Mechanisms of Emerging RNA Virus Adaptation to Hosts

Chelsea Nevins Cereghino

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Biomedical and Veterinary Sciences

James Weger-Lucarelli, Chair

Nisha K. Duggal

Kylene Kehn-Hall

Jeremy Draghi

November 21, 2025

Blacksburg, VA

Keywords: host adaptation, Mayaro virus, severe acute respiratory syndrome coronavirus 2, genetic determinants

Mechanisms of Emerging RNA Virus Adaptation to Hosts

Chelsea Nevins Cereghino

Abstract

Emerging RNA viruses cause acute and chronic diseases that threaten the livelihoods of people in regions where the virus is endemic. As factors such as climate change and increased globalization, travel, and trade alter the co-occurrence of humans and reservoirs or vectors, opportunities for viral geographic expansion or viral spillover increase. Viral adaptation enabling intra- or interspecies transmission can cause outbreaks on the scale of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) during the coronavirus disease 2019 pandemic. Understanding the mechanisms by which emerging RNA viruses adapt to hosts can inform public health strategies targeting the appropriate vectors and reservoirs of the virus or the design of effective therapeutics. Therefore, we sought to identify the viral genetic determinants of host adaptation of two contrasting viruses, Mayaro virus (MAYV) and SARS-CoV-2. We used either selective sweep detection methods or experimental evolution to identify viral mutations with putative adaptive potential. A selective sweep region was identified in the *Spike* gene of human SARS-CoV-2 sequences. A residue at site 519 in this region of *Spike* was identified that differed from that of closely related sarbecoviruses infecting bats and pangolins. The ancestral mutation H519N reduced the entry of pseudotyped viruses in human ACE2 (hACE2)-expressing cells and decreased replication in human lung cells through reduced hACE2 binding. Next, serial passaging of SARS-CoV-2 in cells expressing the animal ACE2 receptor identified the recurring mutation A222V in the *Spike* gene. Spike A222V enhanced replication of SARS-CoV-2 in primary white-tailed deer lung cells through an ACE2-independent mechanism. Lastly, serial passaging of MAYV in cells from two urban mosquito vectors, *Aedes aegypti and Aedes*

albopictus, identified a mutation in E2, T179N, that increased viral fitness in these cells. E2-T179N increased the transmission efficiency of MAYV by *Aedes aegypti* while coming at the cost of reduced fitness and virulence in mice. Taken together, these works highlight the impact of single mutations on virus fitness within and between hosts. Our findings underscore the importance of combining surveillance, limiting specific reservoir and vector exposure, and therapeutic design targeting adaptive residues to prevent the further evolution and emergence of SARS-CoV-2 and MAYV.

Mechanisms of Emerging RNA Virus Adaptation to Hosts

Chelsea Nevins Cereghino

General Audience Abstract

Viruses can cause short- and long-lived diseases that affect the wellbeing of people living in areas where the virus circulates. In recent years, increased globalization, travel, trade, and climate change has led both to the outbreak of new viruses and an increase in the scale of outbreaks for pre-existing viruses. These changes can result from increased exposure to animals that carry disease, such as mosquitoes or wild animals. The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus responsible for the coronavirus disease 2019 pandemic, displayed how virus evolution can lead to new diseases in humans. Understanding how virus evolution impacts the infection of a human, animal, or an insect can aid in the design of vaccines or drugs effective against new virus mutations that may, for example, lead to more severe disease in humans. To identify mutations in viruses that adapt the virus to different hosts, we used two viruses: the mosquito-borne Mayaro virus and SARS-CoV-2. First, we searched sequences of SARS-CoV-2 from humans and identified a region in Spike, the virus protein responsible for binding to the host receptor, that had undergone recent evolution. This region in Spike contained an amino acid unique to SARS-CoV-2 that was not found in related viruses that infect bats. We generated a SARS-CoV-2 mutant containing the bat virus amino acid and measured both a decreased ability of the mutant virus to enter human lung cells and decreased virus replication. We showed that the more bat virus-like SARS-CoV-2 could not bind as well to the host cell receptor to enter human cells. Next, we allowed SARS-CoV-2 to evolve in cells containing the receptor of either dogs, cats, mink, or deer, a proxy for animal infection. The resulting virus acquired a mutation in Spike that increased SARS-CoV-2 replication in deer

lung cells but not in human lung cells. The Spike mutation in SARS-CoV-2 has been identified in nature repeatedly across time, including in SARS-CoV-2 infecting dogs, cats, mink, and deer. Finally, we allowed Mayaro virus to evolve in two cell types from mosquitoes and identified a mutation in a protein that binds to the host cell receptor. This mutation increased the replication of Mayaro virus in mosquito cells and enhanced the transmission of Mayaro virus in live mosquitoes. We infected mice with the mutant virus and observed both decreased amounts of virus in the blood and decreased evidence of disease, suggesting this mutation may negatively impact the virus in a human host. Our studies identified mutations that may have been important for SARS-CoV-2 to initially infect humans and be maintained in wild animal populations such as deer, or for Mayaro virus, to be transmitted more efficiently by mosquitoes. Identification of these mutations in nature could provide evidence that the virus has evolved in a way that could promote outbreaks. Therefore, it is important to design vaccines or antivirals that are effective against viruses with these mutations and to routinely test animals or mosquitoes to identify mutated viruses that could pose threats to humans.

Dedication

This dissertation is dedicated to the young, aspiring scientists who have not yet found the confidence to pursue a doctoral degree. For as much quality science as I hope to leave behind as my legacy, I hope to provide twice the inspiration to the curious-minded to pursue science with unwavering doubt in their abilities.

Secondly, I dedicate this dissertation to the late plant geneticist, molecular biologist and my former mentor, Dr. Wellington Muchero. It was his simple saying “let’s do good science” that will forever echo in the back of my mind and guide my path.

Acknowledgements

Completing a PhD takes a team of people supporting you behind the scenes. I would like to thank those many people who played a role in my personal or scientific growth, technical training, or in providing emotional support during this journey.

I would like to thank my committee members--Jeremy Draghi, Kylene Kehn-Hall, and Nisha Duggal--for their instrumental feedback through all phases of my projects. Committee: I want to thank you for the immense contribution you made to the work I completed during my PhD. I am grateful you helped me through the many hurdles I hit, particularly in rescuing Haitian strains of Mayaro virus without mutations. Your outside perspective on my projects allowed me to move forward after encountering challenges. While I would sometimes enter the committee meetings feeling discouraged or possibly even a little hopeless, your feedback and suggestions left me feeling empowered and eager to get back into the lab. Thank you for taking the time to meet with me, for encouraging me, for renewing hope, and, of course, for refining my critical thinking and scientific skills.

To my lab members: I was incredibly fortunate to have a supportive team for the duration of my PhD. While some team members have come and gone, I am grateful to all of them for their suggestions, for taking time points, for fixing plates, for generating critical data for our manuscripts, for feedback on my many presentations, for training me on the bulk of the technical skills I now have, and for the camaraderie. I'd like to especially thank Christina Chuong, Jeffrey Marano, and Pallavi Rai for the extensive time you spent training me, all while showing patience and encouragement along the way. Katie Tiller: our friendship is so special to me. I look forward to our frequent scientific conversations every day. I love that we see eye to eye on so many

things, and having that support really helped me. I am so lucky you joined the lab and had someone like you to go through my PhD with.

I must acknowledge my mother and father for supporting me emotionally every step of the way. Thank you for understanding the undertaking of a PhD and understanding when I could not commit to an event, trip, etc. because of lab work. Instead of being upset, you have been proud of my accomplishments and tell everyone you know with joy. Having your support in this process may have just been what I needed to keep going even when my spirits were low. I couldn't have done this without you both as my biggest fans.

To my partner, Mike: thank you for listening to me discuss every failed experiment, and every success. Thank you for getting excited with me when something worked. I feel fortunate to have your shared enthusiasm for what I am doing. Thank you for also helping me step back and see the bigger picture and for helping make sure I didn't spend every waking hour in the lab. You've helped me learn about balance, and that has made me a happier and more productive scientist.

Finally, I must thank my advisor, James, for providing me with all of these things: personal growth, scientific growth, technical training, and emotional support. James: At the end of my PhD, I can look back and say that I enjoyed *every* moment of it because of you as my mentor. Thank you for taking a chance on me as your research technician and for your patience with me as I learned how to be an efficient laboratory worker. When I finally told you I was interested in graduate school, thank you for not letting me settle with choosing a master's degree. Your faith in me came at a time when it was most needed, and I am so grateful you provided me with that gentle push to take the leap of faith and begin a PhD. My time working with you has been an absolute joy. I am grateful for every weekly meeting, impromptu discussion, and Teams

message to give me guidance. I appreciate the late night you spent helping me write and edit the MAYV E2-T179N manuscript. You allowed me to pursue several scientific interests and even provided me with an opportunity to travel internationally and learn how to perform field research. Thank you for going above and beyond as a mentor. The way you have shaped me into the scientist I am today is something I could never express enough gratitude for.

Attribution

Chapter 2:

Contributed equally to this work: Chelsea Cereghino and Ferdinand Roesch

Chelsea Cereghino: Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing

Ferdinand Roesch: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing

Lucía Carrau: Conceptualization, Formal analysis, Investigation, Methodology, Writing – review & editing

Alexandra Hardy: Investigation, Methodology

Helder V. Ribeiro-Filho: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing

Annabelle Henrion-Lacritick: Investigation

Cassandra Koh: Data curation, Investigation, Methodology, Writing – review & editing

Jeffrey M. Marano: Formal analysis

Tyler A. Bates: Investigation

Pallavi Rai: Investigation

Christina Chuong: Investigation, Writing – review & editing

Shamima Akter: Investigation

Thomas Vallet: Investigation

Hervé Blanc: Investigation

Truitt J. Elliott: Investigation

Anne M. Brown: Formal analysis, Investigation, Methodology, Writing – review & editing

Pawel Michalak: Methodology, Writing – review & editing

Tanya LeRoith: Formal analysis, Investigation

Jesse D. Bloom: Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing

Rafael Elias Marques: Formal analysis, Methodology, Writing – review & editing

Maria-Carla Saleh: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing

Marco Vignuzzi: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing

James Weger-Lucarelli: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Writing – original draft, Writing – review & editing

Chapter 3:

Contributed equally to this work: Chelsea Cereghino and Kasia Michalak

Chelsea Cereghino: Data curation, Formal analysis, Investigation, Validation, Writing—original draft, Writing—review & editing:

Kasia Michalak: Data curation, Formal analysis, Investigation, Validation, Writing—review & editing

Stephen DiGiuseppe: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—review & editing, Resources, Supervision

Juan Guerra: Investigation

Delaney Yu: Investigation

Ariana Faraji: Investigation

Amanda Sharp: Data curation, Formal analysis, Investigation, Writing—review & editing

Anne M. Brown: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—review & editing, Supervision

Lin Kang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—review & editing

Pawel Michalak: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—review & editing, Funding acquisition, Project administration, Resources, Supervision

James Weger-Lucarelli: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—review & editing, Funding acquisition, Project administration, Resources, Supervision

Chapter 4:

Chelsea Cereghino: Data curation, Formal analysis, Investigation, Validation, Writing—original draft, Writing—review & editing:

Kateland Tiller: Data curation, Investigation, Writing—review & editing:

Lin Kang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—review & editing

Pawel Michalak: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—review & editing, Funding acquisition, Project administration, Resources, Supervision

James Weger-Lucarelli: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—review & editing, Funding acquisition, Project administration, Resources, Supervision

Licensing and Citations

Chapters 2-4 have been previously published and are cited and licensed as follows:

Paper	License
Cereghino C, Roesch F, Carrau L, Hardy A, Ribeiro-Filho HV, Henrion-Lacritick A, Koh C, Marano JM, Bates TA, Rai P, Chuong C, Akter S, Vallet T, Blanc H, Elliott TJ, Brown AM, Michalak P, LeRoith T, Bloom JD, Marques RE, Saleh MC, Vignuzzi M, Weger-Lucarelli J. The E2 glycoprotein holds key residues for Mayaro virus adaptation to the urban Aedes aegypti mosquito. PLoS Pathog. 2023 Apr 5;19(4):e1010491. doi: 10.1371/journal.ppat.1010491. PMID: 37018377; PMCID: PMC10109513.	Creative Commons Attribution 4.0 International (CC BY) license
Cereghino C, Michalak K, DiGiuseppe S, Guerra J, Yu D, Faraji A, Sharp AK, Brown AM, Kang L, Weger-Lucarelli J, Michalak P. Evolution at Spike protein position 519 in SARS-CoV-2 facilitated adaptation to humans. Npj Viruses. 2024 Jul 9;2(1):29. doi: 10.1038/s44298-024-00036-2. PMID: 40295673; PMCID: PMC11721114.	Creative Commons Attribution 4.0 International (CC BY) License
Cereghino C, Tiller K, Kang L, Michalak P, Weger-Lucarelli J. The recurrent Spike A222V mutation in SARS-CoV-2 enhances replication in primary deer lung cells. Virus Evol. 2025 Aug 5;11(1):veaf059. doi: 10.1093/ve/veaf059. PMID: 40874095; PMCID: PMC12378747.	Creative Commons CC BY

Table of Contents

Chapter 1: Introduction

RNA viruses evolve rapidly and cause large-scale outbreaks

Ribonucleic acid (RNA) viruses are obligate intracellular parasites that contain a genome composed of RNA that is replicated by a virally-encoded RNA-dependent RNA polymerase (RdRp). RNA viruses comprise 21 viral families[1], and many RNA viral species cause a significant disease burden in humans and animals. RNA viruses have caused isolated outbreaks, but others have been the causative agents of pandemics. A non-exhaustive list of RNA viruses that have caused pandemics include influenza A virus H1N1 for the influenza pandemic of 1917-1918[2], severe acute respiratory syndrome coronavirus 1 for the severe acute respiratory syndrome pandemic of 2002 to 2003[3], influenza A virus H1N1 for the swine flu pandemic of 2009-2010[4], and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) for the coronavirus disease 2019 (COVID-19) pandemic of 2020-present. Collectively, RNA viruses cause an estimated annual global burden of well over 1 billion cases[5–8], leading to an estimated cost of over 16 billion US dollars for direct healthcare costs alone[9–13]. Therefore, RNA viruses present a significant global health and economic burden, and strategies to mitigate their impacts and understand how to prevent pandemics should be a high priority to public health officials and scientific communities.

RNA viruses have historically caused more large-scale outbreaks compared to DNA viruses. This is in part due to their higher evolutionary rate that is postulated to be caused by the high mutation rate of RNA viruses resulting from lower replication fidelity, smaller average genome size, and shorter generation times in comparison to DNA viruses[14]. RdRps do not have proof-reading activity, though genome proofreading mechanisms exist for *Nidovirales* family members[15]. The outcome of a lack of proof-reading activity is that newly synthesized RNA containing incorrectly incorporated ribonucleotides retain mutations relative to the parent strand.

While not all DNA-dependent DNA polymerases have proofreading activity, some exhibit proofreading activity that contributes to an overall lower average error rate of 10^{-8} to 10^{-6} substitutions per nucleotide per cycle compared to that of an RNA virus at 10^{-6} and 10^{-4} substitutions[15]. Error-prone replication, mechanisms of genome recombination, and reassortment of genome segments leads to genetic diversity within the viral population. Genetic diversity generates novel phenotypes for selection to act upon, contributing to viral evolution[16]. In addition, the comparatively smaller genome size of RNA viruses versus DNA viruses increases the rate at which genomes are produced, leading to more rapid generation of mutated, recombined, and/or reassorted genomes[16]. These mechanisms allow RNA viruses more opportunity to evolve toward phenotypes that allow viruses to cause outbreaks, such as enhanced replication within a host, enhanced transmission between hosts, or the ability to switch host species (i.e., transmission from an animal to a human and vice versa).

Significance of emerging RNA viruses

Emerging viruses are those which have recently appeared in a population, recently increased in incidence, or have recently expanded their geographic range[17]. Most emerging pathogens are zoonoses--pathogens that are transmitted from animals to humans[18]. An analysis of zoonotic spillover events from severe acute respiratory syndrome coronavirus 1, Nipah virus, filoviruses, and Machupo virus from 1963 to 2019 showed annual increases in spillover events and deaths of 4.98% and 8.7%, respectively[19]. Furthermore, outbreaks of all infectious diseases have increased significantly since 1980, and 56% of the outbreaks occurring from 1980-2013 have been caused by zoonotic agents[20]. Factors mediating these outbreaks that increase the overlap of vectors or reservoirs with human populations include climate change, land use changes, increased globalization, human susceptibility to viral infection, travel and trade,

demand for non-traditional meat sources, and deforestation[21,22]. Predictions indicate that zoonotic spillover events will increase in the future and that, assuming a tripling rate of disease emergence, the probability of experiencing a severe pandemic event in one's lifetime will increase from 17% to 44%[23]. Predictions have also been made for the magnitude of outbreaks for specific virus species in the future. For example, one predictive model assessing the impact of climate change on vector distribution and incidence rate for dengue fever indicates that in years 2041-2060 there will be 76.9 million cases annually. This is in contrast to the annual case burden from 2000-2019 which estimated global cases of dengue fever of 49 million[24]. Emerging RNA viruses present challenges to health because effective treatments are lacking and the scale of infections can range from small to explosive outbreaks that are predicted to increase in the future.

The focus of this dissertation is on two emerging, zoonotic RNA viruses--Mayaro virus (MAYV) and SARS-CoV-2. Though unrelated taxonomically and contrasting in disease presentation, immune response, mode of transmission, and outbreak scale, these pathogens pose a significant challenge to public health due to either their emergence potential or adaptive evolution in human and reservoir populations. In the interest of understanding and combating the emergence and evolution of these viruses, we identified the viral genetic determinants contributing to these events and explained mechanisms by which these determinants increase adaptation to hosts.

Mayaro virus overview

MAYV is an emerging, zoonotic, mosquito-borne virus in the *Alphavirus* genus of the *Togaviridae* family. Alphaviruses are positive sense, single-stranded RNA viruses, largely transmitted by mosquitoes of the *Anophelinae* and *Culicinae* subfamilies[25]. Alphaviruses are

characterized by either an encephalitic or arthritogenic disease in humans. MAYV is part of the Semliki Forest virus antigenic complex comprising primarily arthritogenic virus members including Semliki Forest virus, Bebaru virus, chikungunya virus (CHIKV), Getah virus, Sagiyama virus, Ross River virus (RRV), O'nyong nyong virus, and Una virus[26,27]. The virus was identified in 1954 from forest workers in Trinidad[28]. Three canonical genotypes of MAYV have been described: D (dispersed), L (limited) and N (new)[29]. A recent phylogenetic examination of MAYV strains detected a new, recombinant genotype of L/D in a strain from Brazil and Haiti[30]. Mayaro fever is characterized by an acute, febrile illness and prolonged polyarthralgia and myalgia that can last at least a year in approximately 54% of cases[31]. The acute phase of disease results in a fever, malaise, headache, myalgia, retro-orbital pain, anorexia, nausea, vomiting, dizziness, dysgeusia, abdominal pain, rash, sore throat, and photophobia[31]. Severe cases of disease may result in hemorrhagic phenomena[32], encephalopathy[33], or rarely in mortality[33]. Underreporting of cases of Mayaro fever is likely, often because the symptoms are indistinguishable from other diseases caused by endemic alphaviruses like CHIKV and the flaviviruses dengue virus and Zika virus[34]. The cross reactivity of viral antigens from members of the Semliki Forest virus complex with antibodies produced after infection also makes diagnosis of Mayaro fever difficult[35]. As is common with other alphaviruses, treatment options for Mayaro virus disease are limited. Analgesics and nonsteroidal anti-inflammatory drugs are commonly used to help alleviate symptoms of acute and chronic disease[36].

Mayaro virus, like other alphaviruses, has broad tropism and uses various cell entry mechanisms. Entry into cells is mediated by binding of the structural protein E2 to a receptor, such as matrix remodeling associated 8 protein (Mxra8), on the cell membrane which triggers receptor-mediated endocytosis. Other, putative receptors include TIM4, VLDLR, ApoER2,

which increase replication in susceptible cells[37]. Macropinocytosis in vertebrate cells has also been described for MAYV[38]. In the endosome, acidification causes the E1 fusion protein to undergo a conformational change that exposes a fusion loop[39]. Insertion of the fusion loop into the endosomal membrane causes fusion of the viral envelope, releasing the nucleocapsid complex into the cytoplasm. Translation of the positive sense genomic RNA by host ribosomes results in a polyprotein comprising the non-structural proteins nsP1, nsP2, nsP3, and nsP4. The viral nsP2 protease cleaves between nsP3 and nsP4 to form the nsP123/4 replicase complex which synthesizes mainly negative sense RNA intermediates[40]. Maturation of the replicase complex into nsP1/2/3/4 leads to synthesis of exclusively positive sense genomic RNA[40]. Positive sense, subgenomic RNA is produced from a promoter in the negative strand which leads to the synthesis of a structural polyprotein Capsid-precursor to E2(pE2)-6K/TF-E1. Capsid autocleaves away from pE2-6K/TF-E1[40] and forms nucleocapsid structures with the newly synthesized genomic RNA. Polyprotein pE2-6K-E1 traffics to the endoplasmic reticulum where a host signal peptidase cleaves after pE2 and 6K. pE2 and E1 then assemble as heterodimers and are glycosylated in the endoplasmic reticulum and the golgi[41]. Furin processes pE2 into E3 and E2 in the golgi, and the pE2-E1 heterodimers translocate to the plasma membrane, where E3 is released depending on the viral species[42]. Interaction of capsid with the E2 and E1 glycoproteins promotes budding of the particle at the plasma membrane, allowing the virus to acquire an envelope[43].

Mayaro virus pathogenesis in humans

MAYV is transmitted to humans by the bite of a mosquito containing the infectious virus in the saliva. Upon a mosquito bite, the virus may be dispensed directly into the bloodstream or the virus may infect and replicate in the fibroblasts[44] and macrophages[45] of the skin[46]. There is also

evidence that MAYV may infect endothelial cells lining capillaries in humans[47]. Dissemination of MAYV is likely mediated by Ly6C+ monocytes, as CHIKV replicates in infiltrating Ly6C+ monocytes in the skin, facilitating dissemination of the virus via the draining lymph nodes[48]. In this case, the virus enters the bloodstream via the thoracic duct. The virus circulates to secondary sites of replication such as the liver and spleen which then seed the virus into joint and musculoskeletal tissues where replication continues[49]. MAYV infection elicits a robust interferon response[46,50], resulting in chemokines that recruit myeloid cells to sites of replication[46]. Neutrophils and inflammatory monocytes are recruited to sites of MAYV replication via the NLRP3 inflammasome, a mechanism that may contribute to early pain generation[51]. Studies with RRV indicate that it is the innate and not the adaptive immune infiltrate in muscle tissues that mediates arthritogenic disease severity[52]. CHIKV infection induces the recruitment of macrophages to sites of infection, and depletion of macrophages reduces footpad swelling in mice, suggesting their role in promoting CHIKV disease[53]. Macrophages and neutrophils expressing arginase 1 also contribute to persistence of viral RNA in inflamed tissues and are associated with greater muscle tissue pathology in RRV-infected mice [54]. For MAYV, CCR2-/- mice, which do not permit the influx of CCR2+ monocytes to sites of replication, did not lose any bone tissue, suggesting that these monocytes promote MAYV disease[55]. Long-term arthralgia is associated with viral RNA and viral antigens in the joint tissues, and proinflammatory cytokine levels are high in patients with chronic symptoms[56]. There are limited studies on the cytokines associated with patients with chronic MAYV disease versus those with mild disease that recover early. Santiago et al., 2015 observed significantly higher levels of IL-13, IL-7, and VEGF in patients during the acute and/or convalescent phases of MAYV disease that correlated with chronic arthralgia [56]. While no comparison was made between patients with chronic MAYV fever and those that recovered from

the disease, patients with arthralgia 30 days after symptom onset have elevated levels of IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-13, IL-17, interferon-γ–induced protein 10, RANTES, macrophage inflammatory proteins 1α and 1β, granulocyte-macrophage colony-stimulating factor (GM-CSF), and interferon-γ[57]. More extensive studies have been conducted with CHIKV, where Lohachanakul et al., 2012 determined that low levels of IL-8 combined with high levels of IL-6 and MCP-1 are associated with severe disease[58]. Ng et al., 2009 determined that low RANTES combined with high levels of IL-1β and IL-6 correlated with severe CHIKV disease[59]. In another study, four patients with persistent arthralgia had lower levels of eotaxin and hepatocyte growth factor but higher levels of IL-6 and GM-CSF compared to patients that recovered[60]. The elevated cytokines IL-6 and GM-CSF are common to chronic disease caused by MAYV and CHIKV and offer insights into combating the long-term arthralgia associated with MAYV and CHIKV disease.

Ecology, epidemiology and mechanisms of Mayaro virus transmission

MAYV has caused sporadic, small outbreaks of less than 100 cases across South America since it was discovered in 1954. MAYV is primarily maintained in a sylvatic cycle in the Amazon basin region of South America between non-human primates and forest-dwelling mosquitoes of the *Haemagogus janthinomys* species[61], although birds[62], small mammals[63], and reptiles[64] may be implicated in the sylvatic cycle as well. Other potential sylvatic vectors include other *Haemagogus* species[65], *Psorophora* species[66], *Mansonia venezuelensis*[67], *Mansonia humeralis*[68], *Culex* species[65], *Sabethes* species[65], and *Aedes scapularis*[65]. A rural cycle has been proposed for MAYV whereby the more anthropophilic vectors *Haemagogus janthinomys* and *Ae. scapularis* vectors may feed on humans who reside near or work in forested areas of the Amazon basin[69]. Most human cases occur when people visit rural and jungle areas in the Amazon basin

region of South America where the virus is endemic[70]. An urban transmission cycle, though occurring infrequently, is probable since MAYV can be transmitted by the urban *Ae. aegypti* in laboratory studies[37,71,72] and has been isolated from *Ae. aegypti* in Brazil[73]. Smaller, urban outbreaks of less than 100 cases occurred in Belterra, Pará, Brazil in 1978[74] and Boa Vista, Roraima, Brazil in 2021[75].

