

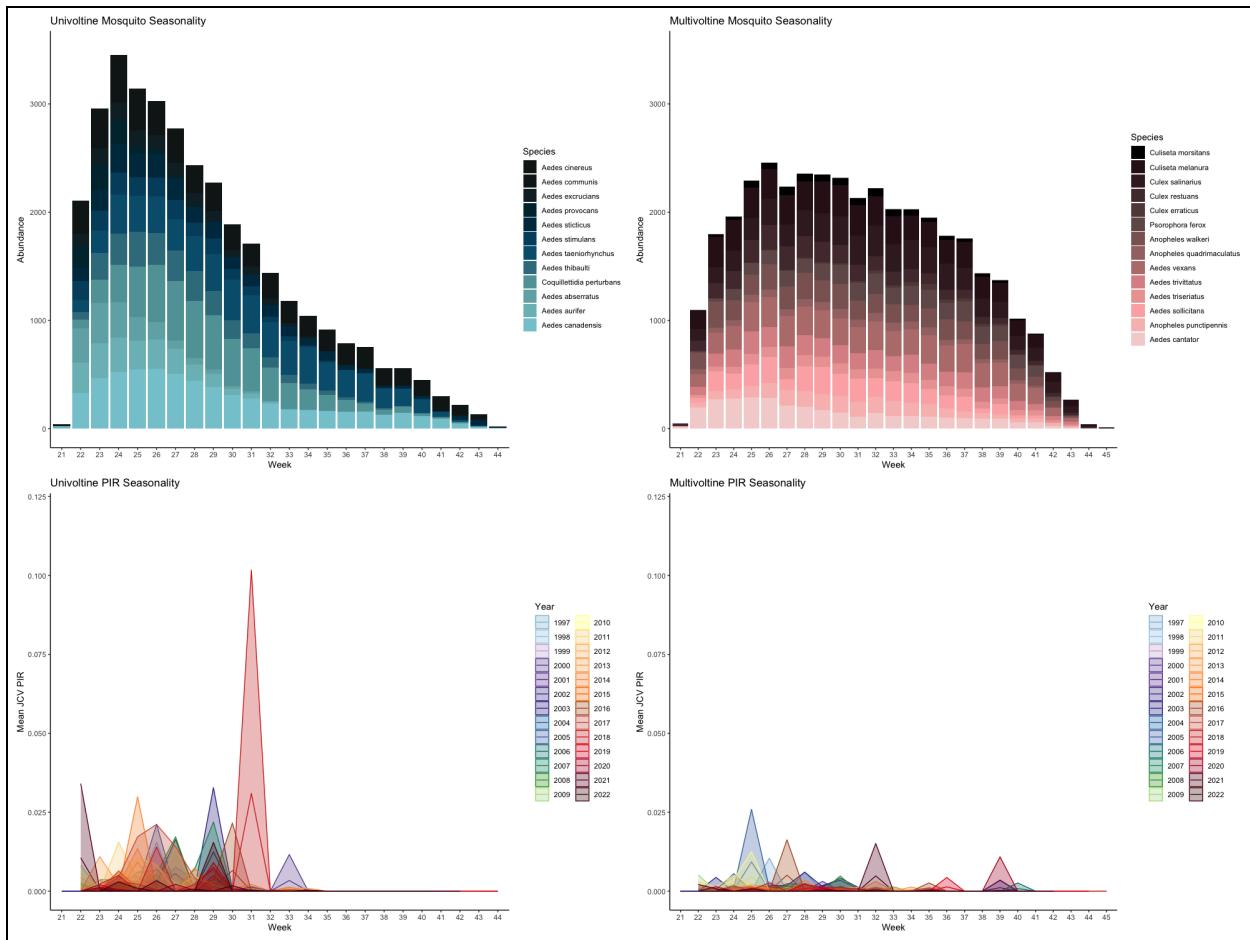
- analysis of Jamestown Canyon virus in New York state: 2001-2022. *Am. J. Trop. Med. Hyg.* 109(6).
17. J. F. Anderson, A. J. Main, P. M. Armstrong, T. G. Andreadis, F. J. Ferrandino. Arboviruses in North Dakota, 2003-2006. *Am J Trop Med Hyg.* 92(2), 377-393.
  18. J. Quick, N. Grubaugh, S. Pullan, I. M. Claro, A. D. Smith, K. Gangavarapu, G. Oliviera, R. Robles-Sikisaka, T. F. Rogers, N. A. Beutler, D. R. Burton, L. L. Lewis-Ximenez, J. Goes de Jesus, M. Giovanetti, S. C. Hill, A. Black, T. Bedford, M. W. Carroll, M. Nunes, L. C. Alcantara Jr., E. C. Sabino, S. A. Baylis, N. R. Faria, M. Loose, J. T. Simpson, O. G. Pybus, K. G. Andersen, N. J. Loman. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.*, 12, 1261-1276.
  19. C. B. F Vogels, V. Hill, M. I. Breban, C. Chaguza, L. M. Paul, A. Sodeinde, E. Taylor-Salmon, I. M. Ott, M. E. Petrone, D. Dijk, M. Jonges, M. R. A. Welkers, T. Locksmith, Y. Dong, N. Tarigopula, O. Tekin, S. Schmedes, S. Bunch, N. Cano, R. Jaber, C. Panzera, I. Stryker, J. Vergara, R. Zimler, E. Kopp, L. Heberlein, A. M. Morrison, S. F. Michael, N. D. Grubaugh. (2023). DengueSeq: A pan-serotype whole genome amplicon sequencing protocol for dengue virus. *medRxiv preprint*.
  20. R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
  21. T. W. Rinker, D. Kurkiewicz. (2018). *pacman: Package Management for R*. version 0.5.0, <http://github.com/trinker/pacman>.
  22. H. Wickham, J. Hester, J. Bryan. (2024). *readr: Read Rectangular Text Data*. R package version 2.1.5, <https://github.com/tidyverse/readr>, <https://readr.tidyverse.org>.