In 2014-2015, MAYV was detected in new geographic areas within the Caribbean nation of Haiti[76]. Evidence of three separate introductions of the virus into Haiti and the identification of five cases over a 10 month period[76] raise the concern that MAYV has adapted to mosquito species or mammalian hosts capable of supporting a transmission cycle within the island of Hispaniola. Given that the primary sylvatic vector *Haemagogus janthinomys* has not been identified in Haiti[77], it is possible that adaptation to urban vectors may have occurred prior to or during the outbreak. The lack of a wild population of nonhuman primates in Hispaniola could also support the hypothesis that Haitian MAYV strains have further adapted toward humans or other vertebrate animals[78]. Further adaptation of MAYV either to a mosquito vector or to humans could result in more frequent urban transmission and outbreaks.

SARS-CoV-2 overview

SARS-CoV-2 is a zoonotic RNA virus that emerged in the human population in 2019 and is the causative agent of the COVID-19 pandemic that began in 2020[79]. SARS-CoV-2 is transmissible by respiratory droplets when infected individuals cough, sneeze, speak, or breathe[80]. COVID-19 is asymptomatic in approximately 80% of infections[82] but otherwise affects the upper and lower respiratory tracts. Fever, fatigue, cough, sputum production, loss of sense of taste and smell, difficulty breathing, sore throat, and chills are the primary acute symptoms of the disease[83]. Severe cases of COVID-19 may result in pneumonia, acute respiratory distress syndrome, or

septic shock, requiring hospitalization[81]. COVID-19 can also involve a chronic phase of disease. Long COVID or post-acute sequelae of COVID-19 is a chronic, non-curable disease that occurs in approximately 10% of SARS-CoV-2 infections and involves multiple organs such as the lungs, heart, blood vessels, immune system, gastrointestinal tract, pancreas, kidneys, spleen, liver, nervous system, and reproductive system[84]. COVID-19 can cause mortality, particularly in the elderly or in immunocompromised patients, and as of July 2025 has resulted in cumulative fatalities in the United States of 1.23 million since the beginning of the pandemic[85]. As a life-saving measure early in the pandemic, a COVID-19 vaccine was made available in December 2020 under the United States Food and Drug Administration emergency use authorization. The first vaccine was an mRNA vaccine encoding the Spike glycoprotein developed by Pfizer-BioNTech, which demonstrated 85% effectiveness in preventing death in older adults after a single dose[86]. Additional vaccines available in the United States were developed by Moderna and Johnson & Johnson. The Moderna vaccine was also an mRNA-based vaccine encoding the Spike glycoprotein sequence[87] while the Johnson & Johnson vaccine was a less effective, virally vectored vaccine[88]. Other companies supported vaccination efforts in countries outside of the United States, and vaccines were based on either mRNA, inactivated SARS-CoV-2, subunit, or adenovirus vectors[89,90]. Vaccination strategies are currently being updated as the virus evolves[91]. The current treatment options for SARS-CoV-2 infection include FDA-approved antivirals nirmatrelvir with ritonavir (Paxlovid), baricitinib with tocilizumab, and remdesivir[92].

SARS-CoV-2 is part of the subfamily *Orthocoronavirinae* of the large *Coronaviridae* family of viruses that contain an enveloped nucleocapsid and a positive sense, singled-stranded RNA genome[93,94]. Coronaviruses are distinctive amongst the RNA viruses for their large genomes of 26-32 kilobases in length[94]. Of the four genera in the *Orthocoronavirinae* subfamily[95],

Betacoronavirus contains the most human pathogens and includes SARS-CoV-2[94]. Betacoronaviruses cause both respiratory and gastrointestinal disease in humans with some species causing mild, seasonal illness[94]. The species HKU1 typically causes mild upper respiratory disease but progression to pneumonia and bronchiolitis sometimes occurs[96]. OC43 causes upper respiratory disease but is associated with neuroinvasion and multiple sclerosis[97]. In addition to mild disease, severe epidemics or pandemics have been caused by betacoronaviruses: Severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2. Both SARS-CoV and SARS-CoV-2 share sequence similarity to other SARS-related coronaviruses (SARSr-CoV) in the *Sarbecovirus* subgenus that infect bats, palm civets, or pangolins, and as such, these animals are often the suspected species of origin for these pandemics[98,99]. Indeed, isolation of SARSr-CoVs from palm civets in a wildlife market identified four viruses with 99.8% sequence similarity to the human SARS-CoV, suggesting these animals may have played a role in the emergence of SARS-CoV[100]. Additionally, recent field investigations have identified SARSr-CoVs in colonies of horseshoe bats in China that have 93.2-96% nucleotide identity to the human SARS-CoV and palm civet-derived SARSr-CoVs. One of these viruses, WIV1, can replicate efficiently in human primary airway epithelial cells[101] Since farmed palm civets do not have high exposure to SARSr-CoVs[102], the current evidence suggests bats may be the natural reservoir of the progenitor to SARS-CoV with palm civets acting as an intermediate host[103].

The genome of SARS-CoV-2 comprises a 5’ cap followed by non-structural proteins, structural and accessory proteins at the 3’ end, and a poly-A tail. The non-structural proteins are involved in critical processes of replication and transcription of the genome[104]. The structural proteins spike, envelope, membrane, and nucleocapsid are key proteins required for virion

assembly[105], and some are involved in pathogenesis and antagonism of the interferon response[106]. The accessory proteins are not critical for virus replication and instead antagonize the host interferon response[107]. The virion itself is composed of a host-derived lipid bilayer membrane decorated with transmembrane proteins envelope, membrane, and spike[108–110]. Infection of cells is mediated by the interaction of the viral spike glycoprotein and angiotensin-converting enzyme 2 (ACE2) on cells. Spike, which is composed of S1 and S2 subunits[111], binds to receptor ACE2 via the receptor binding motif within the receptor binding domain of the S1 subunit. Cleavage by host proteases transmembrane serine protease 2 (TMPRSS2) or cathepsins in S2 potentiates Spike rearrangement and fusion of the viral and host membranes either in the endosome or at low pH, at the surface of the cell[112,113]. The positive-sense viral RNA is uncoated and released into the cell to initiate translation of the viral genome. Polyproteins 1a and 1b are translated from the non-structural protein genes, cleaved by viral proteases, and the non-structural proteins assemble into the replication transcription complex (RTC). The RTC generates full-length negative sense RNA which is used as a template for positive sense genomic RNA and for 10 sgRNAs[114]. Replication of the genome occurs in double-membrane vesicles where the RNA is shielded from recognition by innate sensors MDA5 and RIG-I[115]. Both the structural and accessory proteins are translated from the sgRNAs, and translation of the structural proteins occurs in the endoplasmic reticulum (ER)[116]. Spike, envelope, membrane, and nucleocapsid are implanted in the ER membrane and are transferred to the ER golgi intermediate complex where they join newly formed nucleocapsids and participate in budding. Spike protein maturation occurs in the Golgi where furin and furin-like proprotein convertases process Spike into S1 and S2 subunits that remain non-covalently associated[117]. Virions traffic to the plasma membrane through the secretory pathway[117] and are released into the extracellular space to infect new cells.

SARS-CoV-2 pathogenesis and transmission

SARS-CoV-2 is transmissible primarily by aerosol and respiratory droplets, and as such enters the airway through inhalation. In the upper respiratory epithelium, the virus likely infects multiciliated epithelial cells[118,119]. As the virus is aspirated into the bronchioles and alveoli of the lung, alveolar cells AT1 or AT2 can become infected[119]. During viral replication, cytoplasmic sensors MDA5 and RIG-I recognize double stranded RNA that leads to the production of type I and III interferons, but this response is too late to control early viral replication[120]. Infection of lung epithelial cells leads to the production of cytokines and chemokines, and infection of local tissue-resident macrophages triggers pyroptosis and the production of chemokines and interferons[115]. Macrophages also play a role in the immunopathogenesis of SARS-CoV-2 via Toll-like receptor signaling (TLR). TLR2 recognition of the SARS-CoV-2 envelope and spike proteins by macrophages results in the production of pro-inflammatory cytokines and chemokines IL-1, IL-6, TNF-α, GM-CSF, CXCL8, IL-1β, CXCL10, MCP-1, CCL3, IFN-γ, IL-1α, MCP-1, and G-CSF[121–123]. These cytokines are positively correlated with severe COVID-19[122]. Alternative pathways to cytokine release during COVID-19 can occur in later stages of infection through mitochondrial damage in endothelial cells. Mitochondrial DNA activates the cyclic GMP-AMP synthase–stimulator of interferon genes pathway and leads to type I interferon production. Macrophages that engulf the endothelial cells are also stimulated to produce type I interferons[124]. This leads to severe lung inflammation and tissue damage. In addition, severe clinical outcomes of COVID-19 are mediated by a cytokine storm that results from PANoptosis cell death triggered by TNF-α and IFN-γ. PANoptosis further leads to pro-inflammatory cytokine release and ultimately results in systemic inflammation. In late stages of COVID-19, organs can fail, leading to fatality[125].

Epidemiology of the COVID-19 pandemic and SARS-CoV-2 origin

The first case of COVID-19 occurred in Wuhan city, Hubei province, China on December 10, 2019[126]. By the end of December 2019, there were 174 cases in Wuhan City[127]. Thirty-two percent of the cases were linked to the Huanan Seafood Wholesale Market where live and deceased wild and domestic animals were sold[127]. Given the long-standing link between live animal markets and infectious disease outbreaks and the association of COVID-19 cases with this market, the Chinese government shut down the market on January 1, 2020[128]. By this point in time, however, an epidemic in China had begun that peaked on February 12[129]. During January 2020, SARS-CoV-2 spread to the United States, Thailand, Japan, the Republic of Korea, Germany, and Vietnam[130,131]. The virus continued to spread to additional countries. The World Health Organization declared COVID-19 a pandemic on March 11, 2020[131]. As of July 26, 2025, there have been 778,457,848 reported cases globally[132]. While weekly global cases of COVID-19 and mortality have significantly decreased since its peak in January of 2023[132], the virus continues to evolve, largely due to pressures from either natural or vaccine-induced immunity[133]. Indeed, most circulating lineages as of summer 2025 are sublineages of the Omicron variant that harbor the Spike mutation F456L[134] that evades neutralizing antibodies[135].

The original source of SARS-CoV-2 and its means of spillover to humans remains unknown. Through extensive sampling, sequencing, and phylo-geographic and -dynamic analyses, however, evidence supports a strong hypothesis for the Huanan Seafood Market in Wuhan as the location of the spillover event of SARS-CoV-2. Sixty-six percent of the patients initially infected with SARS-CoV-2 had previous exposure to the Huanan Seafood Market in Wuhan[136]. Environmental samples taken from the market that were located in spaces where live

mammals were sold tested positive for SARS-CoV-2 RNA[137]. Two lineages of SARS-CoV-2, A and B, have been identified from environmental samples from the market and human cases, with lineage B dominating the early cases of those who had visited the market. The two lineage A cases, though reported to have no association with the market, have a statistically higher likelihood of being linked to the market than not based on the geographic location of other COVID-19 cases[128]. The presence of these two lineages of SARS-CoV-2 at the market is consistent with multiple spillover events, and the appearance of these lineages in humans coincide within weeks of each other[138]. Of the animals at the market, civets were a suspected intermediate host for SARS-CoV-2 since it was a proposed source for SARS-CoV[139]. As mentioned previously, SARSr-CoVs that have a 99.8% identity with SARS-CoV have been isolated from palm civets from a wildlife market in Guangdong, China[100]. Since no sampling of SARS-CoVs in wild civets has been performed since the start of the COVID-19 pandemic, there is not enough evidence that civets were harboring SARSr-CoVs similar enough to SARS-CoV-2 to support their case as the source of SARS-CoV-2. Other animals suspected as reservoirs of SARS-CoV-2 include pangolins since they also harbor SARSr-CoVs[140,141]. Pangolin-CoV-2020, a SARSr-CoVs isolated from pangolins, demonstrates a 84.52% nucleotide identity in *Spike* with SARS-CoV-2 and 90.32% identity at the whole genome level[142]. However, based on phylogenetic analysis of the whole genome of Pangolin-CoV-2020, it is not likely that this virus was the progenitor to SARS-CoV-2[142]. In addition, pangolins were not amongst the animals sold at the Huanan Seafood Wholesale Market during the time preceding and during the initial outbreak[143], refuting the idea that it could have been an intermediate host or source of the virus. Rather, bats present a stronger case as the origin of SARS-CoV-2. SARSr-CoVs found in bats display a higher sequence similarity to SARS-CoV-2 than Pangolin-CoV-2020--RaTG13 isolated from a

Rhinolophus affinis is 96.2% similar to SARS-CoV-2[79]. Several bat viruses were also discovered in 2021 from northern Laos, one of which has a high whole-genome similarity of 96.8% with SARS-CoV-2. This virus, BANAL-20-52, has a higher S1 similarity to SARS-CoV-2 than RaTG13. Still, a virus with a similarity higher than 99% identity to SARS-CoV-2 has not been identified, though one exists for SARS-CoV. While the origin of SARS-CoV-2 remains unknown, the bioinformatic and evolutionary analyses of SARSr-CoVs isolated from bats support the hypothesis that SARS-CoV-2 was derived from a progenitor that infects a bat host and spilled over to humans at the Huanan Seafood Market through an unknown intermediate host[144].

Objectives/Aims

The aim of this dissertation is to define the mechanism by which two, diverse emerging RNA viruses--MAYV and SARS-CoV-2--adapt to hosts. Mechanisms of viral adaptation to vertebrates will be described for both viruses, as well as the mosquito vector for Mayaro virus and animal reservoir species for SARS-CoV-2.

References

1. Payne, S. Introduction to RNA viruses. in *Viruses* 97–105 (Elsevier, 2017).
2. Orthomyxoviridae. https://ictv.global/report_9th/RNAneg/Orthomyxoviridae.
3. Cherry, J. D. & Krogstad, P. SARS: the first pandemic of the 21st century. *Pediatr. Res.* **56**, 1–5 (2004).
4. Al Hajjar, S. & McIntosh, K. The first influenza pandemic of the 21st century. *Ann. Saudi Med.* **30**, 1–10 (2010).

5. Ogilvie, I., Khoury, H., El Khoury, A. C. & Goetghebeur, M. M. Burden of rotavirus gastroenteritis in the pediatric population in Central and Eastern Europe: serotype distribution and burden of illness. *Hum. Vaccin.* **7**, 523–533 (2011).
6. CDC. Norovirus Facts and Stats. *Norovirus* https://www.cdc.gov/norovirus/data-research/index.html#:~:text=Norovirus%20worldwide&text=Worldwide%2C%20norovirus%20annually%20causes%20on,total%20cases%20of%20acute%20gastroenteritis (2024).
7. Influenza (seasonal). https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)#:~:text=There%20are%20around%20a%20billion,650%20000%20respiratory%20deaths%20annually.
8. Dengue and severe dengue. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.
9. Bartsch, S. M., Lopman, B. A., Ozawa, S., Hall, A. J. & Lee, B. Y. Global economic burden of Norovirus gastroenteritis. *PLoS One* **11**, e0151219 (2016).
10. Shepard, D. S., Undurraga, E. A., Halasa, Y. A. & Stanaway, J. D. The global economic burden of dengue: a systematic analysis. *Lancet Infect. Dis.* **16**, 935–941 (2016).
11. Rheingans, R. D. *et al.* Economic costs of rotavirus gastroenteritis and cost-effectiveness of vaccination in developing countries. *J. Infect. Dis.* **200 Suppl 1**, S16–27 (2009).
12. de Roo, A. M., Vondeling, G. T., Boer, M., Murray, K. & Postma, M. J. The global health and economic burden of chikungunya from 2011 to 2020: a model-driven analysis on the impact of an emerging vector-borne disease. *BMJ Glob. Health* **9**, e016648 (2024).
13. Putri, W. C. W. S., Muscatello, D. J., Stockwell, M. S. & Newall, A. T. Economic burden of seasonal influenza in the United States. *Vaccine* **36**, 3960–3966 (2018).
14. Stern, A. & Andino, R. Viral Evolution. in *Viral Pathogenesis* 233–240 (Elsevier, 2016).

15. Peck, K. M. & Lauring, A. S. Complexities of viral mutation rates. *J. Virol.* **92**, (2018).
16. Šimičić, P. & Židovec-Lepej, S. A glimpse on the evolution of RNA viruses: Implications and lessons from SARS-CoV-2. *Viruses* **15**, 1 (2022).
17. Johnson, N. A short introduction to disease emergence. in *The Role of Animals in Emerging Viral Diseases* 1–19 (Elsevier, 2014).
18. Taylor, L. H., Latham, S. M. & Woolhouse, M. E. Risk factors for human disease emergence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 983–989 (2001).
19. Meadows, A. J., Stephenson, N., Madhav, N. K. & Oppenheim, B. Historical trends demonstrate a pattern of increasingly frequent and severe spillover events of high-consequence zoonotic viruses. *BMJ Glob. Health* **8**, e012026 (2023).
20. Smith, K. F. *et al.* Global rise in human infectious disease outbreaks. *J. R. Soc. Interface* **11**, 20140950 (2014).
21. Ameni, G. *et al.* A narrative review on the pandemic zoonotic RNA virus infections occurred during the last 25 years. *J. Epidemiol. Glob. Health* **14**, 1397–1412 (2024).
22. Seibold, M. A. *et al.* Risk factors for SARS-CoV-2 infection and transmission in households with children with asthma and allergy: A prospective surveillance study. *J. Allergy Clin. Immunol.* **150**, 302–311 (2022).
23. Marani, M., Katul, G. G., Pan, W. K. & Parolari, A. J. Intensity and frequency of extreme novel epidemics. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2105482118 (2021).
24. Hu, J., Horton, B. P., Yeo, T. W., Sung, J. J. Y. & Steve, Y. H. L. Mosquito and global dengue cases in a warming world. *BMJ Glob. Health* **10**, (2025).

25. Williams, M. C., Woodall, J. P., Corbet, P. S. & Gillett, J. D. O'nyong-nyong fever: An epidemic virus disease in east Africa. 8. Virus isolations from anopheles mosquitoes. *Trans. R. Soc. Trop. Med. Hyg.* **59**, 300–306 (1965).

26. Powers, A. M. *et al.* Evolutionary relationships and systematics of the alphaviruses. *J. Virol.* **75**, 10118–10131 (2001).

27. Forrester, N. L. *et al.* Genome-scale phylogeny of the alphavirus genus suggests a marine origin. *J. Virol.* **86**, 2729–2738 (2012).

28. Anderson, C. R., Downs, W. G., Wattley, G. H., Ahin, N. W. & Reese, A. A. Mayaro virus: a new human disease agent. II. Isolation from blood of patients in Trinidad, B.W.I. *Am. J. Trop. Med. Hyg.* **6**, 1012–1016 (1957).

29. Auguste, A. J. *et al.* Evolutionary and ecological characterization of Mayaro virus strains isolated during an outbreak, Venezuela, 2010. *Emerg. Infect. Dis.* **21**, 1742–1750 (2015).

30. Mavian, C. *et al.* Emergence of recombinant Mayaro virus strains from the Amazon basin. *Sci. Rep.* **7**, 8718 (2017).

31. Halsey, E. S. *et al.* Mayaro virus infection, Amazon Basin region, Peru, 2010-2013. *Emerg. Infect. Dis.* **19**, 1839–1842 (2013).

32. Mourão, M. P. G. *et al.* Mayaro fever in the city of Manaus, Brazil, 2007-2008. *Vector Borne Zoonotic Dis.* **12**, 42–46 (2012).

33. Joel Navarrete-Espinosa, H. G.-D. Arbovirus causales de fiebre hemorrágica en pacientes del Instituto Mexicano del Seguro Social. *Rev. Med. Inst. Mex. Seguro Soc.* **44**, 347–353 (2006).

34. CDC. Clinical Diagnosis, Treatment, and Prevention of Mayaro Virus Disease. *Mayaro Virus* https://www.cdc.gov/mayaro-virus/hcp/clinical-diagnosis-treatment/index.html (2024).

35. Smith, J. L. *et al.* Human antibody responses to emerging Mayaro virus and cocirculating Alphavirus infections examined by using structural proteins from nine New and Old World lineages. *mSphere* **3**, (2018).
36. Abdelnabi, R. & Delang, L. Antiviral strategies against arthritogenic alphaviruses. *Microorganisms* **8**, (2020).
37. Cereghino, C. *et al.* The E2 glycoprotein holds key residues for Mayaro virus adaptation to the urban Aedes aegypti mosquito. *PLoS Pathog.* **19**, e1010491 (2023).
38. Pujhari, S. *et al.* Characterization of Mayaro virus (strain BeAn343102) biology in vertebrate and invertebrate cellular backgrounds. *J. Gen. Virol.* **103**, (2022).
39. Ribeiro-Filho, H. V. *et al.* Cryo-EM structure of the mature and infective Mayaro virus at 4.4 Å resolution reveals features of arthritogenic alphaviruses. *Nat. Commun.* **12**, 3038 (2021).
40. Skidmore, A. M. & Bradfute, S. B. The life cycle of the alphaviruses: From an antiviral perspective. *Antiviral Res.* **209**, 105476 (2023).
41. Ren, S. C., Qazi, S. A., Towell, B., Wang, J. C.-Y. & Mukhopadhyay, S. Mutations at the Alphavirus E1'-E2 interdimer interface have host-specific phenotypes. *J. Virol.* **96**, e0214921 (2022).
42. Holmes, A. C., Basore, K., Fremont, D. H. & Diamond, M. S. A molecular understanding of alphavirus entry. *PLoS Pathog.* **16**, e1008876 (2020).
43. Suomalainen, M., Liljeström, P. & Garoff, H. Spike protein-nucleocapsid interactions drive the budding of alphaviruses. *J. Virol.* **66**, 4737–4747 (1992).
44. Sugasti-Salazar, M., Llamas-González, Y. Y., Campos, D. & González-Santamaría, J. Inhibition of p38 mitogen-activated protein kinase impairs Mayaro virus replication in human dermal fibroblasts and HeLa cells. *Viruses* **13**, 1156 (2021).

45. Hernández-Sarmiento, L. J., Tamayo-Molina, Y. S., Valdés-López, J. F. & Urcuqui-Inchima, S. Mayaro virus infection elicits a robust pro-inflammatory and antiviral response in human macrophages. *Acta Trop.* **252**, 107146 (2024).

46. Esterly, A. T., Lloyd, M. G., Upadhyaya, P., Moffat, J. F. & Thangamani, S. A human skin model for assessing arboviral infections. *JID Innov.* **2**, 100128 (2022).

47. Campos, D. *et al.* Differential susceptibility of human microglia HMC3 cells and brain microvascular endothelial HBEC-5i cells to Mayaro and Una virus infection. *Front. Virol.* **3**, (2024).

48. Holmes, A. C. *et al.* Ly6C+ monocytes in the skin promote systemic alphavirus dissemination. *Cell Rep.* **43**, 113876 (2024).

49. Diagne, C. T. *et al.* Mayaro virus pathogenesis and transmission mechanisms. *Pathogens* **9**, 738 (2020).

50. Figueiredo, C. M. *et al.* Mayaro virus replication restriction and induction of muscular inflammation in mice are dependent on age, type-I interferon response, and adaptive immunity. *Front. Microbiol.* **10**, 2246 (2019).

51. de Castro-Jorge, L. A. *et al.* The NLRP3 inflammasome is involved with the pathogenesis of Mayaro virus. *PLoS Pathog.* **15**, e1007934 (2019).

52. Morrison, T. E. *et al.* Characterization of Ross River virus tropism and virus-induced inflammation in a mouse model of viral arthritis and myositis. *J. Virol.* **80**, 737–749 (2006).

53. Gardner, J. *et al.* Chikungunya virus arthritis in adult wild-type mice. *J. Virol.* **84**, 8021–8032 (2010).

54. Stoermer, K. A. *et al.* Genetic ablation of arginase 1 in macrophages and neutrophils enhances clearance of an arthritogenic alphavirus. *J. Immunol.* **189**, 4047–4059 (2012).

55. Santos, F. M. *et al.* Essential role of the CCL2-CCR2 axis in Mayaro virus-induced disease. *J. Virol.* **98**, e0110223 (2024).

56. Santiago, F. W. *et al.* Long-Term Arthralgia after Mayaro Virus Infection Correlates with Sustained Pro-inflammatory Cytokine Response. *PLoS Negl. Trop. Dis.* **9**, e0004104 (2015).

57. Tappe, D. *et al.* Sustained elevated cytokine levels during recovery phase of Mayaro virus infection. *Emerg. Infect. Dis.* **22**, 750–752 (2016).

58. Lohachanakul, J., Phuklia, W., Thannagith, M., Thonsakulprasert, T. & Ubol, S. High concentrations of circulating interleukin-6 and monocyte chemotactic protein-1 with low concentrations of interleukin-8 were associated with severe chikungunya fever during the 2009-2010 outbreak in Thailand: Determinants of chikungunya severity. *Microbiol. Immunol.* **56**, 134–138 (2012).

59. Ng, L. F. P. *et al.* IL-1beta, IL-6, and RANTES as biomarkers of Chikungunya severity. *PLoS One* **4**, e4261 (2009).

60. Chow, A. *et al.* Persistent arthralgia induced by Chikungunya virus infection is associated with interleukin-6 and granulocyte macrophage colony-stimulating factor. *J. Infect. Dis.* **203**, 149–157 (2011).

61. Hoch, A. L., Peterson, N. E., LeDuc, J. W. & Pinheiro, F. P. An outbreak of Mayaro virus disease in Belterra, Brazil. III. Entomological and ecological studies. *Am. J. Trop. Med. Hyg.* **30**, 689–698 (1981).

62. Calisher, C. H., Gutiérrez, E., Maness, K. S. & Lord, R. D. Isolation of Mayaro virus from a migrating bird captured in Louisiana in 1967. *Bull. Pan Am. Health Organ.* **8**, 243–248 (1974).

63. de Thoisy, B., Gardon, J., Salas, R. A., Morvan, J. & Kazanji, M. Mayaro virus in wild mammals, French Guiana. *Emerg. Infect. Dis.* **9**, 1326–1329 (2003).

64. Woodall, J. P. Virus research in Amazonia. *Atas do Simpósio sobre a Biota Amazônica (Patologia)* **6**, (1967).

65. Centers for Disease Control (U.S.). *Arthropod-Borne Virus Information Exchange*. (Center for Infectious Diseases, Centers for Disease Control, [Atlanta, Ga.?], 1986).

66. Groot, H., Morales, A. & Vidales, H. Virus isolations from forest mosquitoes in San Vicente de Chucuri, Colombia. *Am. J. Trop. Med. Hyg.* **10**, 397–402 (1961).