  23. H. Wickham, J. Bryan. (2023). *readxl: Read Excel Files*. <https://readxl.tidyverse.org>, <https://github.com/tidyverse/readxl>.
  24. H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T.L. Pedersen, E. Miller, S.M. Bache, K. Müller, J. Ooms, D. Robinson, D.P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani. (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686.
  25. S. Firke. (2024). *janitor: Simple Tools for Examining and Cleaning Dirty Data*. R package version 2.2.0.9000, [https://sfirke.github.io/janitor/](https://sfirke.github.io/janitor), <https://github.com/sfirke/janitor>.
  26. G. Grolemund, H. Wickham. (2011). "Dates and Times Made Easy with lubridate." *Journal of Statistical Software*, 40(3), 1–25. <https://www.jstatsoft.org/v40/i03/>.
  27. B. J. Biggerstaff. (2009). PooledInfRate, Version 4.0: a Microsoft® Office Excel© Add-In to compute prevalence estimates from pooled samples. *Centers for Disease Control and Prevention, Fort Collins, CO*.
  28. A. Kassambara. (2022). Ggpubr: 'Ggplot2' Based Publication Ready Plots. <https://cran.r-project.org/web/packages/ggpubr/index.html>.
  29. T. Pedersen. (2024). *patchwork: The Composer of Plots*. R package version 1.2.0.9000, <https://github.com/thomasp85/patchwork>, <https://patchwork.data-imaginist.com>.
  30. V. Hill, R. T. Koch, S. M. Bialosuknia, K. Ngo, S. D. Zink, C. A. Koetzner, J. G. Maffei, A. P. Dupuis, P. B. Backenson, J. Oliver, A. B. Bransfield, M. J. Misencik, T. A. Petruff, J. J.