67. Aitken, T. H. G., Downs, W. G., Anderson, C. R., Spence, L. & Casals, J. Mayaro Virus Isolated from a Trinidadian Mosquito,*Mansonia venezuelensis*. *Science* **131**, 986–986 (1960).

68. De Sousa, F. B. *et al.* Report of natural Mayaro virus infection in Mansonia humeralis (Dyar & Knab, Diptera: Culicidae). *Parasit. Vectors* **16**, 140 (2023).

69. Mota, M. T. de O., Ribeiro, M. R., Vedovello, D. & Nogueira, M. L. Mayaro virus: A neglected arbovirus of the Americas. *Future Virol.* **10**, 1109–1122 (2015).

70. Ganjian, N. & Riviere-Cinnamond, A. Mayaro virus in Latin America and the Caribbean. *Rev. Panam. Salud Publica* **44**, e14 (2020).

71. Fernández, D. *et al.* Differential Susceptibility of Aedes aegypti and Aedes albopictus Mosquitoes to Infection by Mayaro Virus Strains. *Am. J. Trop. Med. Hyg.* **109**, 115–122 (2023).

72. Diop, F. *et al.* Differential Susceptibility and Innate Immune Response of Aedes aegypti and Aedes albopictus to the Haitian Strain of the Mayaro Virus. *Viruses* **11**, 924 (2019).

73. Serra, O. P., Cardoso, B. F., Ribeiro, A. L. M., Santos, F. A. L. dos & Slhessarenko, R. D. Mayaro virus and dengue virus 1 and 4 natural infection in culicids from Cuiabá, state of Mato Grosso, Brazil. *Mem. Inst. Oswaldo Cruz* **111**, 20–29 (2016).

74. Pinheiro, F. P. *et al.* An outbreak of Mayaro virus disease in Belterra, Brazil. I. Clinical and virological findings. *Am. J. Trop. Med. Hyg.* **30**, 674–681 (1981).

75. Forato, J. *et al.* Molecular epidemiology of Mayaro virus among febrile patients, Roraima State, Brazil, 2018-2021. *Emerg. Infect. Dis.* **30**, 1013–1016 (2024).

76. Blohm, G. *et al.* Mayaro as a Caribbean traveler: Evidence for multiple introductions and transmission of the virus into Haiti. *Int. J. Infect. Dis.* **87**, 151–153 (2019).

77. Samson, D. M. *et al.* New baseline environmental assessment of mosquito ecology in northern Haiti during increased urbanization. *J. Vector Ecol.* **40**, 46–58 (2015).

78. Cooke, S. B., Rosenberger, A. L. & Turvey, S. An extinct monkey from Haiti and the origins of the Greater Antillean primates. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2699–2704 (2011).

79. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).

80. Puhach, O., Meyer, B. & Eckerle, I. SARS-CoV-2 viral load and shedding kinetics. *Nat. Rev. Microbiol.* **21**, 147–161 (2023).

81. Mehta, O. P., Bhandari, P., Raut, A., Kacimi, S. E. O. & Huy, N. T. Coronavirus disease (COVID-19): Comprehensive review of clinical presentation. *Front. Public Health* **8**, 582932 (2020).

82. Yang, Q. *et al.* Just 2% of SARS-CoV-2-positive individuals carry 90% of the virus circulating in communities. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2104547118 (2021).

83. Hughes, T. D. *et al.* The effect of SARS-CoV-2 variant on respiratory features and mortality. *Sci. Rep.* **13**, 4503 (2023).

84. Davis, H. E., McCorkell, L., Vogel, J. M. & Topol, E. J. Long COVID: major findings, mechanisms and recommendations. *Nat. Rev. Microbiol.* **21**, 133–146 (2023).

85. CDC. COVID Data Tracker. *Centers for Disease Control and Prevention* https://covid.cdc.gov/covid-data-tracker/#trends_weeklydeaths_select_00 (2020).

86. Lopez Bernal, J. *et al.* Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. *BMJ* **373**, n1088 (2021).

87. Patel, R., Kaki, M., Potluri, V. S., Kahar, P. & Khanna, D. A comprehensive review of SARS-CoV-2 vaccines: Pfizer, Moderna & Johnson & Johnson. *Hum. Vaccin. Immunother.* **18**, 2002083 (2022).

88. Hernández-Bello, J. *et al.* Neutralizing antibody responses to the Delta variant of SARS-CoV-2 following vaccination with Ad5-nCoV (CanSino) in the Mexican population. *PLoS One* **19**, e0299520 (2024).

89. Ikegame, S. *et al.* Neutralizing activity of Sputnik V vaccine sera against SARS-CoV-2 variants. *Nat. Commun.* **12**, 4598 (2021).

90. Chakraborty, C., Bhattacharya, M. & Dhama, K. SARS-CoV-2 vaccines, vaccine development technologies, and significant efforts in vaccine development during the pandemic: The lessons learned might help to fight against the next pandemic. *Vaccines (Basel)* **11**, (2023).

91. FDA Approves and Authorizes Updated mRNA COVID-19 Vaccines to Better Protect Against Currently Circulating Variants. *U.S. Food and Drug Administration* https://www.fda.gov/news-events/press-announcements/fda-approves-and-authorizes-updated-mrna-covid-19-vaccines-better-protect-against-currently (2024).

92. Know Your Treatment Options for COVID-19. *U.S. Food and Drug Administration* https://www.fda.gov/consumers/consumer-updates/know-your-treatment-options-covid-19 (2025).

93. Family: Coronaviridae. https://ictv.global/report/chapter/coronaviridae/coronaviridae.

94. Su, S. *et al.* Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* **24**, 490–502 (2016).

95. Subfamily: Orthocoronavirinae. https://ictv.global/report/chapter/coronaviridae/coronaviridae/orthocoronavirinae.

96. Esper, F., Weibel, C., Ferguson, D., Landry, M. L. & Kahn, J. S. Coronavirus HKU1 infection in the United States. *Emerg. Infect. Dis.* **12**, 775–779 (2006).

97. Arbour, N., Day, R., Newcombe, J. & Talbot, P. J. Neuroinvasion by human respiratory coronaviruses. *J. Virol.* **74**, 8913–8921 (2000).

98. Taxonomy browser (Sarbecovirus). https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=2509511.

99. Wang, L. F. & Eaton, B. T. Bats, civets and the emergence of SARS. *Curr. Top. Microbiol. Immunol.* **315**, 325–344 (2007).

100. Guan, Y. *et al.* Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276–278 (2003).

101. Menachery, V. D. *et al.* SARS-like WIV1-CoV poised for human emergence. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 3048–3053 (2016).

102. Tu, C., Crameri, G. & Kong, X. Antibodies to SARS-Coronavirus in Civets. *Emerging Infectious Diseases* **10**, 2244–2248 (2004).

103. Hu, B. *et al.* Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* **13**, e1006698 (2017).

104. Jin, Y. *et al.* Genome-wide analysis of the indispensable role of non-structural proteins in the replication of SARS-CoV-2. *Front. Microbiol.* **13**, 907422 (2022).

105. McBride, R., van Zyl, M. & Fielding, B. C. The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6**, 2991–3018 (2014).

106. Nieto-Torres, J. L. *et al.* Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS Pathog.* **10**, e1004077 (2014).

107. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170 (2021).

108. Saud, Z. *et al.* The SARS-CoV2 envelope differs from host cells, exposes procoagulant lipids, and is disrupted in vivo by oral rinses. *J. Lipid Res.* **63**, 100208 (2022).

109. Yang, H. & Rao, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nat. Rev. Microbiol.* **19**, 685–700 (2021).

110. Zhang, Z. *et al.* Structure of SARS-CoV-2 membrane protein essential for virus assembly. *Nat. Commun.* **13**, 4399 (2022).

111. Kreutzberger, A. J. B. *et al.* SARS-CoV-2 requires acidic pH to infect cells. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2209514119 (2022).

112. Wang, J. *et al.* Computational insights into the membrane fusion mechanism of SARS-CoV-2 at the cellular level. *Comput. Struct. Biotechnol. J.* **19**, 5019–5028 (2021).

113. Chen, B., Farzan, M. & Choe, H. SARS-CoV-2 spike protein: structure, viral entry and variants. *Nat. Rev. Microbiol.* **23**, 455–468 (2025).

114. Long, S. SARS-CoV-2 subgenomic RNAs: Characterization, utility, and perspectives. *Viruses* **13**, 1923 (2021).

115. Lamers, M. M. & Haagmans, B. L. SARS-CoV-2 pathogenesis. *Nat. Rev. Microbiol.* **20**, 270–284 (2022).

116. Katiyar, H., Arduini, A., Li, Y. & Liang, C. SARS-CoV-2 assembly: Gaining infectivity and beyond. *Viruses* **16**, 1648 (2024).

117. Jackson, C. B., Farzan, M., Chen, B. & Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* **23**, 3–20 (2022).

118. Ahn, J. H. *et al.* Nasal ciliated cells are primary targets for SARS-CoV-2 replication in the early stage of COVID-19. *J. Clin. Invest.* **131**, (2021).

119. Hou, Y. J. *et al.* SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. *Cell* **182**, 429–446.e14 (2020).

120. Thorne, L. G. *et al.* SARS-CoV-2 sensing by RIG-I and MDA5 links epithelial infection to macrophage inflammation. *EMBO J.* **40**, e107826 (2021).

121. Palestra, F. *et al.* SARS-CoV-2 spike protein activates human lung macrophages. *Int. J. Mol. Sci.* **24**, 3036 (2023).

122. Zheng, M. *et al.* TLR2 senses the SARS-CoV-2 envelope protein to produce inflammatory cytokines. *Nat. Immunol.* **22**, 829–838 (2021).

123. Khan, S. *et al.* SARS-CoV-2 spike protein induces inflammation via TLR2-dependent activation of the NF-κB pathway. *Elife* **10**, e68563 (2021).

124. Domizio, J. D. *et al.* The cGAS-STING pathway drives type I IFN immunopathology in COVID-19. *Nature* **603**, 145–151 (2022).

125. Diamond, M. S. & Kanneganti, T.-D. Innate immunity: the first line of defense against SARS-CoV-2. *Nat. Immunol.* **23**, 165–176 (2022).

126. Keusch, G. T. *et al.* Pandemic origins and a One Health approach to preparedness and prevention: Solutions based on SARS-CoV-2 and other RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2202871119 (2022).

127. Crits-Christoph, A. *et al.* Genetic tracing of market wildlife and viruses at the epicenter of the COVID-19 pandemic. *Cell* **187**, 5468–5482.e11 (2024).

128. Worobey, M. *et al.* The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science* **377**, 951–959 (2022).

129. Lu, J. *et al.* Genomic epidemiology of SARS-CoV-2 in Guangdong province, China. *Cell* **181**, 997–1003.e9 (2020).

130. A Timeline of COVID-19 Developments in 2020. *AJMC* https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020 (2021).

131. CDC. CDC Museum COVID-19 Timeline. *Centers for Disease Control and Prevention* https://www.cdc.gov/museum/timeline/covid19.html (2025).

132. COVID-19 cases. *datadot* https://data.who.int/dashboards/covid19/cases?n=c.

133. Liu, W. *et al.* Evolution of the SARS-CoV-2 Omicron variants: Genetic impact on viral fitness. *Viruses* **16**, (2024).

134. auspice. https://nextstrain.org/ncov/gisaid/global/6m?c=gt-S_456.

135. Jian, F. *et al.* Convergent evolution of SARS-CoV-2 XBB lineages on receptor-binding domain 455-456 synergistically enhances antibody evasion and ACE2 binding. *PLoS Pathog.* **19**, e1011868 (2023).

136. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).

137. Liu, W. J. *et al.* Surveillance of SARS-CoV-2 at the Huanan Seafood Market. *Nature* **631**, 402–408 (2024).

138. Pekar, J. E. *et al.* The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. *Science* **377**, 960–966 (2022).

139. Shi, Z. & Hu, Z. A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res.* **133**, 74–87 (2008).

140. Xiao, K. *et al.* Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* **583**, 286–289 (2020).

141. Lam, T. T.-Y. *et al.* Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285 (2020).

142. Liu, P. *et al.* Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog.* **16**, e1008421 (2020).

143. Xiao, X., Newman, C., Buesching, C. D., Macdonald, D. W. & Zhou, Z.-M. Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic. *Sci. Rep.* **11**, 11898 (2021).

144. MacLean, O. A. *et al.* Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. *PLoS Biol.* **19**, e3001115 (2021).

Chapter 2: Genetic determinants of Mayaro virus adaptation to urban mosquito vectors

The E2 glycoprotein holds key residues for Mayaro virus adaptation to the urban *Aedes aegypti* mosquito

Chelsea Cereghino[1,*], Ferdinand Roesch[2,4,*], Lucía Carrau[2,5], Alexandra Hardy[2], Helder V. Ribeiro-Filho[6], Annabelle Henrion-Lacritick[3], Cassandra Koh[3], Jeffrey M Marano[1,7], Tyler A Bates[1], Pallavi Rai[1], Christina Chuong[1], Shamima Akter[1,8], Thomas Vallet[2], Hervé Blanc[3], Truitt Elliot[9], Anne M Brown[9], Pawel Michalak[10], Tanya LeRoith[1], Jesse D Bloom[11], Rafael Elias Marques[6], Maria-Carla Saleh[3], Marco Vignuzzi[2], James Weger-Lucarelli[1,2]

*These authors contributed equally.

[1]Department of Biomedical Sciences and Pathobiology, VA-MD Regional College of Veterinary Medicine, Virginia Tech, Blacksburg, Virginia, USA.

[2]Institut Pasteur, Viral Populations and Pathogenesis Unit, Centre National de la Recherche Scientifique UMR 3569, Paris, France.

[3]Institut Pasteur, Viruses and RNA Interference Unit, Centre National de la Recherche Scientifique UMR 3569, Paris, France.

[4]UMR 1282 ISP, INRAE Centre Val de Loire, Nouzilly, France

[5]Department of Microbiology, New York University Langone Medical Center, New York, United States of America.

[6]Brazilian Biosciences National Laboratory, Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, SP, Brazil.

[7]Translational Biology, Medicine, and Health Graduate Program, Virginia Tech, Roanoke, Virginia, USA.

[8]Department of Bioinformatics and Computational Biology, School of Systems Biology, George Mason University, Fairfax, VA, USA.

[9]Program in Genetics, Bioinformatics, and Computational Biology (GBCB), Virginia Tech, Blacksburg, VA, USA; Research and Informatics, University Libraries, Virginia Tech, Blacksburg, VA, USA; Department of Biochemistry, Virginia Tech, Blacksburg, VA, USA.

[10]Edward Via College of Osteopathic Medicine, Monroe, LA, USA; Center for One Health Research, VA-MD Regional College of Veterinary Medicine, Blacksburg, VA, USA; Institute of Evolution, University of Haifa, Haifa, Israel.

[11]Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America; Howard Hughes Medical Institute, Chevy Chase, Maryland, United States of America.

Corresponding Author: James Weger-Lucarelli; weger@vt.edu

Abstract

Adaptation to mosquito vectors suited for transmission in urban settings is a major driver in the emergence of arboviruses. To better anticipate future emergence events, it is crucial to assess their potential to adapt to new vector hosts. In this work, we used two different experimental evolution approaches to study the adaptation process of an emerging alphavirus, Mayaro virus (MAYV), to *Ae. aegypti*, an urban mosquito vector of many other arboviruses. We identified E2-T179N as a key mutation increasing MAYV replication in insect cells and enhancing transmission after escaping the midgut of live *Ae. aegypti*. In contrast, this mutation decreased viral replication and binding in human fibroblasts, a primary cellular target of MAYV in humans. We also showed that MAYV E2-T179N generates reduced viremia and displays less severe tissue pathology *in vivo* in a mouse model. We found evidence in mouse fibroblasts that MAYV E2-T179N is less dependent

on the Mxra8 receptor for replication than WT MAYV. Similarly, exogenous expression of human apolipoprotein receptor 2 and Mxra8 enhanced WT MAYV replication compared to MAYV E2-T179N. When this mutation was introduced in the closely related chikungunya virus, which has caused major outbreaks globally in the past two decades, we observed increased replication in both human and insect cells, suggesting E2 position 179 is an important determinant of alphavirus host-adaptation, although in a virus-specific manner. Collectively, these results indicate that adaptation at the T179 residue in MAYV E2 may result in increased vector competence – but coming at the cost of optimal replication in humans – and may represent a first step towards a future emergence event.

Author Summary

Mosquito-borne viruses must replicate in both mosquito and vertebrate hosts to be maintained in nature successfully. When viruses that are typically transmitted by forest dwelling mosquitoes enter urban environments due to deforestation or travel, they must adapt to urban mosquito vectors to transmit effectively. For mosquito-borne viruses, the need to also replicate in a vertebrate host like humans constrains this adaptation process. Towards understanding how the emerging alphavirus, Mayaro virus, might adapt to transmission by the urban mosquito vector, *Ae. aegypti*, we used natural evolution approaches to identify several viral mutations that impacted replication in both mosquito and vertebrate hosts. We show that a single mutation in the receptor binding domain of E2 increased transmission by *Ae. aegypti* after bypassing the midgut infection barrier but simultaneously reduced replication and pathology in a mouse model. Mechanistic studies suggested that this mutation decreases the dependence of MAYV on human Mxra8 and the putative MAYV receptor human ApoER2 during replication. This suggests MAYV with this mutation alone is unlikely to be maintained in a natural transmission cycle between mosquitoes

and humans. Understanding the adaptive potential of emerging viruses is critical to preventing future pandemics.

Introduction

Arthropod-borne viruses (arboviruses) impose significant economic and public health costs worldwide. Zika virus (ZIKV), dengue virus (DENV), chikungunya virus (CHIKV) and West Nile virus (WNV) are now globally established and have caused significant outbreaks. In past decades, they mainly affected developing countries in tropical regions of the world and have often been neglected. Recently, local cases of ZIKV, DENV and CHIKV [1,2] have been detected in Europe and the United States, demonstrating that more temperate areas of the world are endangered by arboviruses, and that preventing arboviral emergence is one of the key public health challenges for years to come.

Mayaro virus (MAYV; Genus *Alphavirus*) was first isolated in Trinidad and Tobago in 1954 where it was associated with cases of mild febrile illness [3]. Since then, MAYV has caused sporadic outbreaks in several countries in South and Central America, including Brazil [4], Mexico [5], Peru [6], French Guiana [7], Bolivia [8], Ecuador [9] and Venezuela [10]. Imported cases have been described in multiple European countries including the Netherlands [11], France [12], Germany [13] and Switzerland [6]. Three genotypes have been described for MAYV: D (widely dispersed), L (limited) and N (new) [14]. However, recombinant strains of MAYV may also circulate in the Amazon basin [15]. Despite the increase in epidemiological data and surveillance in recent years, MAYV circulation is still likely under-estimated. Many factors contribute to this underreporting, including antibody cross-reactivity which complicates serological studies [6]; significant overlap in the areas of distribution of MAYV and CHIKV; and the similarities between

the clinical manifestations caused by MAYV and other arboviruses [16]. Like CHIKV and DENV, MAYV infection induces a febrile illness, with symptoms such as fever, rash, headaches, and nausea [17]. MAYV also induces joint pain in most symptomatic patients [18], which sometimes lasts for several months [19] and may be caused by sustained production of pro-inflammatory cytokines [20,21]. In rare cases, Mayaro fever is associated with more serious clinical outcomes, such as neurological complications, hemorrhagic manifestations or death [22]. Antibodies directed against MAYV have been observed in many animal species beyond humans, including primates, birds, and rodents [4,23] which may act as reservoir species.

MAYV is considered a serious candidate for viral emergence [14,24,25]. While most MAYV outbreaks so far have happened in rural areas in close proximity of tropical [26] or in indigenous communities [27], there is increasing concern that MAYV may escape this sylvatic cycle and become urbanized. Indeed, MAYV circulation has recently been observed in urban areas of Haiti [15]. In such areas, efficient transmission by *Ae. aegypti* would likely increase the risk of a sustained outbreak. Thus, the current epidemiological situation of MAYV is reminiscent of that of CHIKV before it emerged in 2006, when a single mutation in E1 allowed for better adaptation to *Ae. albopictus*, ultimately leading to the CHIKV outbreak in the Indian Ocean region [28].

While it is well established that MAYV can infect forest-dwelling *Haemagogus* mosquitoes in the context of its sylvatic cycle of MAYV [29], it is unclear whether MAYV can be transmitted in nature by mosquito vectors suited for urban transmission. Several laboratory studies have suggested that *Ae. aegypti* [30–32] and *Ae. albopictus* [33,34] mosquitoes can be readily infected by MAYV and may support viral transmission, albeit to different levels. One report also suggested that *Anopheles* mosquitoes may also participate in MAYV transmission [35]. While

these studies suggest that MAYV has the potential to be transmitted by urban vectors, definitive evidence that this happens outside of laboratory settings is still lacking.

In this work, we sought to test the hypothesis that similar to CHIKV pre-emergence, MAYV is not yet fully adapted to *Ae. aegypti* transmission. We used two different *in vitro* evolution approaches to identify the genetic blocks that MAYV must overcome to better adapt to the anthropophilic vector *Ae. aegypti*. Deep mutational scanning and serial passaging approaches both identified T179N as a key mutation in the receptor binding protein, E2, resulting in increased MAYV transmission by *Ae. aegypti* after overcoming an initial infection barrier. This mutation decreased viral replication in human and mouse cells and in mice, and during replication was associated with decreased dependence on human Mxra8 and a putative human MAYV receptor apolipoprotein E receptor 2 (ApoER2), suggesting that adaptation of MAYV to this new vector may come at an evolutionary cost. This may explain why this mutation has not arisen yet in nature. Evolution at this residue may constitute an important first step towards increasing MAYV transmission in *Ae. aegypti*, although other barriers to infection of the mosquito midgut and human cells must be overcome to result in sustained transmission.

Results

Natural evolution and deep mutational scanning enrich for similar high frequency mutations in MAYV. We sought to identify mutations enabling MAYV adaptation to mammalian and insect hosts using two complementary approaches: naturally evolving virus through serial passaging and facilitating evolution through deep mutational scanning (DMS). Serial passaging was performed in BHK-21 (baby hamster kidney fibroblast), Aag2 (*Ae. aegypti* mosquito cells), U4.4 (*Ae. albopictus* mosquito cells), and 4a-3A (*Anopheles gambiae* mosquito cells) cells. Our rationale for using *Anopheles* cells was based on a recent report showing them to be competent for MAYV [35].

We sequenced the virus using Illumina next-generation sequencing (NGS) after passage 1, 5 and 10 (**Fig. 1A**). All variants were below our threshold based on coverage depth and frequency in the passage 1 samples. As expected, most high frequency mutations were observed in the passage 10 samples, and the vast majority were found in the genes encoding for the envelope proteins, specifically E2 (**Fig. 1B-D**). We only present mutations in the envelope proteins since few mutations in other protein-coding sequences were identified. Of note, we observed a consensus change in 3/6 replicates from Aag2 cells (with frequencies ranging from 0.58 to 0.97) at nucleotide position 8918 (**Fig. 1D**), which resulted in a change within E2 at amino acid position 179 from a threonine (T) to either an isoleucine (I) or an asparagine (N). Mutations were also observed at this site for U4.4 cells (**Fig. 1C:** 2/6; range of 0.07-0.89), but not mammalian cells (**Fig. 1B**). Following passage in 4a-3A cells, we observed a consensus level mutation within E2 in 4/6 replicates at amino acid position 232 resulting in a change from histidine (H) to proline (P) (**Fig. S1**). While not the focus of these studies, future studies should study the impact of this mutation on adaptation potential for *Anopheles* mosquitoes. **S1 File** presents all the variants that passed our threshold, along with their frequencies and sequence coverage.

MAYV DMS populations (**Fig. 1E**) were generated for all envelope genes (E3-E2-6K-E1; **Fig. S2A**) in 3 independent populations, as previously described for other viruses [36–38]. DMS is a powerful approach to create a diverse viral population that theoretically contains all possible amino acid substitutions at each site within a given genomic region. Following selection in an environment, deep sequencing is performed, and computational analyses then identify sites that are enriched or decreased. As expected, the initial DMS populations were highly diverse (**Fig. S2B**) as assessed by nucleotide diversity in the DMS and WT MAYV populations. We then passaged the DMS viruses in LLC-MK2 (immunocompetent rhesus monkey kidney epithelial) and

Aag2 cells at a MOI of 0.01 for three passages. Virus titers were relatively stable throughout passaging in LLC-MK2 (**Fig. S2C**) and Aag2 (**Fig. S2D**). Following sequencing, we observed several amino acid changes that were enriched in Aag2- and LLC-MK2-passaged DMS populations (**Fig. 1F-G**). **S2 and S3 files** present the full DMS selection outputs, including mutations enriched and selected against. Interestingly, the top three sites under selection (T179N, T179I, and Q183P) in Aag2 cells were in the E2 protein and were found as consensus changes in a least one of the replicates during natural evolution in Aag2 or U4.4 cells, though T179N and T179I were the most abundant in traditional passaging. The top sites under selection in LLC-MK2 were in E1 (Y118V) and E2 (P240F and N238L). We also passaged the DMS populations in live mice and *Ae. aegypti* mosquitoes (**Fig. S2E-F**); however, we observed only weak signals of selection (**Fig. S3**). For the remaining studies, we will focus on mutations that were identified *in vitro*.