- Shepard, J. L. Warren, M. S. Gill, G. Baele, C. B.F. Vogels, G. Gallagher, P. I. Burns, A. Hentoff, S. Smole, C. Brown, M. Osborne, L. D. Kramer, P. M. Armstrong, A. T. Ciota, N. D. Grubaugh. (2023). Dynamics of eastern equine encephalitis virus during the 2019 outbreak in the Northeast United States, *Curr. Biol.* 33(12), 2515-2527.
31. J. Hadfield, C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R. A. Neher. (2018). Nextstrain: real-time tracking of pathogen evolution, *Bioinformatics*, 34(23), 4121–4123.
  32. A. Rambaut. (2014). FigTree v1. 4.2, a graphical viewer of phylogenetic trees. Available from left angle bracket. <http://tree.bio.ed.ac.uk/software/figtree/right.angle.bracket>.
  33. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Lam. (2017). GGTREE: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.*, 8: 28-36.
  34. G. Tonkin-Hill, J. A. Lees, S. D. Bentley, S. D. W. Frost, J. Corander. (2019). Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.*, 47(11), 5539-5549.
  35. J. McBroome, A. de Bernardi Schnieder, C. Roemer, M. T. Wolfinger, A. S. Hinrichs, A. N. O'Toole, C. Ruis, Y. Turakhia, A. Rambaut, & R. Corbett-Detig. (2024). A framework for automated scalable designation of viral pathogen lineages from genomic data. *Nat. Microbiol.*, 9, 550-560.
  36. L. A. Featherstone, A. Rambaut, S. Duchene, W. Wirth. (2024). Clockor2: Inferring global and local strict molecular clocks using root-to-tip regression. *Syst. Biol.*
  37. M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, A. Rambaut. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10, *Virus Evol.*, 4(1).
  38. P. Barrat-Charlaix, T. G. Vaughan, R. A. Neher. (2022). TreeKnit: Inferring ancestral reassortment graphs of influenza viruses. *PLOS Comput. Biol.* 18(8).
  39. C. B. F. Vogels, D. E. Brackney, A. P. Dupuis II, R. M. Robich, J. R. Fauver, A. F. Brito, S. C. Williams, J. F. Anderson, C. B. Lubelczyk, R. E. Lange, M. A. Prusinski, L. D. Kramer, J. L. Gangloff-Kaufmann, L. B. Goodman, G. Baele, R. P. Smith, P. M. Armstrong, A. T. Ciota, S. Dellicour, N. D. Grubaugh. (2023). Phylogeographic reconstruction of the emergence and spread of Powassan virus in the northeastern United States. *PNAS*, 120(16).