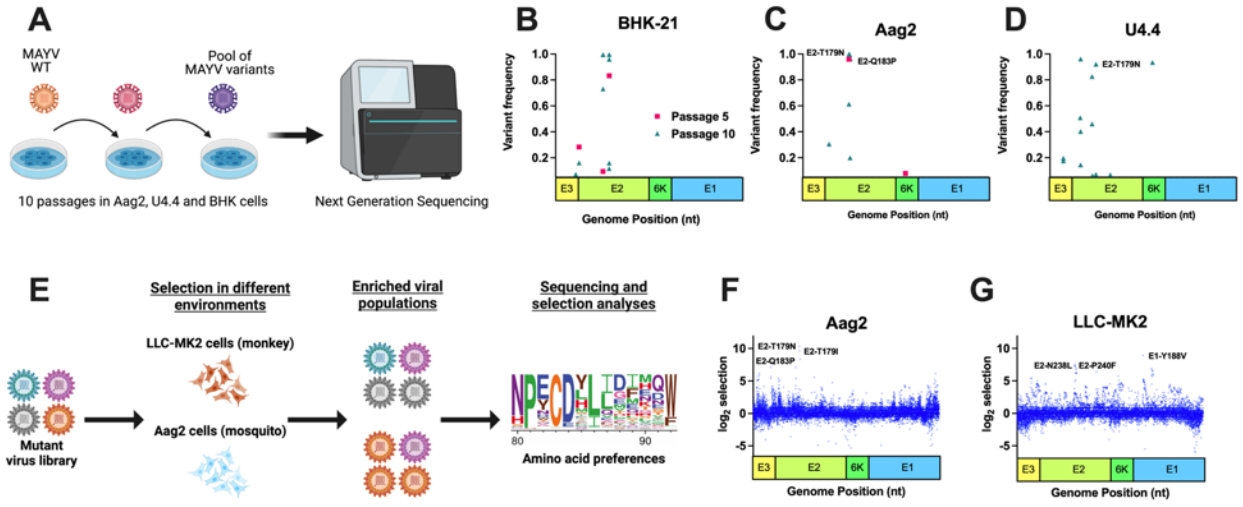


Figure 1: **Natural evolution and deep mutational scanning (DMS) enrich for similar high frequency mutations in Mayaro virus (MAYV).** Experimental evolution was performed using traditional serial passaging and DMS to identify molecular determinants of emergence. **A:** Traditional experimental evolution was performed by serially passaging MAYV at a MOI of 0.01

in BHK-21 cells (hamster) and two insect cell lines, Aag2 (*Ae. aegypti*) and U4.4 (*Ae. albopictus*). A total of ten passages were performed. **B-D:** Following one, five, and ten passages, the viral RNA was sequenced using Illumina NGS to identify potentially adaptive mutations. No high frequency variants were identified following passage one, and as such are not depicted here. **E:** The three MAYV DMS populations, along with WT MAYV, were used to perform three passages in Aag2 (**F**) and LLC-MK2 (**G**) cells. Following passage, the viral RNA was sequenced, and selection analyses were performed to identify enriched variants. The top three variants in each environment, based on selection strength, are presented for each graph. **A and E** were created with BioRender.com whose terms and conditions are at biorender.com/terms.

Adaptive mutations to *Ae. aegypti* cells come at a fitness cost. We next generated several mutants identified in NGS data from both traditional passaging and DMS studies in order to measure the fitness of individual mutations in multiple environments (**Fig. 2A**). The fitness of mutants was measured in a high-throughput fitness screen in several cell lines against a neutral, genetically marked competitor virus, as previously described [39–41]. Two of three mutations that were enriched during Aag2 passaging (E2-T179N, $p<0.0001$; and E2-Q183P, $p=0.0036$) significantly increased viral fitness in Aag2 cells in comparison to WT MAYV (**Fig. 2B**). In U4.4 cells, the fitness of E2-T179N trended higher than WT, although this difference did not reach significance and is considered neutral along with mutation E2-Q183P ($p=0.0624$; **Fig. 2C**). These results suggest that MAYV E2-T179N and, to a lesser extent E2-Q183P, which were identified by two distinct passaging methods, demonstrate increased replication in insect cells but in a species-specific manner. In LLC-MK2 cells, two of the three mutations enriched in LLC-MK2-passaged DMS populations showed significant fitness losses (E1-Y118V, $p<0.0001$; E2-P240F, $p<0.0001$; **Fig. 2D**). E1-Y118 also had lower fitness in MRC-5 human fibroblast cells, ($p<0.0001$; **Fig. 2E**).

We also observed fitness losses for all Aag2-enriched variants in MRC-5 ($p<0.001$ for all comparisons to WT), which we did not observe in LLC-MK2. We then evaluated the fitness of mutations enriched following passage through live mice and mosquitoes and observed mostly modest effects on fitness in Aag2, U4.4, and LLC-MK2 cells (**Fig. S4**). In contrast, 50% of variants enriched in mice and 66% of variants enriched in mosquitoes demonstrated fitness losses in MRC-5 cells (**Fig. S4E**). Notably, the E2-I354D mutant, which was enriched following passage in mice, lost fitness in both mammalian and mosquito environments.

To confirm the effect of the E2-T179N mutation on viral replication, we performed viral growth curves in Aag2, U4.4, and MRC-5 cells. Results were consistent with the competition assays, with a positive effect of E2-T179N on MAYV replication in Aag2 mosquito cells ($p=<0.0001$ at 24h p.i.) and U4.4 mosquito cells ($p=0.0262$ at 1 d.p.i.), coming at the cost of reduced replication in MRC-5 human cells ($p=0.0021$ at 24h p.i.) (**Fig. 3A-C**). No differences in genome:PFU ratios were observed between WT MAYV and MAYV E2-T179N during replication in Aag2 cells (**Fig. S5**).

Next, we assessed whether the E2-T179N mutation similarly impacts CHIKV, a closely related alphavirus that has caused major outbreaks globally in the last decades. Like MAYV, we observed that E2-T179N significantly enhanced CHIKV replication in Aag2 cells (**Fig 3D**; $p=0.0108$ at 120h p.i.). However, in contrast to our MAYV results, we observed an increased replication in MRC-5 cells for the mutant compared to WT CHIKV (**Fig. 3E;** $p=0.0164$ at 48h p.i.). Together, these results indicate that E2 position 179 is a key residue impacting viral replication not only of MAYV but also of other related alphaviruses like CHIKV, albeit with different outcomes in human cells.

Considering the enhanced replication of MAYV E2-T179N in Aag2 cells, we hypothesized E2-T179N, which lies in the receptor binding domain of E2 domain B, may positively impact replication through increased binding of the mutant to Aag2 cells. To test our hypothesis, we performed binding assays of WT MAYV and MAYV E2-T179N with Aag2 cells. We observed significantly reduced binding of MAYV E2-T179N compared to WT (**Fig. S6**; p=0.0005). This result suggested another component of the viral infection cycle may be impacted by E2-T179N such as egress since domain B of E2 plays a role in egress of CHIKV [42]. As such, we interrogated differences in release between WT MAYV and MAYV E2-T179N after the first round of replication at 12 h.p.i. by performing a one-step growth curve in Aag2 cells. Mutant titers were significantly higher than WT MAYV at 12 h.p.i (**Fig. 3F**), and we observed an increase in the mutant titer by 9 h.p.i that was not observed in the WT virus even at 12 h.p.i. These data suggest that the mutant is released earlier from infected cells than WT MAYV.

A

Mutant Name	Method of detection	Phenotype
MAYV WT	N/A	Control
E1-A226V	N/A	Increased CHIKV transmission
E2-T179N	DMS and Serial	Enriched in Aag2
E2-T179I	DMS and Serial	Enriched in Aag2
E2-Q183P	DMS and Serial	Enriched in Aag2
E1-Y118V	DMS	Enriched in LLC-MK2
E2-N238L	DMS	Enriched in LLC-MK2
E2-P240F	DMS	Enriched in LLC-MK2

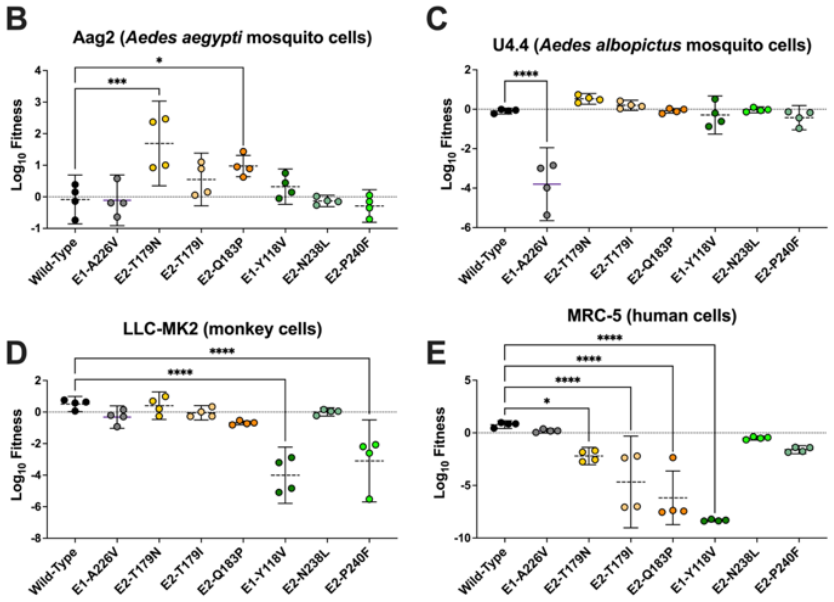


Figure 2. Variants enriched in an *Ae. aegypti* cell line (Aag2) have increased fitness in Aag2 cells but decreased fitness in a human cell line (MRC-5). A: List of the viruses used for the competition assays. The method (serial passaging or DMS) used to identify the mutations, as well

as the phenotypes observed for these different viruses are indicated. **B-E:** Competition assays in mosquito (Aag2 and U4.4) and mammalian (LLC-MK2 and MRC-5) cell lines: Aag2 (**B**)**,** U4.4 (**C**), LLC-MK2 (**D**), and MRC-5 (**E**) cells. Cells were infected at an MOI of 0.01 using a 1:1 ratio based on PFUs for each mutant or WT with a genetically marked MAYV competitor virus. Viral supernatants were harvested at 72h post-infection for Aag2 and 48h post-infection for the other cell lines. Replication of WT and mutant viruses was assessed by RT-qPCR using specific probes labeled with different fluorophores. Log_{10} fitness was calculated by normalizing replication of each virus against a genetically marked reference virus. The mean of 4 independent experiments is represented with standard deviation. Statistical analysis: * = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$; **** = $p<0.0001$ (one-way ANOVA with Dunnett's correction). "Serial" refers to the serial passaging natural evolution experiments to find adaptive mutations.

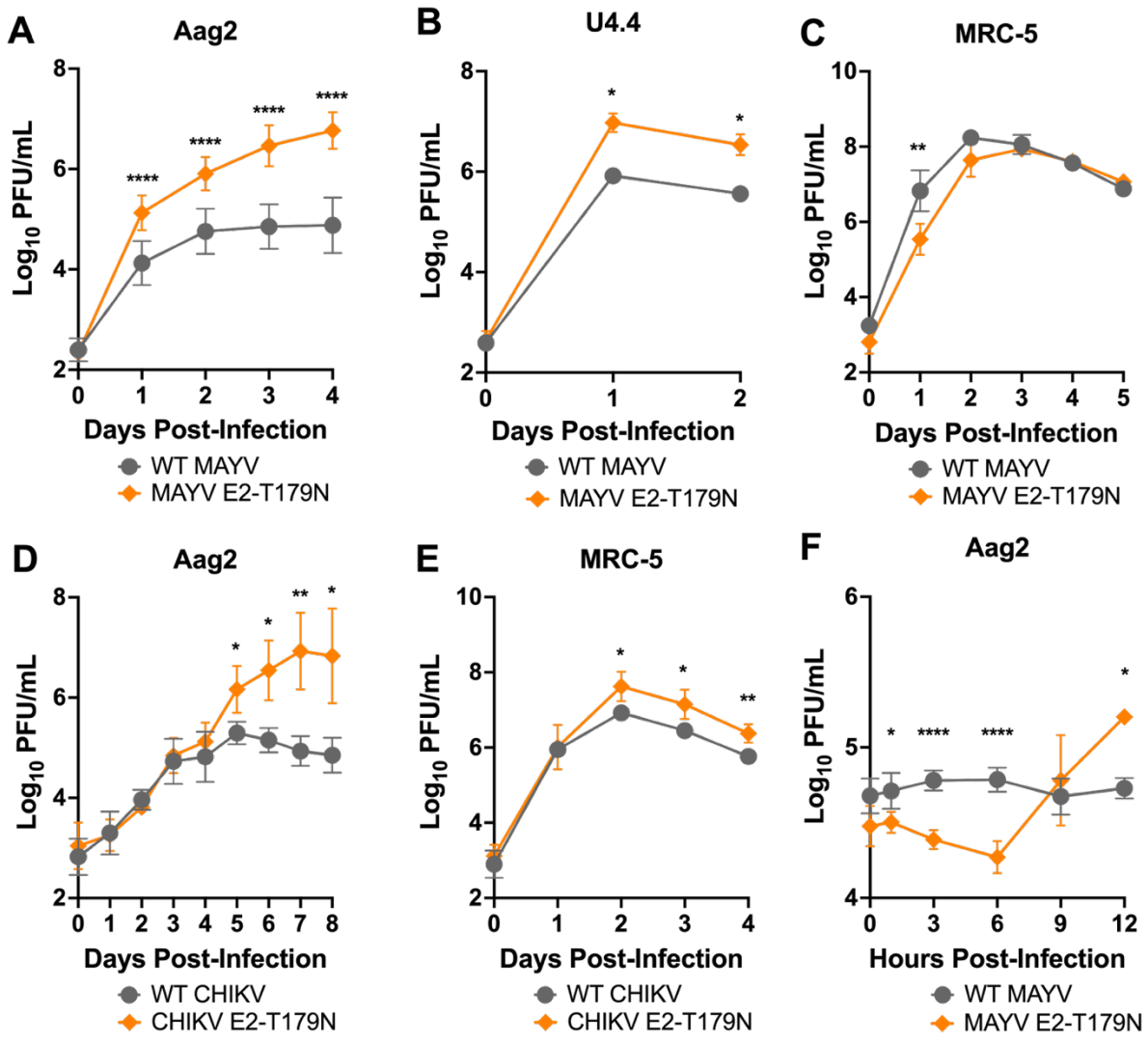


Figure 3. E2-T179N increases viral replication of MAYV and CHIKV in insect cells and causes early release. A-C: Replication of WT MAYV and MAYV E2-T179N in insect and human cells. Aag2 (**A**) and U4.4 (**B**) cells were infected with WT MAYV or MAYV E2-T179N at a MOI of 0.1, while MRC-5 were infected at a MOI of 0.01 (**C**). Replication was assessed over time by plaque assay titration. Data represents the mean of 2 independent experiments apart from the U4.4 growth curve which is representative of three independent experiments. **D-E:** Replication of WT CHIKV and CHIKV E2-T179N in insect and human cells. Aag2 cells were infected with WT CHIKV or CHIKV E2-T179N at a MOI of 0.1 (**D**), while MRC-5 were infected at a MOI of 0.01

(**E**). Replication was assessed over time by plaque assay. Data represents the mean of 2 independent experiments. **F:** One-step growth curve of WT MAYV and MAYV E2-T179N in insect cells. Aag2 cells were infected with WT MAYV and MAYV E2-T179N at an MOI of 10. Replication was assessed over time by plaque assay. Data represent the mean of 2 independent experiments. Statistical analysis: Two-Way ANOVA using Šídák's multiple comparisons; * = $p<0.05$; ** = $p<0.01$; **** = $p<0.0001$. Error bars represent the standard deviation.

E2-T179N negatively impacts MAYV replication in human cells via Mxra8 and ApoER2. While the process by which MAYV enters insect cells remains largely unknown, there is evidence that entry in mammalian cells happens through the endosomal pathway [43] after binding with Mxra8, originally identified as the receptor for CHIKV and several other alphaviruses, including MAYV [44,45]. Thus, we investigated whether the negative effect of E2-T179N on viral replication in human cells (**Fig. 2 and Fig. 3**) is due to decreased viral binding prior to entry and whether binding and replication is altered via the Mxra8 receptor and or other putative MAYV receptors.

To assess viral binding, we inoculated MRC-5 cells with either WT or MAYV E2-T179N at a MOI of 0.1 and incubated the cells at 4°C to allow for binding without internalization of the virus. We then measured the amount of bound virus via RT-qPCR, which showed significantly lower relative binding of MAYV E2-T179N compared to WT MAYV (~4.3 fold less on average, $p<0.0001$, **Fig. 4A**).

To assess the potential impact of E2-T179N on MAYV E2 structure and binding to Mxra8, we used the cryo-EM structure of MAYV (PDB ID: 7KO8 [46]) aligned with the CHIKV structure in complex with Mxra8 (PDB ID: 6JO8 [47]). The E2-T179N change was not expected to have a major impact on the secondary structure of subdomain B or to impair protein-protein contacts

between subdomains in the MAYV E2 (**Fig. S7A**). However, the E2-T179 position is in proximity to the binding site for the Mxra8 receptor (**Fig. S7B**). We thus reasoned that E2-T179N may impair the interaction with Mxra8, which may explain our results on binding to MRC-5 cells. To investigate the E2-T179N-Mxra8 interaction, we conducted several *in silico* analyses. First, we used two N-glycosylation prediction servers, NetNGlyc [48] and NGlycPred [49] to determine whether E2-T179N could create an additional N-glycosylation site via a N-X-S/T motif. Both methods failed to detect any additional N-glycosylation site in E2-T179N. Next, we used molecular modeling and molecular dynamics (MD) simulations with predicted free energy of binding calculations to evaluate if the E2-T179N mutation could affect the interaction between the E2 glycoprotein and the Mxra8 receptor. We obtained a $\Delta\Delta G$ of 0.15 ± 0.66 kcal/mol, indicating that the T179N mutation did not change predicted binding affinity to the Mxra8 D2 domain ($p > 0.05$, one-sample t-test) (**Fig. S7C**). To consider small adjustments in the binding of Mxra8 to MAYV ectodomains, we modeled and performed molecular dynamics simulations in quintuplicate of the full MAYV spike, formed by three pairs of E1 and E2 ectodomains, in complex with 3 copies of the entire Mxra8 ectodomain. We calculated the binding free energy using MMGBSA and computed the difference between the relative binding free energy ($\Delta\Delta G$) of MAYV E2-T179N and WT MAYV spike to Mxra8. The $\Delta\Delta G$ predicted with MM/GBSA was -6.0 ± 26.3 kcal/mol (**Fig. S7D**), which again indicates no significant ($p > 0.05$, one-sample t-test) free energy change associated with the E2-T179N mutation. Our structural analyses failed to detect any differences between binding of WT E2 and E2-T179N to Mxra8.

Given that structural analyses did not detect differences in E2-T179N binding to Mxra8, we sought to validate the results experimentally by performing binding assays in cells expressing Mxra8 and Mxra8 knock-out (KO) cells. WT MAYV and MAYV E2-T179N were inoculated onto

WT 3T3 (mouse fibroblasts) and 3T3 Mxra8 KO cells at a MOI of 0.1, and the cells were incubated at 4°C. Unbound virus was washed away, and virus bound to the cells was quantified by RT-qPCR. We did not detect significantly less binding of MAYV E2-T179N to WT 3T3 cells as compared to WT MAYV, although there was a trend towards decreased binding which is similar to our binding results with MRC-5 cells (**Fig. 4B**). However, both WT MAYV and MAYV E2-T179N bound less to 3T3 Mxra8 KO cells than to WT 3T3 cells. When we compared the reduction in binding to 3T3 Mxra8 KO cells from WT 3T3 cells between both viruses, we found no significant difference (**Fig. 4C**). To assess whether E2-T179N alters replication through Mxra8, we performed growth curves of WT MAYV and MAYV E2-T179N in WT 3T3 and 3T3 Mxra8 KO cells at a MOI of 0.1, measuring infectious virus every 24 hours post-infection. Consistent with previous studies [44], replication of WT MAYV was decreased in 3T3 Mxra8 KO cells compared to WT 3T3 cells at 1 and 2 d.p.i., indicating Mxra8 is important for replication (**Fig. 4D**; p=0.0095, p=0.0007). Similarly, replication of MAYV E2-T179N was reduced at 1 and 2 d.p.i. in 3T3 Mxra8 KO compared to WT 3T3 cells (**Fig. 4E**; p=0.0126, p=0.0485). To compare MAYV E2-T179N to WT MAYV, we calculated the mean change in $\log_{10}$-transformed virus titer between WT MAYV and MAYV E2-T179N in 3T3 Mxra8 KO cells and WT 3T3 cells at 1 d.p.i. WT MAYV had a significantly greater reduction in titer in the absence of Mxra8 compared to the mutant (**Fig. 4F**; p=0.0039). This result suggests the reduced replication of MAYV E2-T179N in human cells is in part explained by its reduced dependence on Mxra8.

We explored other potential receptors that may explain additional mechanisms which attenuate MAYV E2-T179N in mammalian cells by infecting HEK-293A cells transfected with constructs encoding human SLC1A5, hTIM4, VLDLR, and ApoER2 isoforms 1 and 2. SLC1A5 is a component of the CD147 protein complex and was shown to be important for MAYV

replication in HEK-293T cells, and the complex is implicated in the entry of CHIKV and other alphaviruses [50]. TIM4 is a human T-cell immunoglobulin and mucin-domain-containing protein which binds phosphatidylserine and was shown to enhance infection of Eastern equine encephalitis pseudovirus particles [51]. VLDLR and ApoER2 are receptors for many alphaviruses, including Semliki Forest virus, a close relative to MAYV [52]. As such, we hypothesized these putative MAYV receptors could contribute to differences in replication of WT MAYV and MAYV E2-T179N in mammalian cells. We transfected 293A cells with the constructs, including ACE2 and Mxra8 as negative and positive controls, respectively, and infected cells 24 hours later with either WT MAYV or MAYV E2-T179N at a MOI of 0.1. We measured replication by quantifying virus in the supernatant 1 d.p.i. After normalizing the data to replication in HEK-293A cells transfected with ACE2, we found that replication of MAYV E2-T179N was significantly decreased in cells expressing Mxra8 compared to WT MAYV (**Fig. 4G**; p=0.0476). Replication was also significantly increased for WT MAYV and MAYV E2-T179N in cells transfected with the construct encoding hTIM4 (p=0.0031; p=0.01), although there was no significant difference between replication of the two viruses. Compared to MAYV E2-T179N, ApoER2 isoforms 1 and 2, but not VLDLR, increased WT MAYV replication when exogenously expressed (p=0.0016; p=0.0004). These data indicate E2-T179N may negatively impact usage of the putative MAYV receptors ApoER2 isoforms 1 and 2, in addition to Mxra8, in mammalian cells.

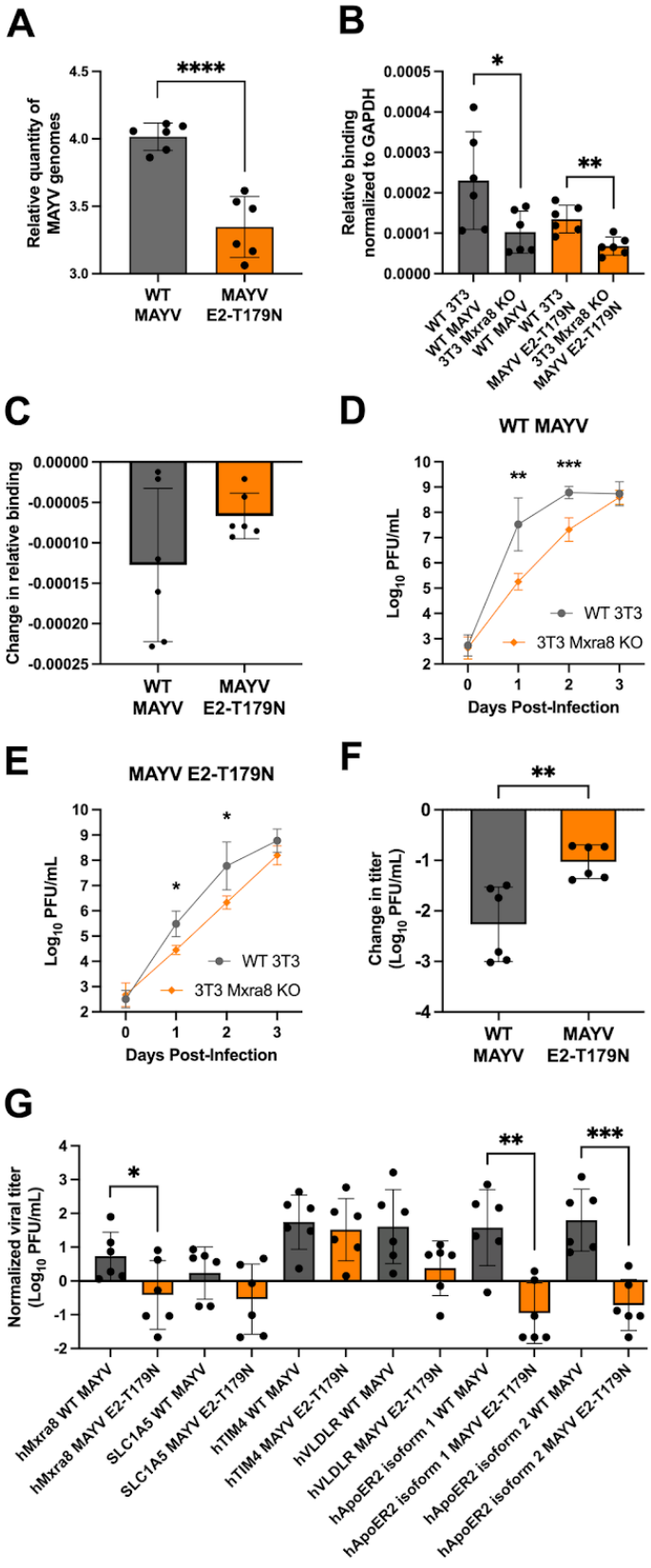

A
Relative quantity of MAYV genomes
4.5
4.0
3.5
3.0
****
WT MAYV
MAYV E2-T179N
B
Relative binding normalized to GAPDH
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
*
**
WT 3T3 WT MAYV
3T3 Mxra8 KO WT MAYV
WT 3T3 MAYV E2-T179N
3T3 Mxra8 KO MAYV E2-T179N
C
Change in relative binding
0.00000
-0.00005
-0.00010
-0.00015
-0.00020
-0.00025
WT MAYV
MAYV E2-T179N
D
WT MAYV
Log10 PFU/mL
10
9
8
7
6
5
4
3
2
**
***
WT 3T3
3T3 Mxra8 KO
0
1
2
3
Days Post-Infection
E
MAYV E2-T179N
Log10 PFU/mL
10
9
8
7
6
5
4
3
2
*
*
WT 3T3
3T3 Mxra8 KO
0
1
2
3
Days Post-Infection
F
Change in titer (Log10 PFU/mL)
0
-1
-2
-3
-4
**
WT MAYV
MAYV E2-T179N
G
Normalized viral titer (Log10 PFU/mL)
4
3
2
1
0
-1
-2
*
**
***
hMxra8 WT MAYV
hMxra8 MAYV E2-T179N
SLC1A5 WT MAYV
SLC1A5 MAYV E2-T179N
hTIM4 WT MAYV
hTIM4 MAYV E2-T179N
hVLDLR WT MAYV
hVLDLR MAYV E2-T179N
hApoER2 isoform 1 WT MAYV
hApoER2 isoform 1 MAYV E2-T179N
hApoER2 isoform 2 WT MAYV
hApoER2 isoform 2 MAYV E2-T179N

Figure 4. Interaction with human receptors Mxra8 and ApoER2 may contribute to the attenuation of MAYV E2-T179N. A: Binding of WT MAYV and MAYV E2-T179N to MRC-5 cells. Pre-chilled MRC-5 cells were infected at a MOI of 0.1, and virus was adsorbed to the cells at 4°C. Unbound virus was washed away, and bound virus was quantified by extracting RNA from the cells and performing RT-qPCR. Relative genome copies of bound virus were determined by normalizing the Ct value of bound virus to the Ct value of the housekeeping gene and virus in the inoculum. Statistical Analysis: unpaired t-test; **** = $p<0.0001$. Data comprise two independent binding assays, each with n=3. The Y-axis is log-transformed for clearer visualization. **B-C:** Binding of WT MAYV and MAYV E2-T179N to WT 3T3 and 3T3 Mxra8 KO cells. Binding assays were performed as above for MRC-5 cells. The change in binding between WT 3T3 and 3T3 Mxra8 KO cells was calculated for each virus (**C**). Statistical Analysis: unpaired t-test; * = $p<0.05$; ** = $p<0.01$. **D-F:** Growth curves of WT MAYV and MAYV E2-T179N in WT 3T3 and 3T3 Mxra8 KO cell lines. WT 3T3 and 3T3 Mxra8 KO cells were infected at a MOI of MOI 0.1 with either WT MAYV (**D**) or MAYV E2-T179N (**E**). Viral titers were quantified each day post-infection by plaque assay. Change in titer for each virus between WT 3T3 and 3T3 Mxra8 KO lines were compared (**F**). Data represent two independent biological replicates performed in triplicate. Statistical Analysis: Two-Way ANOVA with Šídák's multiple comparisons test and unpaired t-test; * =$p<0.05$; ** = $p<0.01$; *** = $p<0.001$. **G:** WT MAYV and MAYV E2-T179N infection in HEK-293A cells transiently expressing endogenous, putative MAYV receptors. Cells were transfected with constructs encoding hMxra8, hACE2, SLC1A5, hTIM4, hVLDLR, and hApoER2 isoforms 1 and 2. Twenty-four hours post-transfection, cells were infected with WT MAYV or MAYV E2-T179N at a MOI of 0.1. Infectious virus in the supernatant 24 h.p.i. was quantified by plaque assay and normalized to infection in cells expressing human ACE2. Data

comprise two independent replicates, each performed in triplicate. Statistical Analysis: unpaired t-test; * = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$. All error bars represent the standard deviation.