## SUPPLEMENTAL MATERIALS



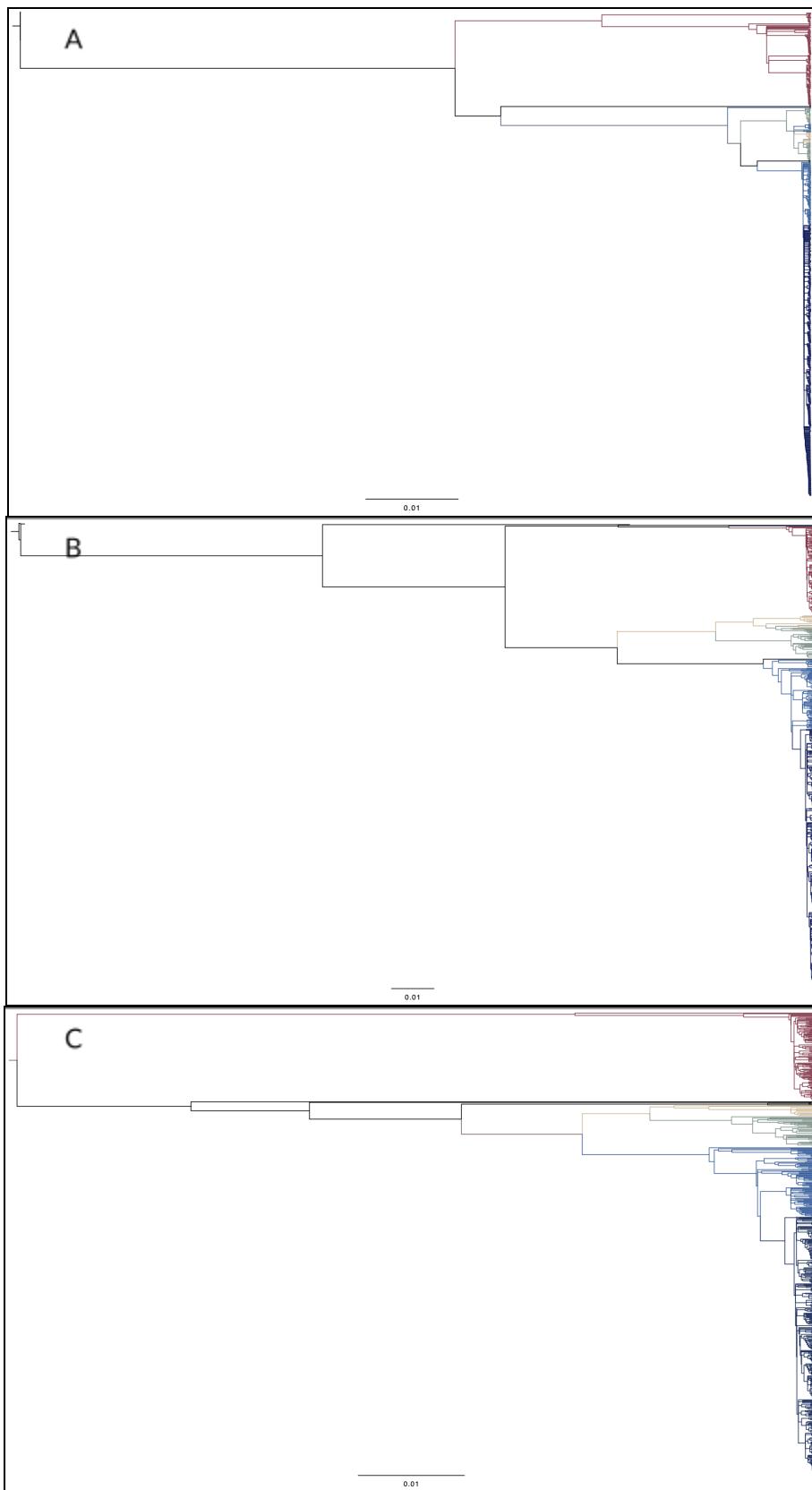
**Supplemental Figure 1.** Mosquitoes collected by the Connecticut Agricultural Experiment Station (CAES), 1992-2022. (A) Abundance of mosquitoes collected per year in Connecticut, colored by mosquito generation number and genus group. (B) Isolations of JCV in mosquitoes collected in Connecticut, colored by mosquito generation number and genus group.



**Supplemental Figure 2.** Seasonality of JCV infection in Connecticut mosquito species from 1997-2022. (top left) Abundance of univoltine mosquitoes in Connecticut per trap per epi-week, averaged from 1997-2022, colored by species. (bottom left) Pooled infection rate of univoltine mosquitoes in Connecticut per generation group per epi-week, colored by year. (top right) Abundance of multivoltine mosquitoes in Connecticut per trap per epi-week, averaged from 1997-2022, colored by species. (bottom right) Pooled infection rate of multivoltine mosquitoes in Connecticut per generation group per epi-week, colored by year.

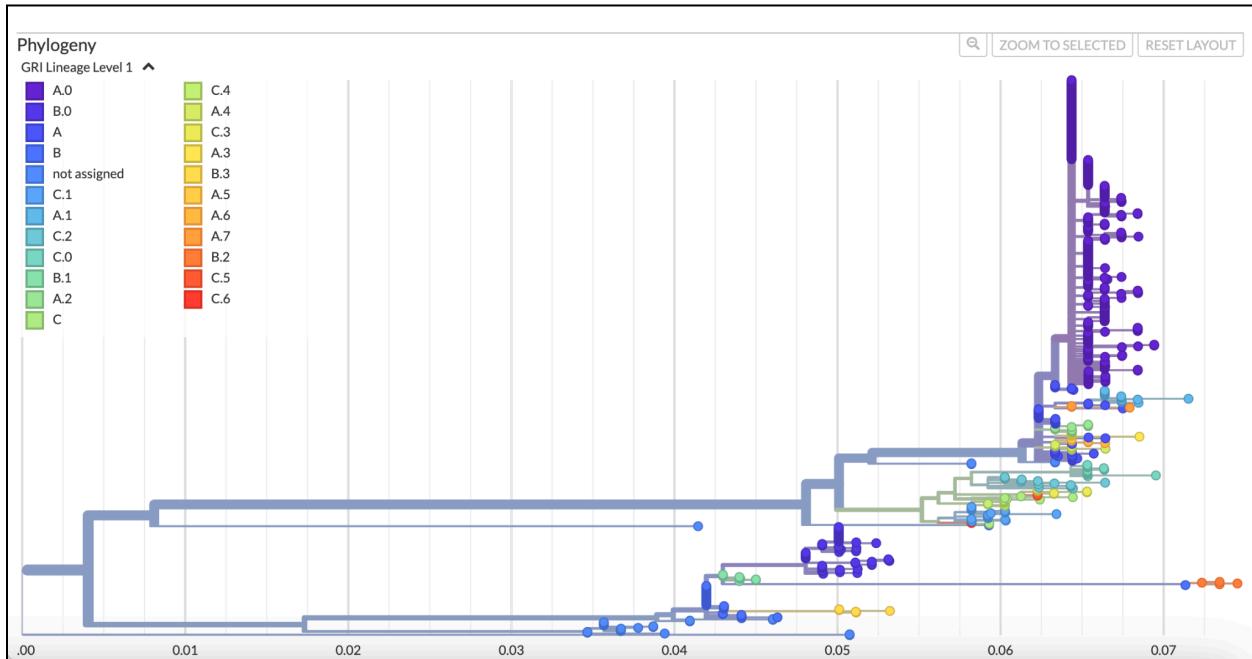
**Supplemental Table 1.** Unassigned JCV samples across genomic data from Connecticut, New York, Massachusetts, and North Dakota, from 1966-2023; samples removed from concatenated dataset

Sample Name	Collection Date	State	Mosquito Species	Small Segment Lineage	Medium Segment Lineage	Large Segment Lineage
ND Yale-JC0225 2005	2005-08-09	North Dakota	<i>Aedes vexans</i>	A-related	A-related	A.1.1
ND Yale-JC0226 2005	2005-08-09	North Dakota	<i>Aedes melanimon</i>	A-related	A-related	A.1.1
CT Yale-JC0251 2006	2006-07-11	Connecticut	<i>Aedes vexans</i>	A.1.1	B-related	A.1.1
CT Yale-JC0297 2007	2007-07-17	Connecticut	<i>Aedes canadensis</i>	B-related	B-related	B-related
CT Yale-JC0392 2011	2011-06-09	Connecticut	<i>Aedes cantator</i>	B.2.1	B.1.2	B-related
CT Yale-JC0522 2016	2016-06-29	Connecticut	<i>Aedes canadensis</i>	B-related	B-related	B-related
CT Yale-JC0528 2017	2016-09-01	Connecticut	<i>Anopheles quadrimaculatus</i>	B.2.1	B.1.2	B-related
CT Yale-JC0529 2017	2017-06-07	Connecticut	<i>Aedes abserratus</i>	A.1.1	B-related	B-related
CT Yale-JC0530 2017	2017-06-08	Connecticut	<i>Aedes abserratus</i>	B-related	B-related	B-related
CT Yale-JC0590 2018	2018-07-23	Connecticut	<i>Anopheles punctipennis</i>	A-related	B-related	A.1.1
NY Yale-JC0701 2012	2012-07-10	New York	<i>Aedes canadensis</i>	A.1.2	A-related	A-related