E2-T179N increases transmission potential for urban vector *Ae. aegypti* after bypassing initial infection barriers. Given that the E2-T179N mutation increased fitness of MAYV in Aag2 cells, we hypothesized that it may impact MAYV transmission potential in *Ae. aegypti* mosquitoes. To test this hypothesis, we performed five independent transmission studies across two study sites and two populations of *Ae. aegypti*: Kamphaeng Phet and Guerrero. We exposed groups of *Ae. aegypti* to an artificial bloodmeal containing either WT or MAYV E2-T179N. To assess infection rates, we collected either midguts, as this is the first site infection, or bodies lacking legs and wings. In our studies with the Kamphaeng Phet population, both WT and MAYV E2-T179N infected mosquitoes to high rates (**Fig. 5A**), although WT MAYV infection rates were higher than for MAYV E2-T179N ($p=0.035$). In the Guerrero population, WT MAYV infection rates were also significantly higher than MAYV E2-T179N (**Fig. 5B**; $p=0.0035$).

We collected legs and wings to assess the ability of MAYV to disseminate to peripheral organs and did not observe significant differences between WT and MAYV E2-T179N in the Kamphaeng Phet population (**Fig. 5A**; $p=0.093$) but did observe significantly lower dissemination rates for MAYV E2-T179N in the Guerrero population (**Fig. 5B**; $p=0.0034$). In contrast, we observed a significant increase in the proportion of mosquitoes able to transmit MAYV E2-T179N in the Kamphaeng Phet population (**Fig. 5A**; $p<0.0001$), but transmission rates were similar in the Guerrero population (**Fig. 5B**).

Since several transmission barriers exist within the mosquito [53], we calculated dissemination and transmission rates based on the mosquitoes that became infected (**Fig. 5C-F**) or those with a disseminated infection (**Fig. 5G-H**) to understand how each barrier contributes to

transmission of MAYV E2-T179N. The rate of infected mosquitoes that developed a disseminated infection was similar for both viruses (**Fig. 5C-D**), indicating that the viruses had similar ability to disseminate from the midgut after successful infection of the midgut. The transmitting of infected rate was higher for MAYV E2-T179N compared to WT MAYV in the Kamphaeng Phet population (**Fig. 5E**; $p<0.0001$), and the Guerrero population trended toward a similar result (**Fig. 5F**; $p=0.0731$). Finally, we calculated the transmission rates for mosquitoes with a disseminated infection and found significantly higher rates for MAYV E2-T179N than WT MAYV in both the Kamphaeng Phet (**Fig. 5G**; $p<0.0001$) and Guerrero populations (**Fig. 5H**; $p=0.0407$). Our results indicate MAYV E2-T179N is less able to overcome the initial midgut infection barrier; however, once MAYV E2-T179N has disseminated, it is more likely to enter the saliva than WT MAYV.

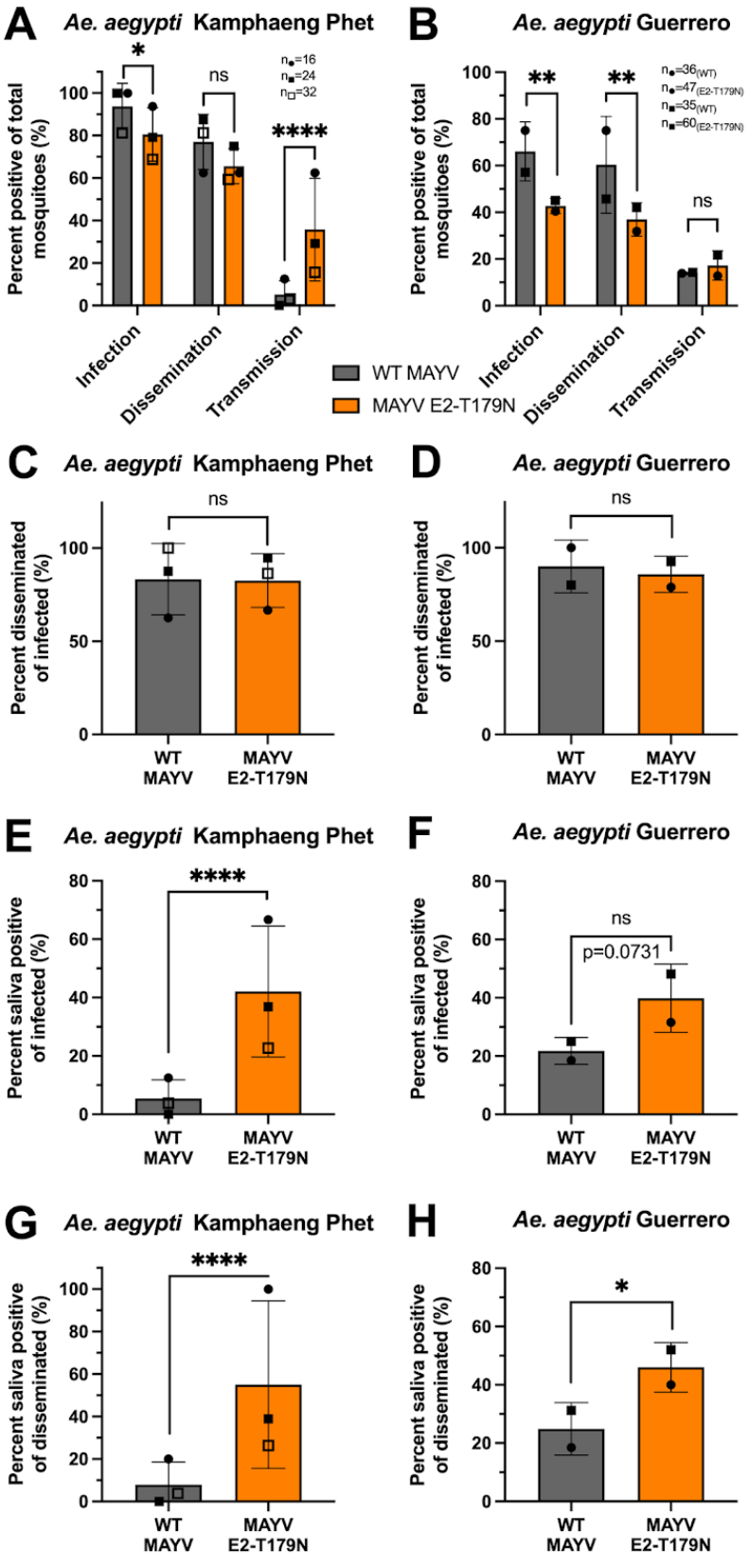

A
Ae. aegypti Kamphaeng Phet
Percent positive of total mosquitoes (%)
*
ns
****
n●=16
n■=24
n□=32
Infection
Dissemination
Transmission
B
Ae. aegypti Guerrero
Percent positive of total mosquitoes (%)
**
**
ns
n●=36(WT)
n●=47(E2-T179N)
n■=35(WT)
n■=60(E2-T179N)
Infection
Dissemination
Transmission
WT MAYV
MAYV E2-T179N
C
Ae. aegypti Kamphaeng Phet
Percent disseminated of infected (%)
ns
WT MAYV
MAYV E2-T179N
D
Ae. aegypti Guerrero
Percent disseminated of infected (%)
ns
WT MAYV
MAYV E2-T179N
E
Ae. aegypti Kamphaeng Phet
Percent saliva positive of infected (%)
****
WT MAYV
MAYV E2-T179N
F
Ae. aegypti Guerrero
Percent saliva positive of infected (%)
ns
p=0.0731
WT MAYV
MAYV E2-T179N
G
Ae. aegypti Kamphaeng Phet
Percent saliva positive of disseminated (%)
****
WT MAYV
MAYV E2-T179N
H
Ae. aegypti Guerrero
Percent saliva positive of disseminated (%)
*
WT MAYV
MAYV E2-T179N

Figure 5. Vector competence of *Ae. aegypti* infected by WT and MAYV E2-T179N. A-B: Adult *Ae. aegypti* Kamphaeng Phet colony mosquitoes were infected with a bloodmeal containing 10^6 PFU/mL (**A**) *Ae. aegypti* Guerrero colony mosquitoes with 10^7 PFU/mL (**B**) of either WT MAYV or MAYV E2-T179N. Fully engorged mosquitoes were kept at 28°C with 80-90% relative humidity, 12hr light:12hr dark, and permanent access to a 10% sucrose solution. Mosquitoes were cold-anesthetized and dissected at 7 d.p.i. (**A**) or 10 d.p.i. (**B**). Viral titers in the midgut (**A**) or bodies (**B**) and legs and wings were quantified by plaque assay titration. During dissection, saliva was harvested from live mosquitoes and viral loads were either amplified on C6/36 or BHK-21 cells (**A**). After 3 days, presence of virus was detected using RT-qPCR or direct visualization of CPEs. For the experiments with *Ae. aegypti* Guerrero mosquitoes, virus in the saliva was quantified by plaque assay, and plaques as little as one in neat saliva were validated by RT-qPCR (**B**). The infection, dissemination, and transmission rates are represented as positive mosquitoes of total bloodfed mosquitoes and span 3 independent experiments (with n=16, n=24, and n=32 per group for the first, second, and third replicates, respectively; n=72 total/group) are represented as positive mosquitoes of total bloodfed mosquitoes (**A**). Infection, dissemination, and transmission rates are also represented as positive mosquitoes of total bloodfed mosquitoes and span 2 independent experiments (with n=36/n=47 and n=35/n=60 for WT/mutant-infected groups for replicates one and two, respectively; total n=71 for WT and total n=107 for mutant) (**B**). **C-D:** Percentage of mosquitoes positive for virus in legs and wings of infected mosquitoes as measured by RT-qPCR on amplified saliva samples (**C**) or plaque assay and RT-qPCR (**D**). **E-F:** Percentage of mosquitoes positive for virus in saliva of infected mosquitoes as measured by RT-qPCR on amplified saliva samples (**E**) or plaque assay and RT-qPCR (**F**). **G-H**: Percentage of mosquitoes positive for virus in saliva of mosquitoes with disseminated virus in legs and wings as measured by RT-qPCR on

amplified saliva samples (**G**) or plaque assay and RT-qPCR (**H**). Statistical analysis: ns = not significant, * = $p<0.05$; ** = $p<0.01$; **** = $p<0.0001$ (Two-tailed Fisher's exact test). Sample sizes for independent experiments are described panels A and B and relate to unique symbols for data points on the graphs. All error bars represent the standard deviation.

MAYV E2-T179N generates reduced viremia and tissue pathology in mice. Since MAYV E2-T179N exhibited increased replication in insect cells but reduced replication in mammalian cells compared to WT MAYV, we sought to explore whether adaptation to *Ae. aegypti* comes at the cost of reduced viral fitness in a mammalian host. Thus, we infected CD-1 mice with WT MAYV and MAYV E2-T179N. Mice infected with MAYV E2-T179N had significantly lower viremia than mice infected with WT MAYV at all timepoints tested (**Fig. 6A**) but mice consistently gained weight over the course of the study in both groups (**Fig. 6B**). Mice infected with MAYV E2-T179N had significantly lower footpad swelling relative to WT MAYV at five days post infection (**Fig. 6C**; $p=0.0006$) and delayed peak swelling, although swelling between the groups was similar. To determine whether MAYV E2-T179N infection caused less tissue damage compared to WT MAYV, we collected the inoculated footpad of each mouse at seven days post-infection and processed them for histological analysis. Based on the severity of lymphoplasmacytic myositis, scores ranging from zero to three were assigned to samples, with zero denoting normal tissue and three the greatest extent of myofiber loss across all samples. Mice infected with MAYV E2-T179N yielded significantly lower footpad scores than those infected with WT MAYV (**Fig. 6D**; $p=0.0150$), indicating lower levels of tissue damage at this timepoint. Images taken from the stained cross-sections of footpads show the degree of muscle fiber degeneration and immune cell infiltration, with this symptom being more severe in the group infected with WT MAYV (**Fig. 6E, panel 1**) than MAYV E2-T179N (**Fig. 6E, panel 2**). The primary infiltrating immune cells were

lymphocytes, and to a lesser extent, plasma cells and macrophages. A representative image from mock-infected mice is presented in **Fig. 6E, panel 3**. These results demonstrate that MAYV E2-T179N replicates to lower levels and causes less severe tissue pathology compared to WT MAYV in the mammalian host.

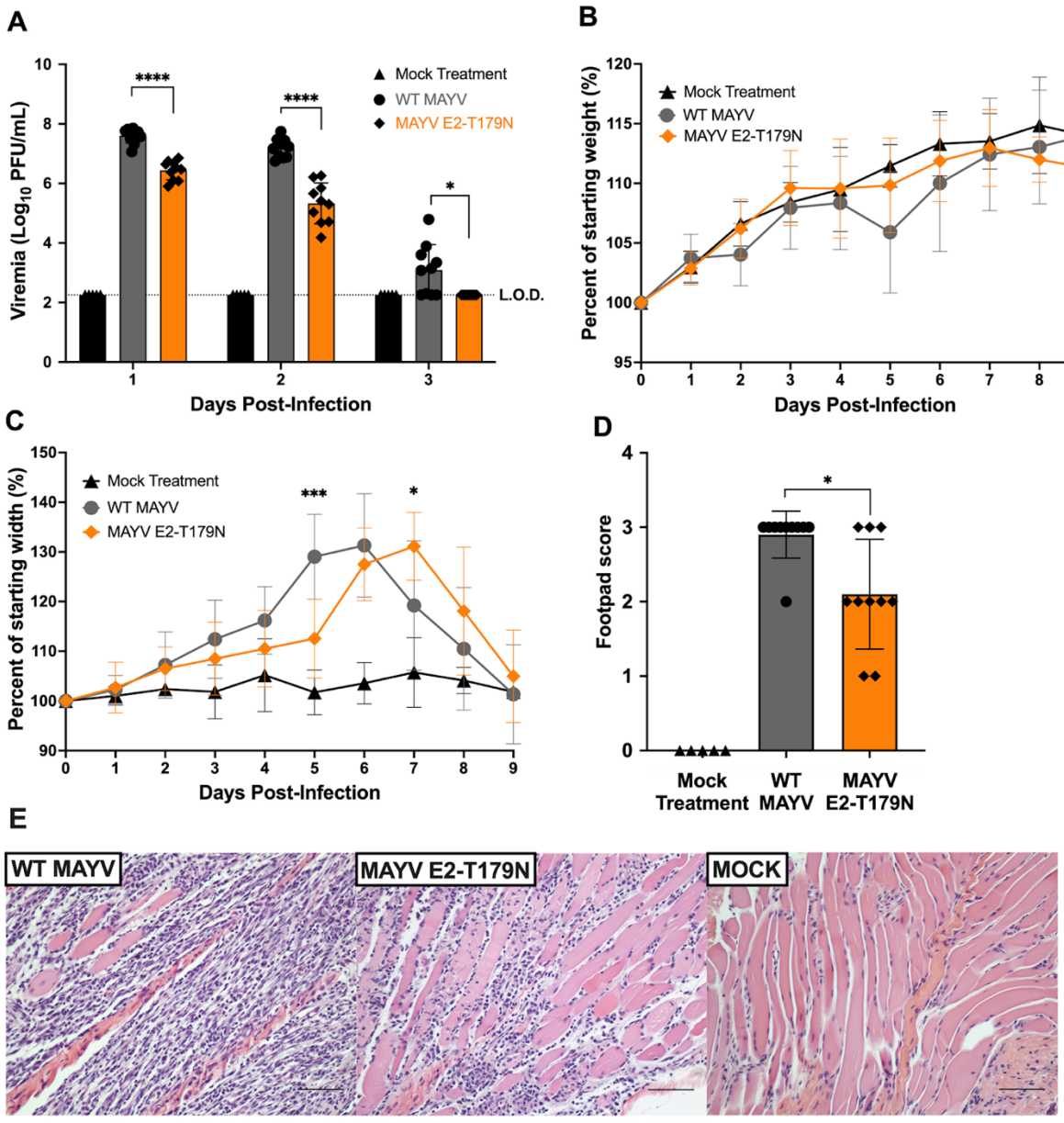


Figure 6. E2-T179N attenuates MAYV replication and tissue pathology in mice. Four-week-old CD-1 mice were infected with 10^5 PFU in 50 µL via injection to the left, hind footpad. **A:** Viral

load in serum was assessed daily by plaque assay. Statistical comparisons were made using multiple unpaired t tests with correction for multiple comparisons using the false discovery rate method of Benjamini, Krieger, and Yekutieli. **B**: Mice were weighed daily following infection to assess weight change. **C:** Footpad swelling was measured daily using a digital caliper. **B-C:** Statistical comparisons were made using a two-way ANOVA with Dunnett's correction for multiple comparisons. Data represent the percent of the weight or footpad width before infection. **D:** Footpad swelling score as determined by histopathology whereby a score of 0 indicates a normal state, 1 mild (<25%) myofiber loss, 2 moderate (25-50%) myofiber loss, and 3 severe (>50%) myofiber loss. Statistical comparisons were made by Mann-Whitney test. **E:** Hematoxylin and eosin stain of muscular tissue in left, hind footpad of mice seven days post infection with WT (left) MAYV E2-T179N (center) or after a mock treatment (right). Footpad scores of 3 and 1 are shown for images of WT MAYV and MAYV E2-T179N, respectively. The scale bar in the images is 200 microns. Studies were performed in two independent experiments, using 10 mice per infected group and five mock-treated mice. Error bars represent standard deviation from the mean.

Discussion

MAYV is an emerging viral threat with epidemic potential. The main goal of this study was to determine whether evolution of MAYV may lead to viral emergence through adaptation to the urban vector *Ae. aegypti*. We identified E2-T179N as a mutation that confers increased replication in *Ae. aegypti* cells (**Fig. 2-3**) and enhances transmission rates after initially overcoming the midgut infection barrier in two distinct populations of *Ae. aegypti* (**Fig. 5**). Thus, we demonstrate MAYV has the potential to adapt to *Ae. aegypti* and enter the urban transmission cycle in some capacity.

Transmission studies of WT MAYV and MAYV E2-T179N in *Ae. aegypti* Kamphaeng Phet and Guerrero populations demonstrated a reduced infection rate for MAYV E2-T179N, however this was paired with an increased transmission rate of mosquitoes with a disseminated infection compared to WT MAYV. However, we observed similar dissemination rates and a trend toward higher transmission potential of the mosquitoes that became infected. These data suggest E2-T179N has an initial fitness reduction at the midgut which is later overcome as the virus either infects and or escapes the salivary gland. The mechanism by which this occurs should be explored further by investigating differences in receptor populations across mosquito tissue types, and the translatability of these data to other alphaviruses such as CHIKV should also be explored.

The same E2-T179N mutation caused a fitness reduction in human cells (**Fig. 2-3**) and mice (**Fig. 6**), suggesting that this mutation alone would not favor a complete, urban MAYV transmission cycle. Other roadblocks remain in the way of MAYV emergence: notably, it is still unclear whether the viremia induced by MAYV in humans is sufficient to sustain human-to-human amplification. With reduced viremia in mice (**Fig. 6A**), it is even less likely MAYV E2-T179N generates enough viremia for human transmission. Serum viral loads of 10^4-10^5 PFU/mL MAYV have been observed in a study involving 21 humans with confirmed MAYV infection during the three days following symptom onset [4]. In comparison, infectious titers for CHIKV range from $10^{3.9}$ to $10^{6.8}$ PFU/mL [54]. Thus, titers in humans appear to be slightly lower for MAYV than the successful human pathogen, CHIKV, which may limit emergence potential without further adaptation. Furthermore, the geographical overlap between MAYV and CHIKV may prevent MAYV emergence, as studies in mice have suggested that a prior CHIKV infection confers cross-protection against MAYV infection [55,56].

The two different experimental evolution approaches we used (serial passaging and DMS) led to the identification of E2-T179N as a key mutation enhancing MAYV replication in *Ae. aegypti* and *Ae. albopictus* cells. Notably, we observed that only three serial passages of DMS viruses in Aag2 cells allowed identification of this adaptive mutation (**Fig. 1F**), while we observed this mutation in only 50% of the replicates in our serial passaging experiment after ten passages (**Fig. 1D**). This confirms that DMS passaging speeds up the process of selection, although DMS can only be targeted to a portion of the viral genome. Thus, we conclude that combining these two approaches represents an efficient strategy to identify adaptive mutations with high confidence. Perhaps unsurprisingly, most mutations detected with both approaches occurred within the receptor-binding protein (E2), highlighting the importance of this protein for host range. Furthermore, the residues with the strongest impact on fitness (positions 179 and 183 in E2) both reside within domain B, which we previously showed was critical for cell tropism [57]. A surprising finding in our DMS studies was that variants enriched during passage in LLC-MK2 either had neutral or deleterious impacts on fitness in mammalian cells. However, these data are consistent with results observed in ZIKV [38] where mammalian fitness gains were only observed with a double mutant. A possible explanation for this is the importance of cooperative interactions with other residues (either on the same genome or on other genomes) occurring within the viral population. Moreover, a mutation that was slightly enriched during mouse passage (E2-I354D) decreased fitness appreciably in both mammalian and insect cells; thus, this mutation represents a broadly attenuating mutation and suggests that DMS studies may be useful for developing rationally designed attenuated vaccine viruses. For example, in future studies, selecting the residues most differentially selected against may lead to a highly attenuated virus and could help increase understanding of fundamental virus biology using a safer loss-of-function approach.

Given the reduced fitness, replication, and binding of MAYV E2-T179N to human fibroblasts MRC-5, we sought to explore known and putative human MAYV receptors which may be mediating these differences. Binding assays to WT 3T3 and 3T3 Mxra8 KO cells showed no difference in the reduction in binding posed by the loss of Mxra8 between the viruses, however, there was a larger reduction in replication of WT MAYV than MAYV E2-T179N in 3T3 Mxra8 KO cells. We conclude the human MAYV receptor Mxra8 contributes to the replication attenuation of MAYV E2-T179N in mammalian cells.

To date, there is no evidence TIM-family receptors, VLDLR, and ApoER2 are used specifically by MAYV. We tested a panel of putative MAYV receptors by infecting HEK-293A cells transiently expressing exogenous human TIM4, VLDLR, and ApoER2 isoforms 1 and 2 with WT MAYV and found increased replication in the presence of these host factors. These results warrant further studies to elucidate the precise role they play in infection. MAYV E2-T179N displayed significantly decreased replication than WT MAYV in the presence of exogenous Mxra8 and ApoER2 isoforms 1 and 2 (**Fig. 4G**). While there may be other factors responsible for the attenuation of MAYV E2-T179N such as binding to heparan sulfate [58], these data suggest that MAYV E2-T179N's attenuation may in part be mediated through interactions with these host factors.

While we showed E2-T179N negatively impacts binding and viral replication in MRC-5 cells, and replication through the human receptors Mxra8 and ApoER2, the mechanisms underlying the impact of E2-T179N on binding and replication in Aag2 cells are still unclear. The significantly reduced binding of MAYV E2-T179N to Aag2 cells was unexpected given its increased replication in these cells. We hypothesize E2-T179N compensates reduced binding via increased fitness later in the viral life cycle. Our data showed MAYV E2-T179N is released earlier

than WT MAYV in Aag2 cells (**Fig. 3F**), suggesting there may be differences in egress or some other stage during replication. This may be through interactions with other viral proteins; for example, it is known that the cytoplasmic domain of E2 of alphaviruses such as Sindbis virus interact with the nucleocapsid core to promote budding [59]. Structural analyses of E2-T179N in complex with MAYV nucleocapsid could elucidate changes to E2 that may positively impact encapsidation and budding efficiency. Other possible explanations for the phenotypes observed are disrupted E1:E2 interactions, since this residue resides in domain B, which interacts with E1 and could disrupt endosomal fusion [60]. Future studies will help shed light on the mechanisms responsible for the phenotypes observed.