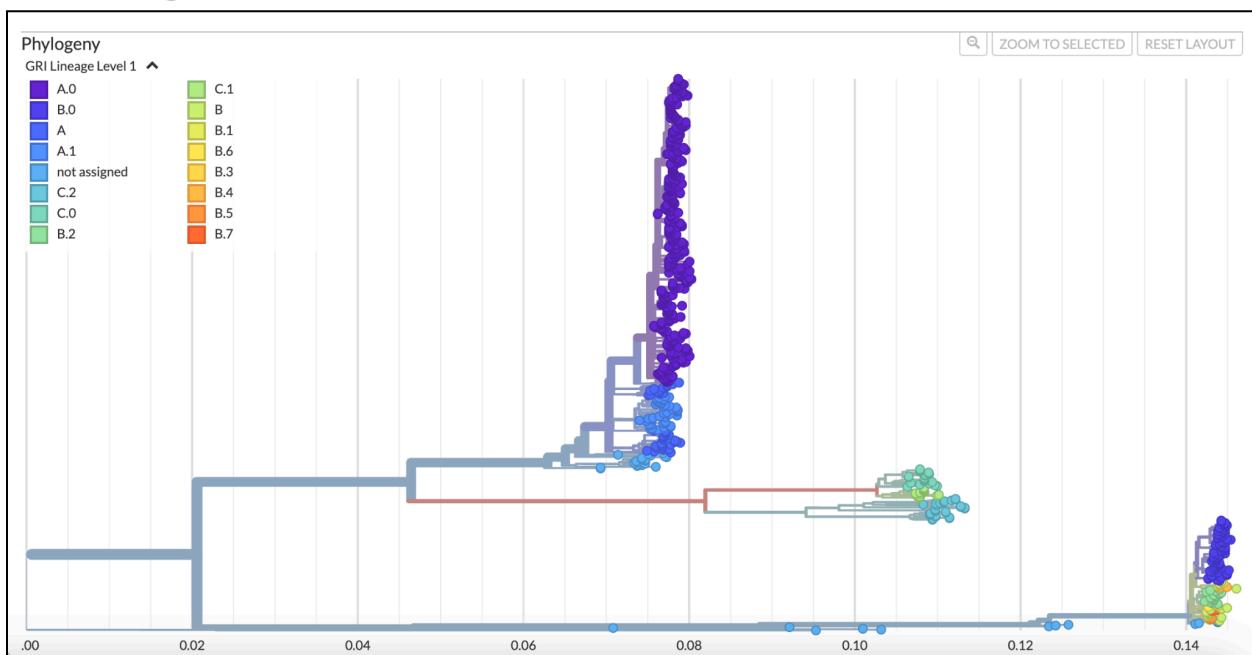


**Supplemental Figure 3.** Designation of JCV lineages in the northern United States, as determined by fastBAPS. (A) Maximum likelihood tree of the JCV small segment, colored by fastBAPS cluster. (B) Maximum likelihood tree of the JCV medium segment, colored by fastBAPS cluster. (C) Maximum likelihood tree of the JCV large segment, colored by fastBAPS cluster.

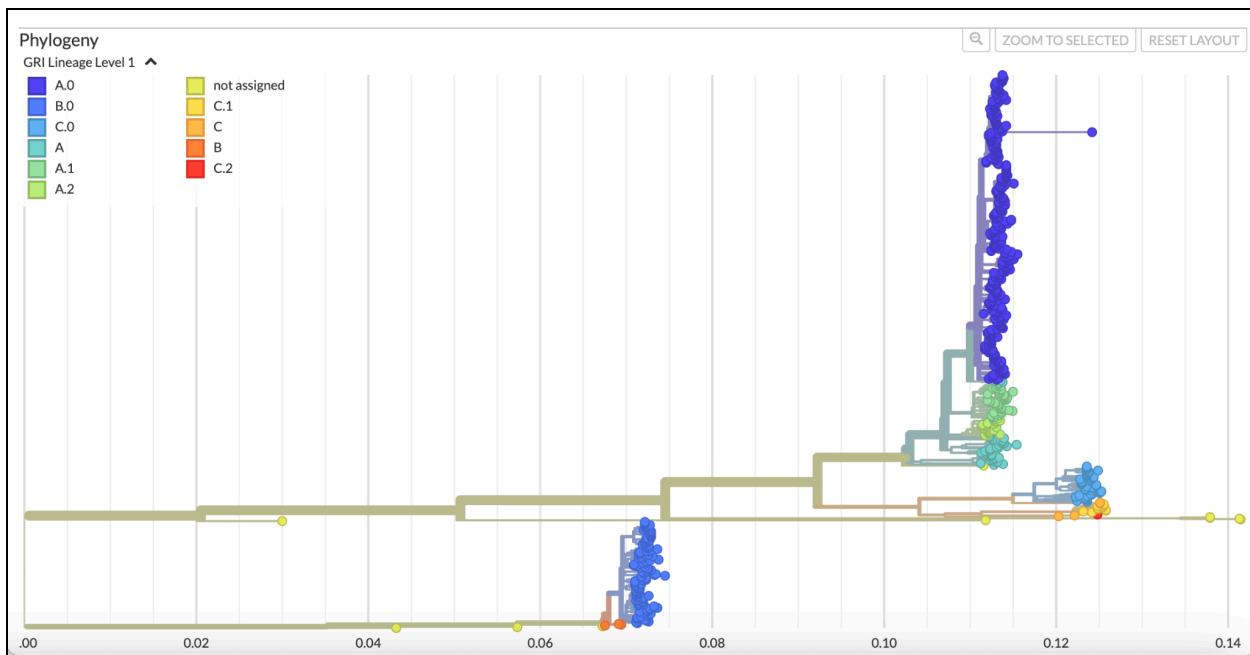
### A. Small segment



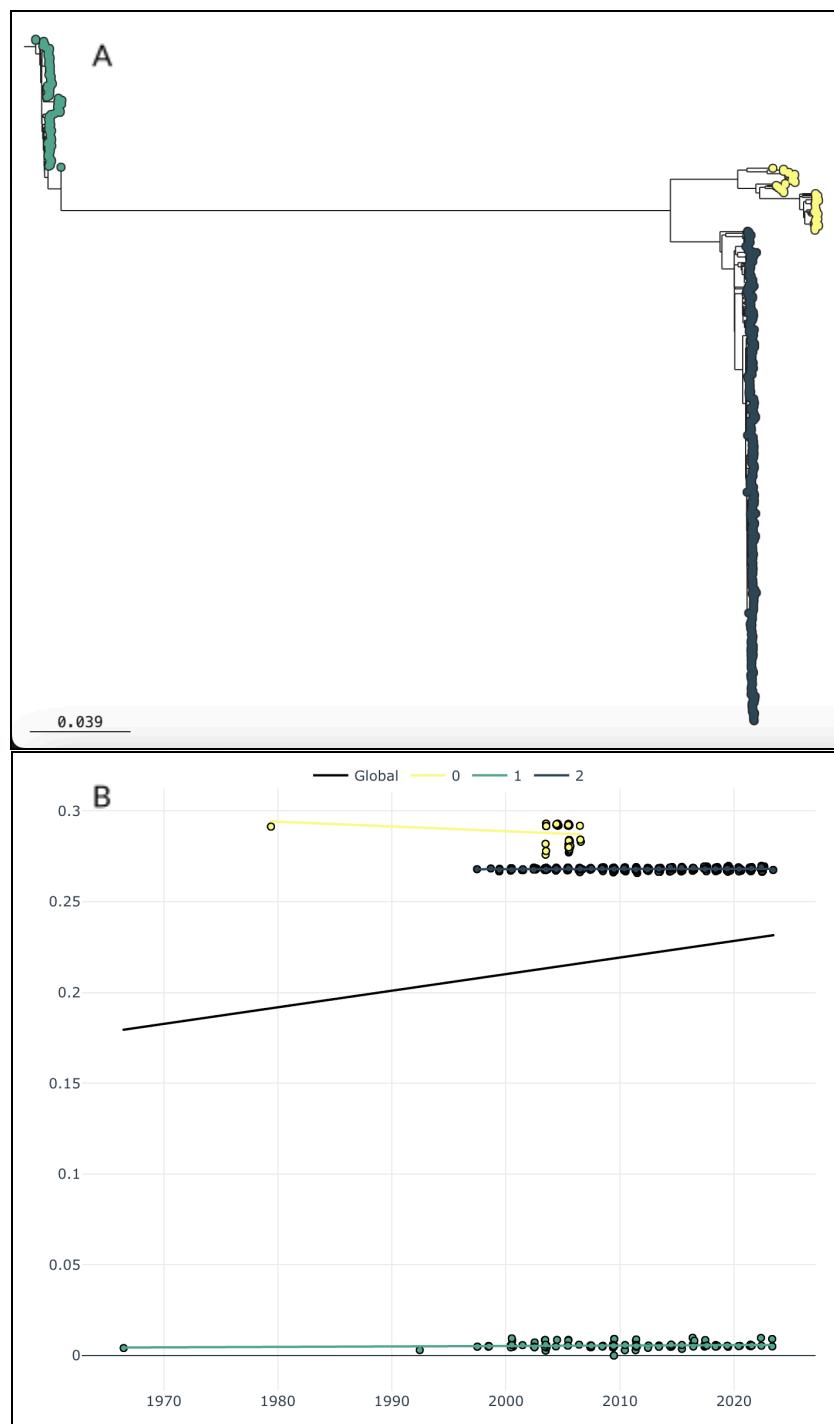
### B. Medium segment



## C. Large Segment



**Supplemental Figure 4.** Designation of JCV lineages in the northern United States, as determined by Autolin. (A) Maximum likelihood tree of the JCV small segment, colored by Autolin lineage. (B) Maximum likelihood tree of the JCV medium segment, colored by Autolin lineage. (C) Maximum likelihood tree of the JCV large segment, colored by Autolin lineage.



**Supplemental Figure 5.** Estimation of JCV evolutionary rates across the concatenated genome using Clockor2.(A) Maximum likelihood tree of the JCV concatenated genome, colored by clock group. (B) Root-to-tip plot of JCV evolution, colored by clock group.

**Supplemental Table 2.** Calculated local clock rates and associated R<sup>2</sup> scores for clock group of the concatenated JCV genome. BIC score was used as a comparison metric for the calculation of evolutionary rates.

<b>Local Clock</b>					
CLOCK	#TIPS	RATE	R <sup>2</sup>	RESIDUAL MEAN SQUARED	INTERCEPT
0	62	-2.625 x 10 <sup>-4</sup>	0.024	3.280 x 10 <sup>-5</sup>	3100.559
1	127	2.480 x 10 <sup>-5</sup>	0.015	2.454 x 10 <sup>-6</sup>	1788.551
2	483	9.033 x 10 <sup>-6</sup>	0.008	3.986 x 10 <sup>-7</sup>	-27664.199