Alphaviruses, like other arboviruses, navigate evolutionary trade-offs because of their dual-host nature. Our studies exemplified this concept: MAYV cannot optimize its replication in both insect vectors and mammalian hosts simply through the E2-T179N mutation, possibly due to the need to use different receptors in each host or even that a mutation in one environment may increase replication through efficient egress while simultaneously negatively impacting receptor binding in another host. Mutations well-suited for increased fitness in one part of the viral infection cycle may negatively influence fitness in another part of the cycle or within or between hosts. However, it is possible that additional mutations in E1 or E2 may compensate for this defect, and/or that evolution at the T179 towards other residues may confer increased replication in both hosts. Interestingly, we observed that two distinct mutations (T179N and T179I) were enriched in mosquito cells. Therefore, testing other T179 mutation combinations may provide additional insights into MAYV viral evolution potential. A limitation of this work is our use of a single strain and genotype of MAYV, and future studies should assess the impact of E2 T179N on other MAYV genotypes. Importantly, previous reports have shown a viral strain-dependent impact of the

CHIKV A226V mutation that enhances *Ae. albopictus* transmission for the ECSA genotype, but not the Asian genotype [61]. To that end, we found that E2-T179N CHIKV mutant similarly displayed increased fitness in *Ae. aegypti* cells (**Fig. 3C**), but, in contrast to MAYV, also showed increased fitness in human fibroblasts (**Fig. 3D**). Thus, this residue is important for fitness for both viruses but the differential impact on fitness suggests different mechanisms or epistatic interactions. This result is consistent with the effect we observed with the E1-A226V MAYV mutant, which is known to enhance CHIKV's fitness in *Ae. albopictus* [62] but greatly reduced MAYV's fitness in U4.4 cells (**Fig. 2C**). Finally, given that E2-T179N in MAYV did not alter fitness in monkey cells (**Fig. 2B-C**), we cannot conclude that this is a mammalian- or insect-specific phenotype; rather, it appears to be species-specific, likely dependent on the presence of various entry factors and receptors.

In addition, the E2-T179N mutation led to reduced viral replication and tissue damage *in vivo* in a mouse model, which was consistent with the attenuation of MAYV E2-T179N in human and mouse fibroblasts (**Fig. 3 and 6**). Future work will establish whether reduced tissue pathology of MAYV-induced disease is solely due to lower viral replication, or to other immune parameters such as inflammation, cytokine response, IFN signaling, etc. It will also be important to study why no differences in footpad swelling were observed despite differences in tissue damage, which may shed light on the inflammatory processes induced by alphaviruses. Furthermore, future studies will assess whether viremia levels in mice are still sufficient to cause productive infection in mosquitoes. Our work suggests that mutations in the E2 of MAYV may favor its dissemination through the usage of alternate mosquito vectors and should therefore be monitored closely with appropriate surveillance programs.

Materials and Methods

Cells, viruses and plasmids. Vero, BHK-21, LLC-MK2, and MRC-5 cells were grown in Dulbecco's modified Eagle's medium (DMEM, Gibco), containing 10% fetal calf serum (FBS; Gibco), 1% penicillin/streptomycin (P/S; Thermo Fisher) in a humidified atmosphere at 37°C with 5% CO_2. U4.4 and Aag2 cells were maintained in Leibovitz's L-15 medium (Gibco) with 10% FBS, 1% P/S, 1% non-essential amino acids (Sigma) and 1% tryptose phosphate (Sigma) in a dry atmosphere without CO_2 at 28°C. 3T3 WT and Mxra8 KO cells were a kind gift of Dr. Mike Diamond and were maintained in a humidified atmosphere at 37°C with 5% CO_2 in DMEM (Genesee Scientific) supplemented with 10% fetal bovine serum (R&D Systems), 1% non-essential amino acids, 50 µg/mL gentamicin sulfate, and 25 mM HEPES. HEK-293A cells were grown in DMEM (Genesee Scientific) containing 5% FBS (R&D Systems), 1% non-essential amino acids, 50 µg/mL gentamicin sulfate, and 25 mM HEPES in a humidified atmosphere at 37°C with 5% CO_2 For some studies, Aag2 were maintained in Schneider's insect medium (Genesee Scientific) with 7% FBS, 1% non-essential amino acids, 2.5 µg/µL Amphotericin B, 50 µg/mL gentamicin sulfate, and 5% tryptose phosphate broth in a humified atmosphere without CO_2 at 28°C. MAYV strain TRVL 4675 was derived from an infectious clone we previously described that contains an SP6 promoter to generate infectious RNA [63]. To generate the MAYV infectious clones under the CMV promoter, we amplified the entire MAYV genome from the previously described SP6-containing plasmid along with the pcDNA3.1-based CMV vector [41] containing overlapping ends. All clone sequences were confirmed using next generation sequencing (NGS). The generation of the infectious molecular clone for CHIKV strain SL-CK1 was described previously [41]. The expression construct encoding hTIM4 was provided by H. Choe [51], constructs for hVLDLR, hApoER2 isoform 1, hApoER2 isoform 2, and hMxra8 were provided by

J. Abraham [52], and constructs for expression of hACE2 receptor were provided by G. Larson from the United States Centers for Disease Control and Prevention. The pLX304 vector containing SLC1A5 was obtained from the Arizona State University DNASU Plasmid Repository.

Viral stocks. To generate MAYV viral stocks for the serial passaging experiments, a T25 flask of BHK-21 cells at 75% confluency was transfected with 10 µg of MAYV CMV-promoter driven plasmid using TransIT-LT1 Transfection Reagent (Mirus) according to the manufacturer's instructions. Viral supernatants were collected 48h later and used to perform one blind passage on C6/36 cells. The high titer viral stock used for viral growth curves and mosquito experiments was generated the same way, except that the virus was passaged twice on Vero cells (MOI 0.01). The generation of MAYV and CHIKV mutant stocks was performed by transfecting 500 ng of DNA into a 24-well plate well using JetOptimus (Polyplus, France). The mutants were then passaged once in BHK-21 cells at a MOI of 0.01. Viral stocks were titrated by plaque assay.

Plaque assay. Vero cells were seeded in 24-well plates (10^5 cells per well) and infected with serial dilutions of infectious supernatant diluted in RPMI-1640 with 2% FBS and 10 mM HEPES (viral diluent). After 1h at 37°C, a semi-solid overlay consisting of 1.5% methylcellulose, 2X EMEM, 4X L-glutamine, 0.735% sodium bicarbonate, 0.2 mg/mL gentamicin sulfate, 4% heat-inactivated FBS, and 20 mM HEPES was added to the cells. Cells were fixed with 4% formalin solution, pH 6.2 at 3 or 4 d.p.i. and stained with 0.1% crystal violet.

Serial passaging. We infected cells at a MOI of 0.01 as indicated above for viral growth curves, except infections were carried out with 6 replicates instead of 3. Viral supernatants were collected (2 days p.i. for BHK-21 cells; 3 days p.i. for U4.4 and Aag2 cells) aliquoted, and frozen. After

each passage, all supernatants were titrated by plaque assay. Supernatants collected at passages 1, 5 and 10 were sequenced using NGS as described below.

NGS for serial passaged samples. RNA of 100 μL of each sample supernatant was extracted using TRIzol reagent (Invitrogen) following the manufacturer's protocol. RNA was eluted in 30 μL of nuclease-free water. After quantification using Quant-IT RNA assay kit (Thermo Fisher Scientific), viral RNA was enriched using polyA selection NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). Libraries were prepared with NEBNext Ultra II RNA Library Prep Kit for Illumina. The quality of the libraries was verified using a High Sensitivity DNA Chip (Agilent) and quantified using the Quant-IT DNA assay kit (Thermo Fisher Scientific). Sequencing of the libraries was performed on a NextSeq 500 sequencer (Illumina) with a NextSeq Mid Output kit v2 (151 cycles).

Generation of MAYV envelope DMS populations. DMS populations were created in the genes encoding E3, E2, 6K, and E1 using the SP6-driven MAYV infectious clone as previously described [36–38] with modifications. Notably, we used a bacteria-free cloning approach that we have previously described [64,65]. DMS mutagenesis primers are listed in **S4 file**. The forward mutagenesis primers pool was used with a corresponding MAYV reverse end primer (5' CCCGCATTACACGGTACTTATGAT 3') The reverse mutagenesis primers pool was used with a corresponding MAYV forward end primer (5' CGGAAGGCACAGAGGAGTGG 3'). We performed one round of mutagenesis with ten PCR cycles. All PCRs were performed with SuperFi II PCR master mix (Invitrogen). The fragments were then joined by PCR. The vector containing the remaining MAYV genome was amplified to create overlapping ends with the mutagenized fragments that are compatible with Gibson assembly. The primer sequences used to amplify the

vector are forward (5' CCACTCCTCTGTGCCTTCCG 3') and reverse (5' ATCATAAGTACCGTGTAATGCGGG 3'). Amplicons were run on a 0.6% agarose gel containing GelGreen nucleic acid stain, excised, and then purified using the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel). Amplicons were then assembled (1:1 insert:vector molar ratio) into circular molecules using the NEBuilder HiFi DNA Assembly Master Mix incubated at 50°C for two hours. To confirm that no parental plasmid vector was carried through the process, for each mutant, we included a control containing the DNA fragments but no assembly mix; this was then treated identically to the other samples for the remainder of the process. The assembly was then digested with exonuclease I, lambda exonuclease, and DpnI (all from NEB) to remove single-stranded DNA, double-stranded DNA, and bacterial-derived plasmid DNA, respectively. This product was then amplified by rolling circle amplification (RCA) using the Repli-g mini kit (Qiagen). The RCA product was linearized with SgrAI (NEB) and purified. Capped RNA was generated using the mMESSAGE mMACHINE SP6 kit (Invitrogen) and then transfected into a T150 flask of BHK-21 cells using JetMessenger (Polyplus). Viral supernatants were harvested 48h later, clarified by centrifugation, and then precipitated using polyethylene glycol (PEG) 8000 [66]. We constructed three replicate libraries by performing all the steps independently for each replicate starting with individual PCR reactions.

DMS passaging. Passaging *in vitro* was performed in Aag2 and LLC-MK2 at a MOI of 0.01 for a total of three passages. Following one passage, the virus was titrated by plaque assay and the next passage was initiated at the same MOI. Only one passage was performed for *in vivo* studies. Three week-old female C57BL/6J mice (Jackson laboratory) were inoculated with 10^6 PFU of WT MAYV or DMS populations in the left hind footpad. The mice were bled on days one and two post-infection to measure viremia and collect virus for sequencing. *Ae. aegypti* females originally

collected in Guerrero (Mexico) were exposed to an infectious bloodmeal containing $3x10^7$ PFU/mL. Fully bloodfed mosquitoes were separated and allowed to incubate at 28°C for ten days. Following incubation, salivary secretions were collected in viral diluent containing RPMI-1640 media supplemented with 25 mM HEPES, 1% BSA, 50 µg/mL gentamicin, and 2.5 µg/mL amphotericin B. We collected only saliva samples because this best represents the virus with potential to transmit to the mammalian host.

DMS sequencing and analysis. Viral RNA for sequencing was extracted using the Direct-zol RNA extraction kit (Zymo Research). RNA was converted to cDNA using the Maxima H Minus cDNA Synthesis Master Mix (Invitrogen). PCR amplicons containing the DMS region were amplified using SuperFi II master mix with MAYV specific primers: forward (5' AGTGGGTAAGCCTGGCGACA 3') and reverse (5' TAAATCGGTCCGCATCATGCAC 3'). The fragments were purified using a 1x ratio of AMPure XP (Beckman). Libraries were prepared using Nextera XT and sequenced on an Illumina NextSeq 500. Data was analyzed using the dms_tools2 software, which is available at https://jbloomlab.github.io/dms_tools2/. Differential selection was calculated by comparing the starting virus stock to the post-passage samples.

NGS data analysis. Raw sequencing data was deposited under the Sequence Read Archive (SRA, NCBI) under bioproject number PRJNA796940. The raw NGS reads were first trimmed using BBDuk, a tool from the BBMap tool kit, to remove the sequencing adapters and reads with a quality score of less than 30 [67]. The sequences were then aligned to the MAYV genome using the Burrow Wheeler Aligner (BWA) tool with no added parameters [68]. The resulting .bam files were then sorted using Sambamba [69]. Indels and variants were called using LoFreq [70] and then filtered to remove variants present at less than 5% frequency. We then used snpdat to annotate

the variants [71]. A variant threshold was calculated by comparing the coverage depth to the allele frequency. Specifically, we took the reciprocal frequency of a given variant and multiplied by 10; if the sequencing coverage at this site was greater than that number, it was considered positive; otherwise, it was discarded [72].

Site-directed mutagenesis and virus rescue. All mutants were created using our bacteria-free cloning approach described above with some modifications. The CMV-promoter-driven plasmids for MAYV and CHIKV were used to make mutants. Mutations were incorporated into PCR primers that created 20-30 bp overlaps for Gibson assembly; all primer sequences are available upon request. PCR fragments were amplified using Platinum SuperFi II PCR master mix (Invitrogen). The assembled product was amplified by RCA with the FemtoPhi DNA amplification kit (Evomic Science). Virus rescue was performed by transfecting RCA products directly into BHK-21 cells using JetOptimus (Polyplus), as we have previously described [73]. Viral titers were measured by plaque assay. To sequence the viral genome, we extracted RNA, performed DNase treatment to remove residual plasmid DNA, and generated cDNA using Maxima RT (ThermoFisher). PCR amplicons were generated using SuperFi II PCR master mix. Following gel or PCR purification, amplicons were submitted for Sanger sequencing at the Virginia Tech Genomics Sequencing Center.

Viral growth curves. One or two days prior to infection, cells were seeded in a 24-well plate and infected at 60-80% confluency. Cells were infected in triplicate at a MOI of 0.01 for all cells except Aag2, WT 3T3, and 3T3 Mxra8 KO cells, which were infected at MOI 0.1. Virus was diluted in RPMI-1640 (Genesee Scientific) media containing 2% FBS and 10 mM HEPES. One hour post-infection, viral inoculum was removed, cells were washed once with PBS, and the appropriate

culture media was added to the cells. Cells were incubated at either 37°C or 28°C, according to cell type and conditions listed above. Supernatant was collected every 24 hours for the indicated time points and frozen until viral titration. One-step growth curves were performed by infecting Aag2 cells at a MOI of 10. Cells were chilled at 4°C for 30 minutes and adsorption was performed at 4°C for 30 minutes to synchronize infection. Unbound virus was washed away with cold PBS six times, and internalization and infection were initiated by adding media to cells and incubating at 28°C. Supernatants were harvested every 3 hours post-infection, and all infectious virus was quantified by plaque assay.

Genome:PFU ratio. A viral growth curve was performed in Aag2 cells, as above, with a MOI of 0.1 for both WT MAYV and MAYV E2-T179N. Supernatant was harvested each day post infection and titrated by plaque assay. Viral genomes were quantified each day post infection via RT-qPCR on a Bio-Rad CFX96 Touch Real-Time PCR Detection System. Briefly, supernatant was diluted 1:5 in nuclease-free water before performing reverse transcription using the New England Biolabs Universal One-Step RT-qPCR kit followed by SYBR green-based qPCR using the following primers: MAYV 5028For. (5' CCTCTGTTAGTCCTGTGCAATAC 3'); MAYV 5108Rev. (5' AAGGTGCTTAGGGAGCTACT 3'). A standard curve was generated using full-length MAYV genomic RNA derived from a SP6-containing MAYV infectious clone [73] to calculate MAYV genomes per milliliter of supernatant, and the genome:PFU ratio was calculated by dividing the genome concentration by the PFUs per milliliter of supernatant achieved via plaque assay.

Competition assays. *In vitro* competition assays were performed against a genetically marked reference virus essentially as previously described [39–41]. Briefly, the competitor and test virus

(in this case WT or mutant MAYV) were mixed at an equal PFU:PFU ratio and then used to infect cells at a MOI of 0.01. Virus-containing supernatant was harvested at two days p.i. for all cell lines used: Aag2, U4.4, MRC-5 and LLC-MK2. The proportion of each virus was measured using TaqMan RT-qPCR with probes specific to each virus containing ZEN / Iowa Black FQ quenchers. The primers used are as follows: MAYV 5028For. (5' CCTCTGTTAGTCCTGTGCAATAC 3'); MAYV 5108Rev. (5' AAGGTGCTTAGGGAGCTACT 3'); WT probe (5' FAM CACAGTGAAACTACTGTAAGCTTGAGCTCG 3'); Competitor probe (5' JOE CACAGTGAAACTACTGTttcCcTttcCTCG 3'). The genome copies of each virus in a given sample were calculated using a standard curve of viral RNA derived from the SP6 promoter-based clones for both WT and the marked reference virus. The relative fitness was determined with methods previously described [39,74] using the genome copies for each virus. Briefly, the formula $W = [R(t)/R(0)]^{1/t}$ represents the fitness (W) of the mutant genotype relative to the common competitor virus, where *R(0)* and *R(t)* represent the ratio of mutant to competitor virus in the inoculation mixture and at *t* days post-inoculation, respectively.

Binding assays in MRC-5 and Aag2 cells. Pre-chilled MRC-5 or Aag2 cells were washed with 4°C PBS containing 5% w/v BSA prior to inoculation with WT MAYV or MAYV E2-T179N at a MOI of 0.1. Adsorption of the virus proceeded for 30 minutes at 4°C. Inoculum was removed, and cells were washed six times with the 4°C PBS solution before RNA extraction. New England Biolabs Universal One-Step RT-qPCR kit was used to prepare cDNA from total RNA. MAYV genomes in the total RNA were quantified through SYBR green-based RT-qPCR on a Bio-Rad CFX96 Touch Real-Time PCR Detection System using the following primers: MAYV 5028For. (5' CCTCTGTTAGTCCTGTGCAATAC 3'); MAYV 5108Rev. (5' AAGGTGCTTAGGGAGCTACT 3'); GAPDHFor. (5' CCAGGTGGTCTCCTCTGACTT 3');

GAPDHRev. (5’ GTTGCTGTAGCCAAATTCGTTGT 3’); rp49For. (5’ AAGAAGCGGACGAAGAAGT 3’); rp49Rev. (5’ CCGTAACCGATGTTTGGC 3’). The housekeeping primers for GAPDH and rp49 were used with MRC-5-derived RNA and Aag2-derived RNA, respectively. Relative MAYV genomes were calculated by normalizing the Ct values of MAYV to the inoculum and the housekeeping gene.

Binding assays in WT 3T3 and 3T3 Mxra8 KO cells. WT 3T3 and 3T3 Mxra8 KO cells at near confluency were pre-chilled at 4℃ for 30 minutes. Cells were washed on ice twice with cold PBS containing 5% w/v BSA then inoculated with either WT MAYV or MAYV E2-T179N diluted in PBS with 5% w/v BSA at a MOI of 0.1. Adsorption occurred at 4℃ for 30 minutes. After adsorption, the inoculum was removed, and cells were washed six times with cold PBS with 5% w/v BSA. To remove cells, a solution of 0.1% BSA and 0.3% Igepal CA 630 was added to the cells. After 5 minutes on ice, the cells were taken off the wells by pipetting. To analyze virus bound to the cells, viral genomes were quantified using a SYBR green-based RT-qPCR assay on an Applied Biosystems QuantStudio 3 using the following primers: MAYV 5028For. (5’ CCTCTGTTAGTCCTGTGCAATAC 3’); MAYV 5108Rev. (5’ AAGGTGCTTAGGGAGCTACT 3’); GAPDHFor. (5’ CCAGGTGGTCTCCTCTGACTT 3’); GAPDHRev. (5’ GTTGCTGTAGCCAAATTCGTTGT 3’). Relative MAYV genomes were calculated by normalizing the Ct values of MAYV to the housekeeping gene.

Infection in HEK-293A cells exogenously expressing receptors. HEK-293A cells were transfected when at 70% confluency using JetOPTIMUS reagent (Polyplus) according to the manufacturer’s protocols. Expression constructs encoding SLC1A5, hTIM4, hApoER2 isoform 1, hApoER2 isoform 2, hVLDLR, hMxra8, and hACE2 receptor were supplied to the cells at a 1:1

DNA:reagent ratio. Twenty-four hours post-transfection, cells were pre-chilled then inoculated with WT MAYV or MAYV E2-T179N at a MOI of 0.1 while on ice. The cells were placed at 4℃ for 30 minutes to synchronize binding. The inoculum was removed and cells washed three times with PBS to remove unbound virus. Media was added to the cells, and an initial supernatant sample was collected. The cells were incubated at 37℃ for the next 24 hours post-infection before collecting another supernatant sample. Infectious virus in the supernatant was quantified by plaque assay.

Mouse infection. Four-week-old, female CD-1 mice were inoculated with 10^5 PFU/mL of either WT MAYV or MAYV E2-T179N in 50 μL of RPMI-1640 via injection to the left, hind footpad; in the same manner, a mock-infected group was injected with RPMI-1640. At days 1-3 p.i. serum was collected via submandibular bleed and titrated for viremia. Weights of the mice were taken and footpad widths were measured using digital calipers daily following infection. At seven d.p.i. a subset of mice was euthanized for collection of the left, hind footpad for histopathology. Footpad tissue was fixed in 4% formalin solution, pH 6.2 before decalcification in 10% EDTA pH, 7.3 at 4°C. Tissues were hematoxylin- and eosin-stained, imaged and assigned scores by a board-certified pathologist according to the extent of myofiber loss. The remaining mice were euthanized at 9 d.p.i., and footpads were collected and processed as mentioned above.

Mosquito rearing. Laboratory colonies of *Ae. aegypti* were established from field collections in Kamphaeng Phet Province, Thailand or from a colony isolated from Guerrero, Mexico. All the experiments were performed within 20 generations of laboratory colonization. The insectary conditions for Kamphaeng Phet colony maintenance were 28°C, 70% relative humidity, and a 12h light:12h dark cycle. For Guerrero colonies, maintenance was 28°C, 70-80% relative humidity,

and a 12h light:12h dark cycle. All adults were maintained with permanent access to a 10% sucrose solution. Kamphaeng Phet adult females were offered commercial rabbit blood (BCL) twice a week through a membrane feeding system (Hemotek).

***Ae. aegypti* Kamphaeng Phet population infections.** 6-8 days old Kamphaeng Phet female mosquitoes were selected and starved overnight prior to the bloodmeal. They were fed with an infectious bloodmeal consisting of 2 mL of previously washed human blood (ICareB platform, Institut Pasteur), 1 mL of MAYV viral solution ($3x10^6$ PFU diluted in L15 media with P/S, NEAA, 10% FBS and 1% sodium bicarbonate) and 5 mM ATP (Sigma). Mosquitoes were offered the infectious bloodmeal for 20 min through a membrane feeding system (Hemotek) set at 37°C with a piece of hog gut as the membrane. Following the bloodmeal, fully engorged females were counted, selected, and kept at 28°C in 70% relative humidity and under a 12 h light: 12 h dark cycle with permanent access to 10% sucrose. 7 days after the bloodmeal, mosquitoes were cold-anesthetized and dissected. Legs and wings were collected on ice. Mosquitoes were salivated for 20 min in a tip containing 20 µL of FBS (Gibco), and midguts were collected. Body parts were collected in microtubes (Qiagen) containing 5 mm diameter stainless steel beads (Qiagen) and 150 µL DMEM supplemented with 2% FBS and homogenized using a TissueLyser II (2 cycles at 30 Hz, 1 min, Qiagen). Homogenates were clarified by centrifugation and frozen until plaque assay titration. Viral loads in the saliva were further amplified on BHK-21 or C6/36 for 3 days, and presence of the virus was determined by direct visualization of cytopathic effects or RT-qPCR detection in the conditions described above for competition assays.

***Ae. aegypti* Guerrero population infections.** 3-5 days old Guerrero [53] female mosquitoes were cold anesthetized and sorted 2 days prior to infection. Mosquitoes were starved for 24 hours then

fed with an infectious bloodmeal of defibrinated sheep blood consisting of 0.05 mM ATP and 10^7 PFU/mL of either WT or MAYV E2-T179N diluted in viral diluent. The mosquitoes were fed for 30 minutes through a membrane feeder supplying the infectious bloodmeal at 37°C with a piece of hog gut. Fully engorged mosquitoes were sorted and maintained in an environmental chamber set to 26°C, 70-80% relative humidity, and a 12hr light:12hr dark cycle. Mosquitoes were anesthetized at 10 d.p.i., and bodies, legs and wings, and saliva were collected. Bodies and legs and wings were collected in tubes containing a metal bead and RPMI-1640 with 2% FBS, 10 mM HEPES, Amphotericin B (2.5 μg/mL), and gentamicin sulfate (50 μg/mL) (mosquito diluent). Live salivation occurred for one hour in Type A immersion oil which was collected in mosquito diluent and frozen. Tissues were homogenized using a TissueLyser II (30 Hz/s, 2 min, Qiagen). Homogenates were clarified by centrifugation and frozen until plaque assay titration or RT-qPCR on all samples.

Human blood and ethics statement. Human blood used to feed mosquitoes was obtained from healthy volunteer donors. Healthy donor recruitment was organized by the local investigator assessment using medical history, laboratory results, and clinical examinations. All adult subjects provided written informed consent. Biological samples were supplied through participation of healthy volunteers at the ICAReB biobanking platform (BB-0033-00062/ICAReB platform/Institut Pasteur, Paris/BBMRI AO203/[BIORESOURCE]) of the Institut Pasteur to the CoSImmGen and Diagmicoll protocols, which have been approved by the French Ethical Committee (Comité de Protection des Personnes; CPP) Ile-de-France I. The Diagmicoll protocol was declared to the French Research Ministry under the reference: DC 2008–68 COL 1.

Biosafety protocols. The research protocols at Virginia Tech were approved by the Institutional Animal Care and Use Committee (IACUC #18-084) and the Institutional Biosafety Committee of Virginia Tech (IBC #18-026). All studies performed at the Institut Pasteur were approved by the Committee of Health, Safety, and Working Conditions under protocol number 75-1448. Biosafety practices at Virginia Tech were conducted according to the recommendations of the National Institutes of Health and the Centers for Disease Control and Prevention which included performing mosquito infections at ACL3 and mouse infections at ABSL2. The mouse studies were carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Handling of MAYV and CHIKV was performed at BSL2 and BSL3, respectively. At Institut Pasteur, all work with MAYV was performed in dedicated BSL3 facilities.

***In silico* N-glycosylation predictions:** N-glycosylation prediction was performed using two webservers: NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) and NGlycPred (https://bioinformatics.niaid.nih.gov/nglycpred/). For both servers, we used the MAYV E2 glycoprotein sequence from strain IQT4235. With the NetNGlyc server, we used the default option to run predictions only on the Asn-Xaa-Ser/Thr sequons and used a 0.5 threshold. In NGlycPred server, we used the preferred option to consider structural properties and patterns.

Molecular dynamics simulations and calculation of binding free energy values: All the structural models of MAYV in complex with the human Mxra8 receptor were built based on the coordinates position of CHIKV crystallographic structure in complex with Mxra8 (PDB ID: 6JO8 [47]). To model the complexes between MAYV E2 subdomain and Mxra8 D2 domain, we removed from 7KO8 and 6JO8 structures all residues not included in those domains. The

asymmetric unit of MAYV Cryo-EM structure (PDB ID: 7KO8) has 4 E1-E2 heterodimer copies whereas CHIKV crystallographic structure (PDB ID: 6JO8) has three copies. Thus, we cross-aligned, one by one, the E2 B subdomains of MAYV structure to each E2 B subdomains of CHIKV with the Chimera software [75]. This cross-alignment returned 12 complexes between MAYV E2 B subdomains and Mxra8 D2 domain. We estimated predicted free energy changes upon T179N mutation for all 12 MAYV complexes using FoldX 5 software [76]. Before energy calculations, each structure was treated to optimize sidechain rotamers using the FoldX function *RepairPDB*. In FoldX, wild-type structures were mutated and the ΔΔG (in kcal/mol) was estimated by subtracting the resulting energy of the mutated complex by the wild-type complex. We reported the FoldX result as mean ± standard deviation of ΔΔG for the 12 MAYV complexes. These same WT and T179N complexes were used as initial structures for molecular dynamics simulations using Amber 20 suit of programs with the *ff14SB* force field. Using *tleap*, we added Na+ and Cl- counter ions to reach net-neutralization with salt excess to reach 150 mM NaCl [77]. Each structure was solvated in a truncated octahedral box (15 A from the solute) filled with *TIP3P* water. The *PBRadii* was set to *mbondi2* with *tleap* program. The system was minimized by 2500 steps of steepest descent minimization and 2500 steps of conjugate gradient. The equilibration was performed using the NVT ensemble (200 ps) followed by NPT ensemble (200 ps), both with harmonic restraints in protein atoms. A last NPT equilibration step without restraints was performed for 500 ps. For each complex, the production run was performed at 298 K for 2 ns with a time step of 0.002 fs in triplicate. Hydrogen-containing bonds were constrained using SHAKE [78]. Long-range electrostatic interactions were calculated using particle Mesh Ewald and short-range nonbonded interactions were calculated with a 9 A cutoff. In all simulation steps we applied a harmonic restraint of 10 kcal·mol−1 ·Å −2 to the backbone atoms to prevent an overall

displacement of the complex from its initial position. We estimated the relative binding free energy using 1-trajectory Molecular Mechanics Generalized Born Surface Area (MM/GBSA) calculations. For this, we used 20 frames from entire 2 ns simulations with MMPBSA.py script [79]. The individual topology files for the calculations were generated using ante-MMPBSA.py script. Since our objective was only to compare wild-type and mutant interactions, we did not include the entropy term. As result, the binding free energy estimate reported a relative value, not an absolute. In MMGBSA calculations, we used igb model 5 (all the other parameters were default). We reported the MMGBSA results as a ΔΔG of the difference between the relative ΔG estimated for T179N MAYV:Mxra8 and WT MAYV:Mxra8. To model the entire MAYV spike, formed by three E1-E2 heterodimers (not including transmembrane domains), in complex with Mxra8, we aligned individually the MAYV E1-E2 heterodimers (from PDB ID: 7KO8) with CHIKV E1-E2 heterodimers (from PDB ID: 6JO8) in Chimera. Then, we created the MAYV T179N mutated structure with the YASARA software [80]. Both wild-type and mutant MAYV trimeric complexes were submitted to energy minimization in solvent with YASARA to obtain an initial model. Then, we performed a 20 ns molecular dynamics simulation in quintuplicate for WT and T179N MAYV complexes, following the same simulation protocol described above without backbone restraints. The MMGBSA calculation was performed using 100 frames from the last 4 ns with the same parameters described above. Before any calculations or simulations, the protonation state at pH 7.4 for each residue of the structural models was assigned using the *pdb2pqr30* script [81] and *propka* as the titration state method [82]. For all cases, we also modeled the missing loop of Mxra8 D2 domain using YASARA software [80]. The convergence of the simulations was accessed by the RMSD of the Mxra8 backbone in a trajectory aligned by the MAYV backbone (**Fig. S7**).

Statistical analysis. All data were analyzed using the Prism 9 software (GraphPad) and are presented as mean ± standard deviation unless indicated otherwise. The statistical tests used are described in each figure legend and were performed using GraphPad Prism. All experiments (except serial and DMS passaging) were performed in two to three independent replicates with at least three technical replicates per group.

Acknowledgements. The authors thank Rodrigo Guabiraba for helpful discussions. Illustrative figures in this manuscript were created with BioRender.com. This work was funded by the DARPA PREEMPT program managed by Dr. Rohit Chitale and Dr. Kerri Dugan and administered through DARPA Cooperative Agreement HR001118S0017 (the content of the information does not necessarily reflect the position or the policy of the U.S. government, and no official endorsement should be inferred). This work also received funding from Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (grant ANR-10-LABX-62-IBEID) to M-C.S. and M.V. Further support was provided by startup funds awarded to J.W-L by the Virginia-Maryland College of Veterinary Medicine and a grant from the One Health Research Funding Program awarded to J.W-L and P.M.

References

1. Brady OJ, Hay SI. The first local cases of Zika virus in Europe. The Lancet. 2019;394: 1991–1992. doi:10.1016/s0140-6736(19)32790-4

2. Philip C, Novick CG, Novick LF. Local Transmission of Zika Virus in Miami-Dade County: The Florida Department of Health Rises to the Challenge. Journal of Public Health Management and Practice. 2019;25: 277–287. doi:10.1097/phh.0000000000000990

3. Anderson CR, Wattley GH, Ahin NW, Downs WG, Reese AA. Mayaro Virus: A New Human Disease Agent: II. Isolation from Blood of Patients in Trinidad, B.W.I.1. The American Journal of Tropical Medicine and Hygiene. 1957;6: 1012–1016. doi:10.4269/ajtmh.1957.6.1012

4. Freitas RB, Rosa JFT da, LeDuc JW, Pinheiro FP, Gabbay YB, Mello WA. An Outbreak of Mayaro Virus Disease in Belterra, Brazil: I. Clinical and Virological Findings*. The American Journal of Tropical Medicine and Hygiene. 1981;30: 674–681. doi:10.4269/ajtmh.1981.30.674

5. Navarrete-Espinosa J, Gómez-Dantés H. [Arbovirus causing hemorrhagic fever at IMSS]. Revista médica del Instituto Mexicano del Seguro Social. 2006;44: 347–53. Available: https://www.ncbi.nlm.nih.gov/pubmed/16904038

6. Neumayr A, Gabriel M, Fritz J, Günther S, Hatz C, Schmidt-Chanasit J, et al. Mayaro Virus Infection in Traveler Returning from Amazon Basin, Northern Peru. Emerging Infectious Diseases. 2012;18: 695–696. doi:10.3201/eid1804.111717

7. Talarmin A, Chandler LJ, Kazanji M, de Thoisy B, Debon P, Lelarge J, et al. Mayaro virus fever in French Guiana: isolation, identification, and seroprevalence. Am J Trop Med Hyg. 1998;59: 452–456. doi:10.4269/ajtmh.1998.59.452

8. Lema AB, Gajdusek DC, Eichenwald H, Schaeffer M. Epidemic Jungle Fevers Among Okinawan Colonists in the Bolivian Rain Forest: I. Epidemiology. The American Journal of Tropical Medicine and Hygiene. 1959;8: 372–396. doi:10.4269/ajtmh.1959.8.372

9. Izurieta RO, Macaluso M, Watts DM, Tesh RB, Guerra B, Cruz LM, et al. Hunting in the rainforest and mayaro virus infection: An emerging alphavirus in Ecuador. Journal of Global Infectious Diseases. 2011;3: 317–323. doi:10.4103/0974-777x.91049

10. Auguste AJ, Liria J, Forrester NL, Giambalvo D, Moncada M, Long KC, et al. Evolutionary and Ecological Characterization of Mayaro Virus Strains Isolated during an Outbreak, Venezuela, 2010. Emerg Infect Dis. 2015;21: 1742–1750. doi:10.3201/eid2110.141660

11. Hassing R-J, Leparc-Goffart I, Blank SN, Thevarayan S, Tolou H, van Doornum G, et al. Imported Mayaro virus infection in the Netherlands. Journal of Infection. 2010;61: 343–345. doi:10.1016/j.jinf.2010.06.009

12. Receveur MC, Grandadam M, Pistone T, Malvy D. Infection with Mayaro virus in a French traveller returning from the Amazon region, Brazil, January, 2010. Euro surveillance : bulletin Européen sur les maladies transmissibles = European communicable disease bulletin. 2010;15. Available: https://www.ncbi.nlm.nih.gov/pubmed/20460093

13. Friedrich-Jänicke B, Emmerich P, Tappe D, Günther S, Cadar D, Schmidt-Chanasit J. Genome analysis of Mayaro virus imported to Germany from French Guiana. Emerg Infect Dis. 2014;20: 1255–1257. doi:10.3201/eid2007.140043

14. Mackay IM, Arden KE. Mayaro virus: a forest virus primed for a trip to the city? 2016;18: 724–734. doi:10.1016/j.micinf.2016.10.007

15. Mavian C, Rife BD, Dollar JJ, Cella E, Ciccozzi M, Prosperi MCF, et al. Emergence of recombinant Mayaro virus strains from the Amazon basin. Scientific Reports. 2017;7: 8718. doi:10.1038/s41598-017-07152-5

16. Paniz-Mondolfi AE, Rodriguez-Morales AJ, Blohm G, Marquez M, Villamil-Gomez WE. ChikDenMaZika Syndrome: the challenge of diagnosing arboviral infections in the midst of

concurrent epidemics. Annals of Clinical Microbiology and Antimicrobials. 2016;15: 42. doi:10.1186/s12941-016-0157-x

17. Levi LI, Vignuzzi M. Arthritogenic Alphaviruses: A Worldwide Emerging Threat? Microorganisms. 2019;7: 133. doi:10.3390/microorganisms7050133

18. Aguilar-Luis MA, Del Valle-Mendoza J, Silva-Caso W, Gil-Ramirez T, Levy-Blitchtein S, Bazán-Mayra J, et al. An emerging public health threat: Mayaro virus increases its distribution in Peru. Int J Infect Dis. 2020;92: 253–258. doi:10.1016/j.ijid.2020.01.024

19. Taylor SF, Patel PR, Herold TJS. Recurrent Arthralgias in a Patient With Previous Mayaro Fever Infection. Southern Medical Journal. 2005;98: 484–485. doi:10.1097/01.smj.0000145879.14102.f4

20. Mourão MPG, Bastos M de S, Figueiredo RP de, Gimaque JBL, Galusso E dos S, Kramer VM, et al. Mayaro Fever in the City of Manaus, Brazil, 2007–2008. Vector-Borne and Zoonotic Diseases. 2012;12: 42–46. doi:10.1089/vbz.2011.0669

21. Tappe D, Perez-Giron J, Just-Nubling G, Schuster G, Gomez-Medina S, Gunther S, et al. Sustained Elevated Cytokine Levels during Recovery Phase of Mayaro Virus Infection. 2016;22: 750–752. doi:10.3201/eid2204.151502

22. Chen Z, Lang D. The effectiveness of disease management interventions on health-related quality of life of patients with established arthritogenic alphavirus infections: a systematic review protocol. JBI Database of Systematic Reviews and Implementation Reports. 2013;11: 56–72. doi:10.11124/jbisrir-2013-934

23. Thoisy B, Gardon J, Salas R, Morvan J, Kazanji M. Mayaro Virus in Wild Mammals, French Guiana. Emerg Infect Dis. 2003;9: 1326–1329. doi:10.3201/eid0910.030161

24. Acosta-Ampudia Y, Monsalve DM, Rodriguez Y, Pacheco Y, Anaya J-M, Ramirez-Santana C. Mayaro: an emerging viral threat? 2018;7: 163. doi:10.1038/s41426-018-0163-5

25. Izurieta R, DeLacure D, Izurieta A, Hoare I, Ortiz M. Mayaro virus: the jungle flu. Virus Adapt Treat. 2018;10: 9–17. doi:10.2147/vaat.s128711

26. Hozé N, Salje H, Rousset D, Fritzell C, Vanhomwegen J, Bailly S, et al. Reconstructing Mayaro virus circulation in French Guiana shows frequent spillovers. Nature Communications. 2020;11: 2842. doi:10.1038/s41467-020-16516-x

27. Pérez JG, Carrera J-P, Serrano E, Pittí Y, Maguiña JL, Mentaberre G, et al. Serologic Evidence of Zoonotic Alphaviruses in Humans from an Indigenous Community in the Peruvian Amazon. The American Journal of Tropical Medicine and Hygiene. 2019;101: 1212–1218. doi:10.4269/ajtmh.18-0850

28. Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney M-C, et al. Genome Microevolution of Chikungunya Viruses Causing the Indian Ocean Outbreak. PLoS Medicine. 2006;3: e263. doi:10.1371/journal.pmed.0030263

29. LeDuc JW, Pinheiro FP, Peterson NE, Hoch AL. An Outbreak of Mayaro Virus Disease in Belterra, Brazil: III. Entomological and Ecological Studies*. The American Journal of Tropical Medicine and Hygiene. 1981;30: 689–698. doi:10.4269/ajtmh.1981.30.689

30. Pereira TN, Carvalho FD, Mendonça SFD, Rocha MN, Moreira LA. Vector competence of Aedes aegypti, Aedes albopictus, and Culex quinquefasciatus mosquitoes for Mayaro virus. PLOS Neglected Tropical Diseases. 2020;14: e0007518. doi:10.1371/journal.pntd.0007518

31. Tesh RB, Higgs S, Hausser NL, Thangamani S, Kochel TJ, Ziegler SA, et al. Experimental Transmission of Mayaro Virus by Aedes aegypti. The American Journal of Tropical Medicine and Hygiene. 2011;85: 750–757. doi:10.4269/ajtmh.2011.11-0359

32. Wiggins K, Eastmond B, Alto BW. Transmission potential of Mayaro virus in Florida Aedes aegypti and Aedes albopictus mosquitoes. 2018;32: 436–442. doi:10.1111/mve.12322

33. Dieme C, Ciota AT, Kramer LD. Transmission potential of Mayaro virus by Aedes albopictus, and Anopheles quadrimaculatus from the USA. Parasites & Vectors. 2020;13: 613. doi:10.1186/s13071-020-04478-4

34. Smith GC, Francy DB. Laboratory studies of a Brazilian strain of Aedes albopictus as a potential vector of Mayaro and Oropouche viruses. Journal of the American Mosquito Control Association. 1991;7: 89–93. Available: https://www.ncbi.nlm.nih.gov/pubmed/1646286

35. Brustolin M, Pujhari S, Henderson CA, Rasgon JL. Anopheles mosquitoes may drive invasion and transmission of Mayaro virus across geographically diverse regions. PLOS Neglected Tropical Diseases. 2018;12: e0006895. doi:10.1371/journal.pntd.0006895

36. Ashenberg O, Padmakumar J, Doud MB, Bloom JD. Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA. PLOS Pathogens. 2017;13: e1006288. doi:10.1371/journal.ppat.1006288

37. Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, et al. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. Proc Natl Acad Sci USA. 2018;115: E8276. doi:10.1073/pnas.1806133115

38. Setoh YX, Amarilla AA, Peng NYG, Griffiths RE, Carrera J, Freney ME, et al. Determinants of Zika virus host tropism uncovered by deep mutational scanning. Nature Microbiology. 2019;4: 876–887. doi:10.1038/s41564-019-0399-4

39. Carrasco P, Daròs JA, Agudelo-Romero P, Elena SF. A real-time RT-PCR assay for quantifying the fitness of tobacco etch virus in competition experiments. Journal of Virological Methods. 2007;139: 181–188. doi:10.1016/j.jviromet.2006.09.020

40. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Spike mutation D614G alters SARS-CoV-2 fitness. Nature. 2021;592: 116–121. doi:10.1038/s41586-020-2895-3

41. Weger-Lucarelli J, Carrau L, Levi LI, Rezelj V, Vallet T, Blanc H, et al. Host nutritional status affects alphavirus virulence, transmission, and evolution. PLoS Pathog. 2019;15: e1008089. doi:10.1371/journal.ppat.1008089

42. Fox JM, Long F, Edeling MA, Lin H, van Duijl-Richter MKS, Fong RH, et al. Broadly Neutralizing Alphavirus Antibodies Bind an Epitope on E2 and Inhibit Entry and Egress. Cell. 2015;163: 1095–1107. doi:10.1016/j.cell.2015.10.050

43. Carvalho CAM, Silva JL, Oliveira AC, Gomes AMO. On the entry of an emerging arbovirus into host cells: Mayaro virus takes the highway to the cytoplasm through fusion with early endosomes and caveolae-derived vesicles. PeerJ. 2017;5: e3245. doi:10.7717/peerj.3245

44. Zhang R, Kim AS, Fox JM, Nair S, Basore K, Klimstra WB, et al. Mxra8 is a receptor for multiple arthritogenic alphaviruses. 2018;557: 570–574. doi:10.1038/s41586-018-0121-3

45. Zhang R, Earnest JT, Kim AS, Winkler ES, Desai P, Adams LJ, et al. Expression of the Mxra8 Receptor Promotes Alphavirus Infection and Pathogenesis in Mice and Drosophila. Cell Rep. 2019;28: 2647-2658.e5. doi:10.1016/j.celrep.2019.07.105

46. Ribeiro-Filho HV, Coimbra LD, Cassago A, Rocha RPF, Guerra JV da S, Felicio R de, et al. Cryo-EM structure of the mature and infective Mayaro virus at 4.4 Å resolution reveals features of arthritogenic alphaviruses. Nature Communications. 2021;12: 3038. doi:10.1038/s41467-021-23400-9

47. Song H, Zhao Z, Chai Y, Jin X, Li C, Yuan F, et al. Molecular Basis of Arthritogenic Alphavirus Receptor MXRA8 Binding to Chikungunya Virus Envelope Protein. Cell. 2019;177: 1714-1724.e12. doi:10.1016/j.cell.2019.04.008

48. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002; 310–322. Available: https://www.ncbi.nlm.nih.gov/pubmed/11928486

49. Chuang G-Y, Boyington JC, Joyce MG, Zhu J, Nabel GJ, Kwong PD, et al. Computational prediction of N-linked glycosylation incorporating structural properties and patterns. Bioinformatics. 2012;28: 2249–2255. doi:10.1093/bioinformatics/bts426

50. De Caluwé L, Coppens S, Vereecken K, Daled S, Dhaenens M, Van Ostade X, et al. The CD147 Protein Complex Is Involved in Entry of Chikungunya Virus and Related Alphaviruses in Human Cells. Front Microbiol. 2021;12: 615165. doi:10.3389/fmicb.2021.615165

51. Jemielity S, Wang JJ, Chan YK, Ahmed AA, Li W, Monahan S, et al. TIM-family proteins promote infection of multiple enveloped viruses through virion-associated phosphatidylserine. PLoS Pathog. 2013;9: e1003232. doi:10.1371/journal.ppat.1003232

52. Clark LE, Clark SA, Lin C, Liu J, Coscia A, Nabel KG, et al. VLDLR and ApoER2 are receptors for multiple alphaviruses. Nature. 2022;602: 475–480. doi:10.1038/s41586-021-04326-0

53. Garcia-Luna SM, Weger-Lucarelli J, Rückert C, Murrieta RA, Young MC, Byas AD, et al. Variation in competence for ZIKV transmission by Aedes aegypti and Aedes albopictus in Mexico. PLoS Negl Trop Dis. 2018;12: e0006599. doi:10.1371/journal.pntd.0006599

54. Lanciotti RS, Kosoy OL, Laven JJ, Panella AJ, Velez JO, Lambert AJ, et al. Chikungunya virus in US travelers returning from India, 2006. Emerg Infect Dis. 2007;13: 764–767. doi:10.3201/eid1305.070015

55. Fumagalli MJ, de Souza WM, de Castro-Jorge LA, de Carvalho RVH, Castro Í de A, de Almeida LGN, et al. Chikungunya Virus Exposure Partially Cross-Protects against Mayaro Virus Infection in Mice. Heise MT, editor. J Virol. 2021;95: e01122-21. doi:10.1128/JVI.01122-21

56. Webb EM, Azar SR, Haller SL, Langsjoen RM, Cuthbert CE, Ramjag AT, et al. Effects of Chikungunya virus immunity on Mayaro virus disease and epidemic potential. Sci Rep. 2019;9: 20399. doi:10.1038/s41598-019-56551-3

57. Weger-Lucarelli J, Aliota MT, Wlodarchak N, Kamlangdee A, Swanson R, Osorio JE. Dissecting the role of E2 protein domains in Alphavirus pathogenicity. J Virol. 2015;90: 2418–2433. doi:10.1128/JVI.02792-15

58. Byrnes AP, Griffin DE. Binding of Sindbis virus to cell surface heparan sulfate. J Virol. 1998;72: 7349–7356. doi:10.1128/JVI.72.9.7349-7356.1998

59. Jose J, Przybyla L, Edwards TJ, Perera R, Burgner JW 2nd, Kuhn RJ. Interactions of the cytoplasmic domain of Sindbis virus E2 with nucleocapsid cores promote alphavirus budding. J Virol. 2012;86: 2585–2599. doi:10.1128/JVI.05860-11

60. Voss JE, Vaney M-C, Duquerroy S, Vonrhein C, Girard-Blanc C, Crublet E, et al. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. Nature. 2010;468: 709–712. doi:10.1038/nature09555

61. Tsetsarkin KA, Chen R, Leal G, Forrester N, Higgs S, Huang J, et al. Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. Proc Natl Acad Sci U S A. 2011;108: 7872–7877. doi:10.1073/pnas.1018344108

62. Tsetsarkin KA, Chen R, Yun R, Rossi SL, Plante KS, Guerbois M, et al. Multi-peaked adaptive landscape for chikungunya virus evolution predicts continued fitness optimization in Aedes albopictus mosquitoes. Nature Communications. 2014;5: 4084. doi:10.1038/ncomms5084

63. Chuong C, Bates TA, Weger-Lucarelli J. Infectious cDNA clones of two strains of Mayaro virus for studies on viral pathogenesis and vaccine development. Virology. 2019;535: 227–231. doi:10.1016/j.virol.2019.07.013

64. Kang L, He G, Sharp AK, Wang X, Brown AM, Michalak P, et al. A selective sweep in the Spike gene has driven SARS-CoV-2 human adaptation. Cell. 2021;184: 4392-4400.e4. doi:10.1016/j.cell.2021.07.007

65. Weger-Lucarelli J, Garcia SM, Rückert C, Byas A, O'Connor SL, Aliota MT, et al. Using barcoded Zika virus to assess virus population structure in vitro and in Aedes aegypti mosquitoes. Virology. 2018;521: 138–148. doi:10.1016/j.virol.2018.06.004

66. Kutner RH, Zhang X-Y, Reiser J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. Nat Protoc. 2009;4: 495–505. doi:10.1038/nprot.2009.22

67. Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner. 2014. Available: https://www.osti.gov/biblio/1241166

68. Rodgers MA, Wilkinson E, Vallari A, McArthur C, Sthreshley L, Brennan CA, et al. Sensitive Next-Generation Sequencing Method Reveals Deep Genetic Diversity of HIV-1 in the Democratic Republic of the Congo. J Virol. 2017;91. doi:10.1128/JVI.01841-16

69. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics. 2015;31: 2032–2034. doi:10.1093/bioinformatics/btv098

70. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26: 589–595. doi:10.1093/bioinformatics/btp698

71. Doran AG, Creevey CJ. Snpdat: easy and rapid annotation of results from de novo snp discovery projects for model and non-model organisms. BMC Bioinformatics. 2013;14: 45. doi:10.1186/1471-2105-14-45

72. McCrone JT, Lauring AS. Measurements of Intrahost Viral Diversity Are Extremely Sensitive to Systematic Errors in Variant Calling. J Virol. 2016;90: 6884–6895. doi:10.1128/JVI.00667-16

73. Marano JM, Chuong C, Weger-Lucarelli J. Rolling circle amplification: A high fidelity and efficient alternative to plasmid preparation for the rescue of infectious clones. Virology. 2020;551: 58–63. doi:10.1016/j.virol.2020.08.016

74. Gnädig NF, Beaucourt S, Campagnola G, Bordería AV, Sanz-Ramos M, Gong P, et al. Coxsackievirus B3 mutator strains are attenuated in vivo. Proc Natl Acad Sci U S A. 2012;109: E2294-2303. doi:10.1073/pnas.1204022109

75. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera?A visualization system for exploratory research and analysis. J Comput Chem. 2004;25: 1605–1612. doi:10.1002/jcc.20084

76. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Research. 2005;33: W382–W388. doi:10.1093/nar/gki387

77. Schmit JD, Kariyawasam NL, Needham V, Smith PE. SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation. J Chem Theory Comput. 2018;14: 1823–1827. doi:10.1021/acs.jctc.7b01254

78. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics. 1977;23: 327–341. doi:10.1016/0021-9991(77)90098-5

79. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. *MMPBSA.py* : An Efficient Program for End-State Free Energy Calculations. J Chem Theory Comput. 2012;8: 3314–3321. doi:10.1021/ct300418h

80. Krieger E, Vriend G. YASARA View—molecular graphics for all devices—from smartphones to workstations. Bioinformatics. 2014;30: 2981–2982. doi:10.1093/bioinformatics/btu426

81. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Research. 2007;35: W522–W525. doi:10.1093/nar/gkm276

82. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K_a Predictions. J Chem Theory Comput. 2011;7: 525–537. doi:10.1021/ct100578z

Supplementary Figures

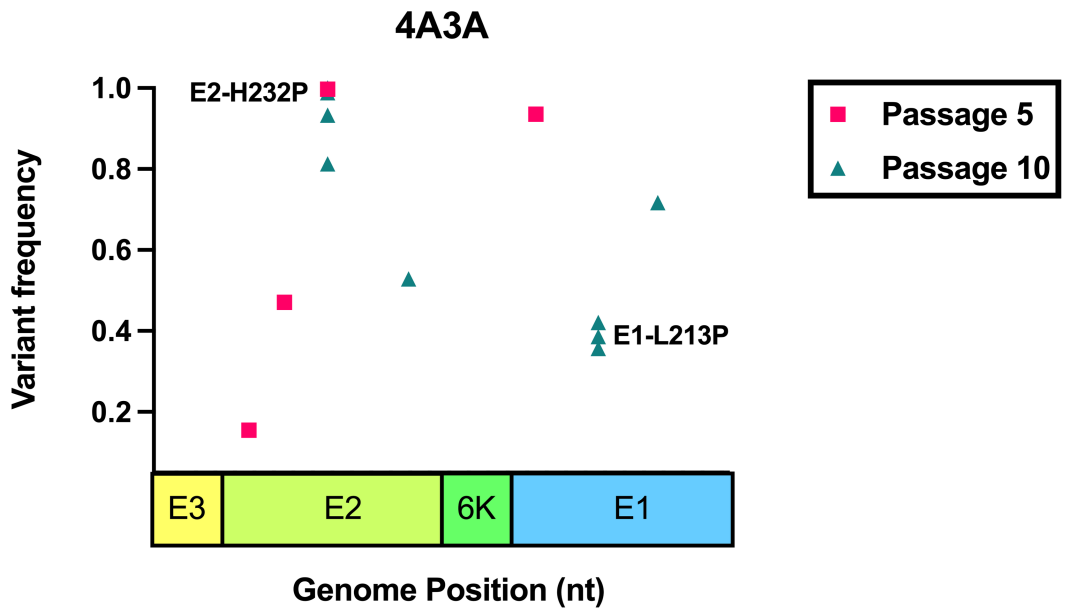


Figure S1: Natural evolution of Mayaro virus (MAYV) in *Anopheles gambiae* cells. Experimental evolution was performed using traditional serial passaging in 4a-3A cells. We serially passaged MAYV at a MOI of 0.01 for a total of ten passages. Following one, five, and ten

passages, the viral RNA was sequenced using Illumina NGS to identify potentially adaptive mutations. No high frequency variants were identified following one passage, and as such are not depicted here.

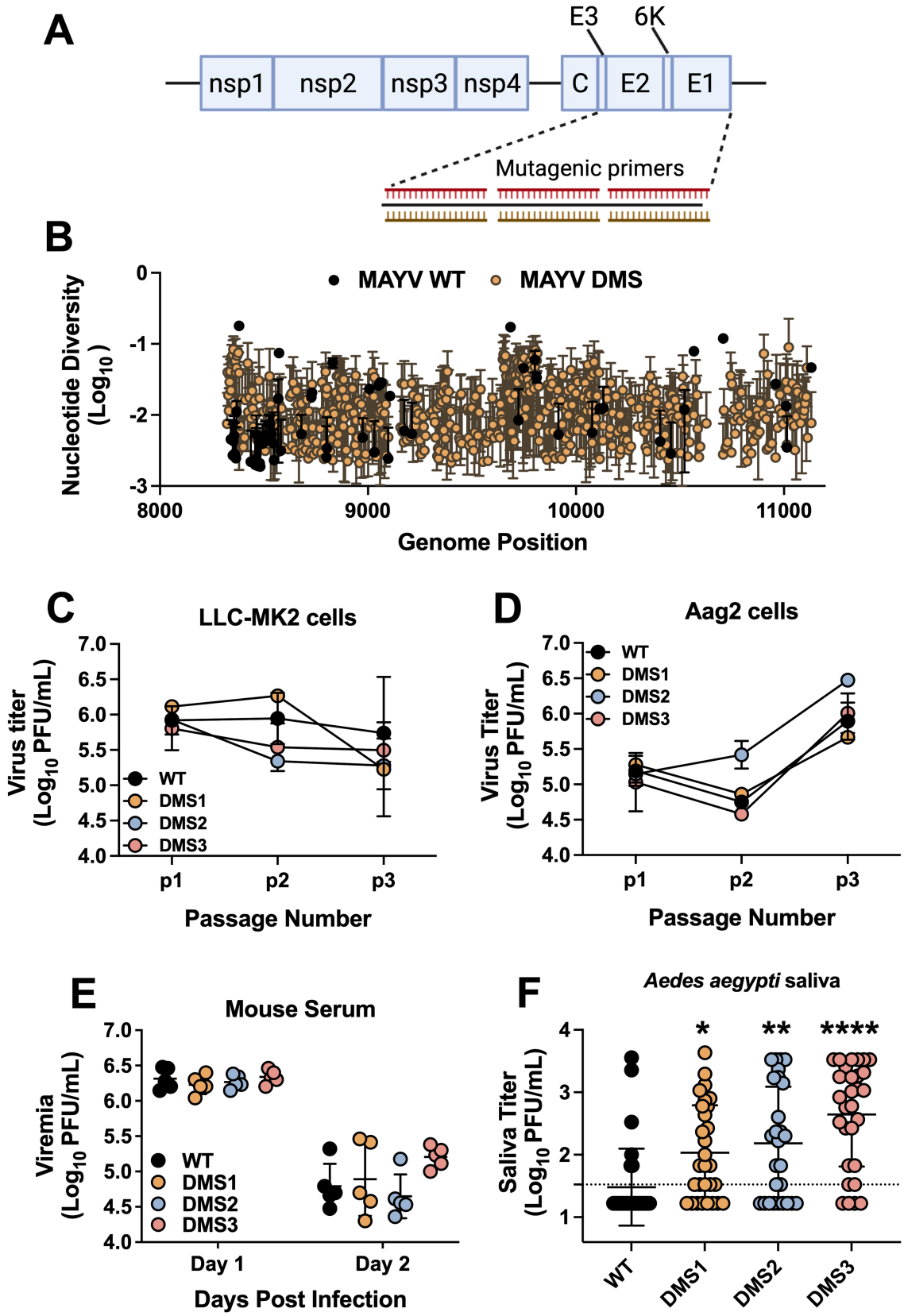

A
E3
6K
nsp1
nsp2
nsp3
nsp4
C
E2
E1
Mutagenic primers
B
MAYV WT
MAYV DMS
Nucleotide Diversity (Log10)
0
-1
-2
-3
8000
9000
10000
11000
Genome Position
C
LLC-MK2 cells
Virus titer (Log10 PFU/mL)
7.0
6.5
6.0
5.5
5.0
4.5
4.0
WT
DMS1
DMS2
DMS3
p1
p2
p3
Passage Number
D
Aag2 cells
Virus Titer (Log10 PFU/mL)
Passage Number
E
Mouse Serum
Viremia (Log10 PFU/mL)
Day 1
Day 2
Days Post Infection
F
Aedes aegypti saliva
Saliva Titer (Log10 PFU/mL)
*
**
****
4
3
2
1
WT
DMS1
DMS2
DMS3

Figure S2: MAYV deep mutational scanning (DMS) populations are highly diverse and replicate *in vitro* and *in vivo*. A. Genome organization of MAYV DMS viruses. Created with BioRender.com. **B.** Non-synonymous nucleotide diversity for wild-type (WT) and MAYV DMS. The three independent MAYV DMS populations were combined to aid visualization. **C-D**. Three passages were performed *in vitro* at an MOI of 0.01 in LLC-MK2 (monkey kidney; **C**) and Aag2 (mosquito; **D**). **E.** Statistical analysis: * = $p<0.05$; ** = $p<0.01$; **** = $p<0.0001$ (one-way ANOVA with Dunnett's correction).

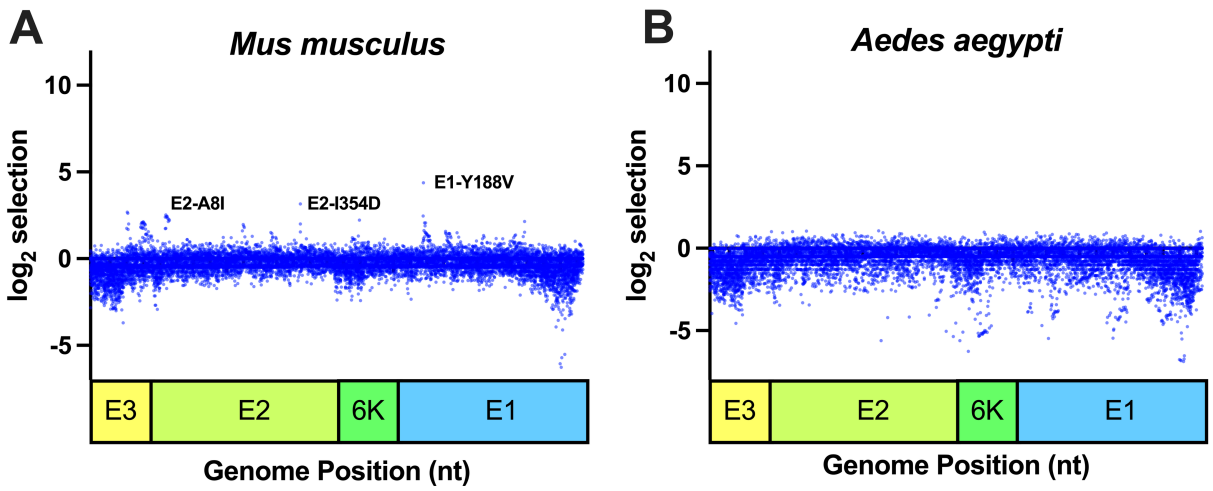


Figure S3: **Deep mutational scanning (DMS) of Mayaro virus (MAYV) in live mice and mosquitoes. A-B.** The three MAYV DMS populations, along with WT MAYV, were used to perform three passages in mice (**A**) and *Ae. aegypti* mosquitoes (**B**). Following passage, the viral RNA was sequenced, and selection analyses were performed to identify enriched variants. The top three variants based on selection strength, are presented for mice (**A**). Given the low level of enrichment observed in *Ae. aegypti* mosquitoes, we do not present individual variants.

A

Mutant Name	Method of detection	Phenotype
MAYV WT	N/A	Control
E2-I354D	DMS	Enriched in mice
E2-A8I	DMS	Enriched in mice
E2-S87T	DMS	Enriched in mice
E2-A92V	DMS	Enriched in mice
E2-E272Q	DMS	Enriched in *Aedes aegypti*
E2-A128G	DMS	Enriched in *Aedes aegypti*
E1-T433S	DMS	Enriched in *Aedes aegypti*
E2-L414I	DMS	Consensus change in some mosquito saliva
E3-P34T	DMS	Consensus change in some mosquito saliva
E3-K65A	DMS	Consensus change in some mosquito saliva

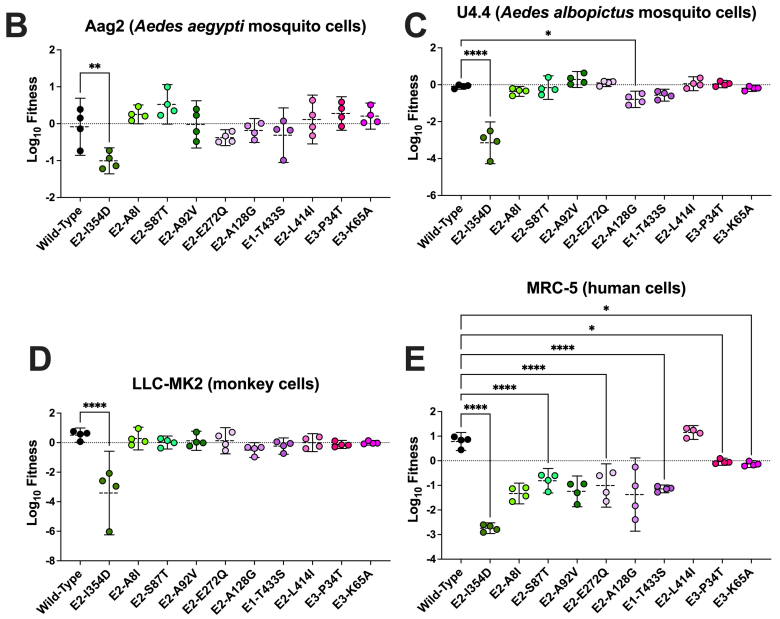


Figure S4. Variants enriched in live mice and mosquitoes have little effect on viral fitness. A: List of the viruses used for the competition assays. All mutations were identified from deep mutational scanning data, and the phenotypes observed for the different viruses are indicated. The wild-type data—included as a comparison—is the same as is presented in Figure 2 in the main text. **B-E:** Competition assays in Aag2 (**B**)**,** U4.4 (**C**), LLC-MK2 (**D**), and MRC-5 (**E**) cells. Cells were infected at a MOI of 0.01 using a 1:1 ratio based on PFUs for each mutant or WT with a genetically marked MAYV competitor virus. Viral supernatants were harvested at 72h post-infection for Aag2 and 48h post-infection for the other cell lines. Replication of WT and mutant viruses was assessed by RT-qPCR using specific probes labeled with different fluorophores. Log_{10} fitness was calculated by normalizing replication of each virus against a genetically marked reference virus. The mean of 4 independent experiments is represented with standard deviation. Statistical analysis: * = $p<0.05$; ** = $p<0.01$; **** = $p<0.0001$ (one-way ANOVA with Dunnett's correction). Consensus change in some mosquito saliva means that the mutation was found at a

above 50% in one mosquito saliva sample: E2 L414I was found at 89%, E3-P34T was found at 99.7%, and E3-K65A was found at 99.7%.

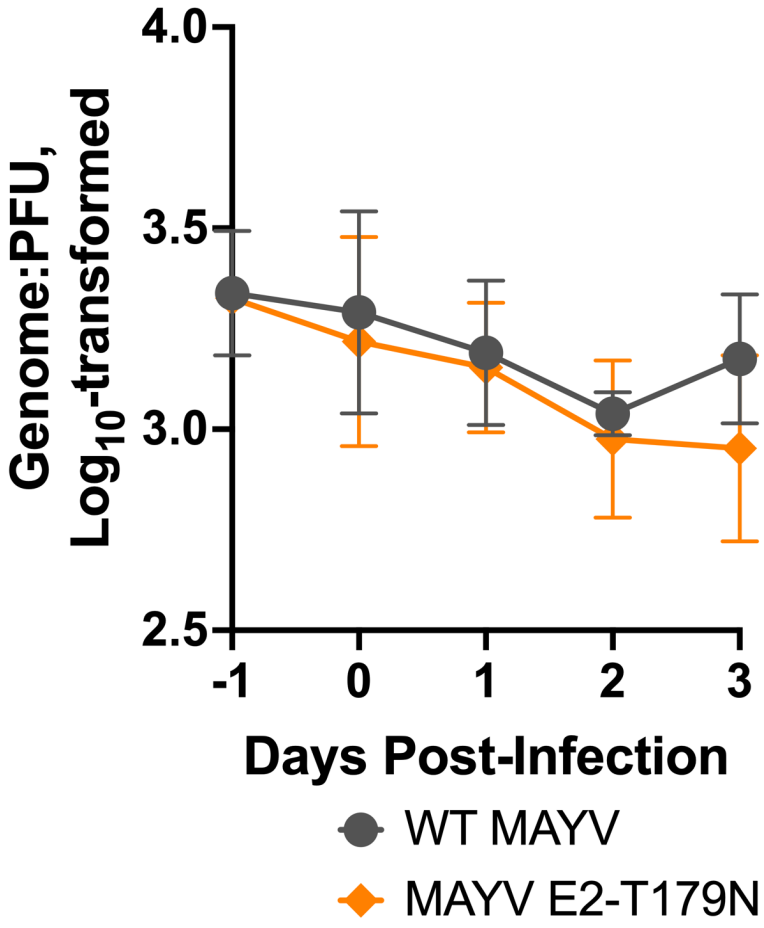


Figure S5. WT MAYV and MAYV E2-T179N have similar genome:PFU ratios following Aag2 infection. Aag2 cells were infected with either WT MAYV or MAYV E2-T179N at a MOI of 0.1, and genome:PFU ratios were measured each day post infection via RT-qPCR and plaque assay. The -1 day post infection represents the genome:PFU ratio of the inoculum used for the infection. Genome:PFU ratios represented are $\log_{10}$-transformed. The means of two independent experiments are represented, and error bars represent standard deviation. Statistical analysis: non-significant (two-way ANOVA with Šídák's multiple comparisons test).

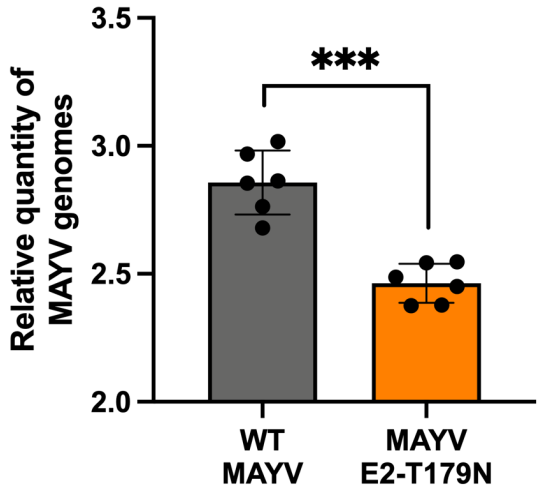


Figure S6. Binding affinity of WT and MAYV E2-T179N MAYV to Aag2 cells. Binding assays of WT MAYV and MAYV E2-T179N were performed by inoculating chilled Aag2 cells with virus at a MOI of 0.1 prior to adsorption at 4°C. Unbound virus was washed away, and bound virus was quantified by RT-qPCR using RNA extracted from the inoculated cells. The relative quantities of MAYV genomes were determined by normalizing to the Ct values of a housekeeping gene and the Ct value of virus in the inoculum. Data represent the means of two independent experiments. Y-axis is log-transformed for clearer visualization. Statistical analysis: *** = $p<0.002$ (unpaired t-test).

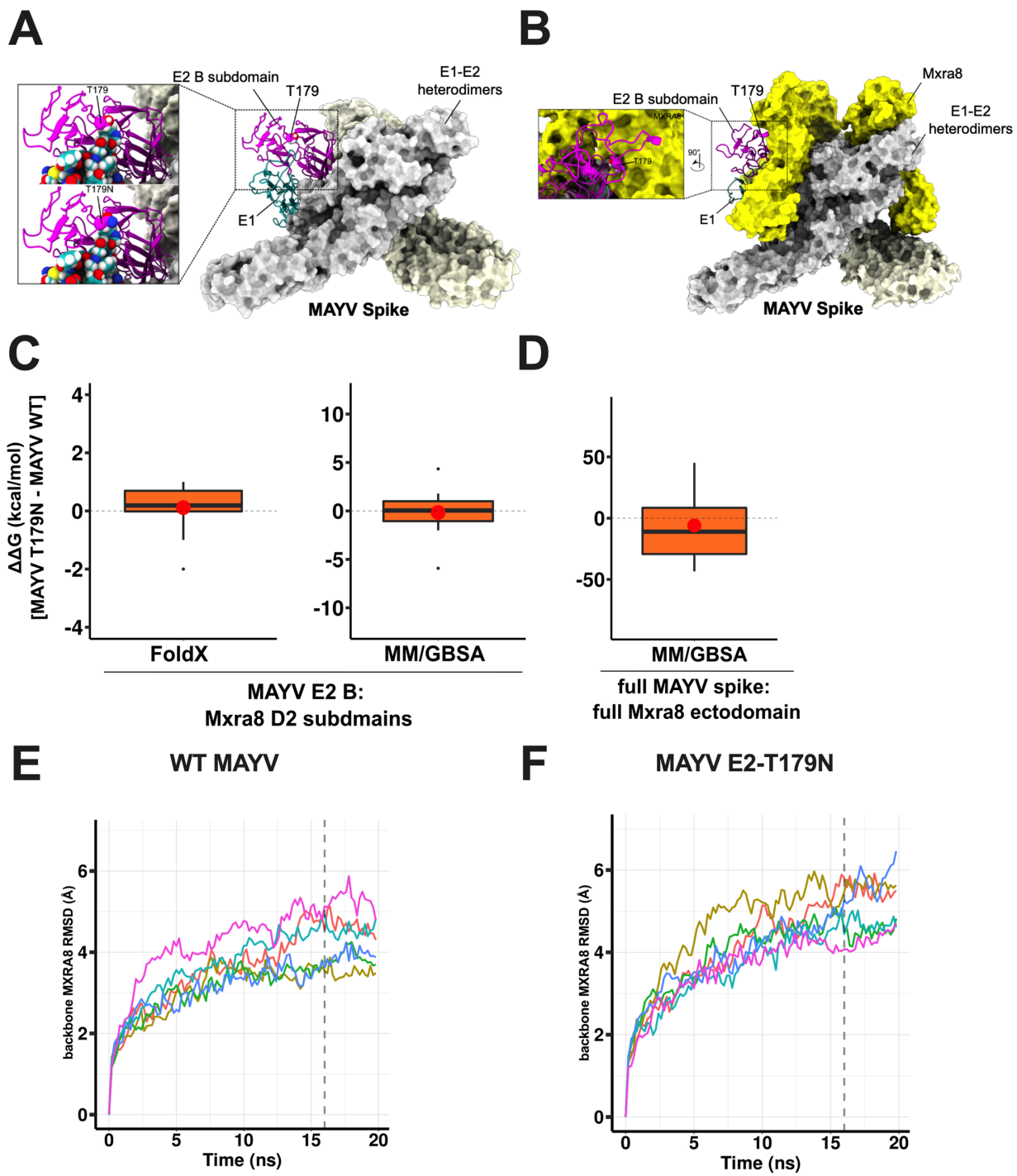


Figure S7. Structural analysis of WT MAYV and MAYV E2-T179N Spike and complexation with Mxra8. A: MAYV spike structure (PDB ID: 7KO8) composed of three E1-E2 heterodimers. Only the E1-E2 ectodomains are presented. For clarity, we showed only one heterodimer (E1 in

green and E2 in magenta) in cartoon representation, the other two are represented as a surface. Zoomed-in views of T179 and the T179N mutation are shown in insets. T179, T179N and the E1 protein are represented in spheres. **B**: MAYV spike in complex with three copies of the human Mxra8 receptor (in yellow) obtained from the last frame of the 20 ns molecular dynamics simulations. For clarity, only one E1-E2 heterodimer is represented as a cartoon. Inset shows a side view of the interface between the E2 B subdomain and the Mxra8 receptor. T179 is shown in spheres. All images were generated using the ChimeraX software. **C:** Binding free energy changes (ΔΔG in kcal/mol) of 12 MAYV E2 B:Mxra8 D2 subdomain complexes (n = 12) upon MAYV T179N mutation using FoldX and MM/GBSA methods. ΔΔG was obtained by subtracting the ΔG estimated for T179N MAYV:Mxra8 interaction from the ΔG obtained for WT MAYV:Mxra8 interaction. MMGBSA calculations were performed from 2 ns molecular dynamics simulation (in triplicate) of each complex. In none of the applied methods, did the estimated ΔΔG significantly differ from zero ($p > 0.05$ - one sample t-test). The box plot central line indicates the median and the red circle indicates the mean. **D**: Binding free energy changes (ΔΔG in kcal/mol) of the full trimeric MAYV spike in complex with the full Mxra8 ectodomain upon MAYV T179N mutations using MM/GBSA method. The calculations were performed from a quintuplicate of 20 ns molecular dynamics simulations (n = 5) (see methods for details). ΔΔG was obtained by subtracting the relative ΔG estimated for T179N MAYV:Mxra8 interaction from the ΔG obtained for WT MAYV:Mxra8 interaction. The estimated ΔG did not significantly differ from zero ($p > 0.05$ - one sample t-test). The box plot central line indicates the median and the red circle indicates the mean. **E-F:** Root-mean-square deviation (RMSD) of Mxra8 receptor in complex with (**E**) WT MAYV spike or (**F**) MAYV T179N during 20 ns molecular dynamics simulations. To calculate the RMSD of Mxra8 backbone atoms, the trajectory was aligned by the MAYV backbone atoms

using the first frame as reference in VMD. Each line corresponds to a replicate and the vertical dashed line indicates the last 4 ns of the simulation.

Chapter 3: Genetic determinants of spillover of SARS-CoV-2 to humans

Evolution At Spike Protein Position 519 in SARS-CoV-2 Facilitated Adaptation to Humans

Cereghino C.[1,2,*], Michalak K.[3*], DiGiuseppe S.[3], Guerra J.[3], Yu D.[3], Faraji A.[3], Sharp A.K.[4], Brown A.M.[4,5], Kang L.[3,6,7], Weger-Lucarelli J.[1,2,§], Michalak P.[3,7,8,§]

[1]*Department of Biomedical Sciences and Pathobiology, Virginia Tech, Blacksburg, VA*

[2]*Center for Emerging, Zoonotic and Arthropod-borne Pathogens, Virginia Tech, Blacksburg, VA*

[3]*Department of Biomedical Research, Edward Via College of Osteopathic Medicine, Monroe, LA*

[4]*Department of Biochemistry, Virginia Tech, Blacksburg, VA*

[5]*Research and Informatics, University Libraries, Virginia Tech, Blacksburg, VA*

[6]*College of Pharmacy, University of Louisiana Monroe, Monroe, LA*

[7]*Center for One Health Research, VA-MD College of Veterinary Medicine, Blacksburg, VA*

[8]*Institute of Evolution, University of Haifa, Haifa, Israel.*

*These authors contributed equally to this work.

§Co-corresponding authors.

Abstract

As the COVID-19 pandemic enters its fourth year, the pursuit of identifying a progenitor virus to SARS-CoV-2 and understanding the mechanism of its emergence persists, albeit against the backdrop of intensified efforts to monitor the ongoing evolution of the virus and the influx of new mutations. Surprisingly, few residues hypothesized to be essential for SARS-CoV-2 emergence and adaptation to humans have been validated experimentally, despite the importance that these mutations could contribute to the development of effective antivirals. To remedy this, we searched

for genomic regions in the SARS-CoV-2 genome that show evidence of past selection around residues unique to SARS-CoV-2 compared with closely related coronaviruses. In doing so, we identified a residue at position 519 in Spike within the receptor binding domain that holds a static histidine in human-derived SARS-CoV-2 sequences but an asparagine in SARS-related coronaviruses from bats and pangolins. In experimental validation, the SARS-CoV-2 Spike protein mutant carrying the putatively ancestral H519N substitution showed reduced replication in human lung cells, suggesting that the histidine residue contributes to viral fitness in the human host. Structural analyses revealed a potential role of Spike residue 519 in mediating conformational transitions necessary for Spike prior to binding with ACE2. Pseudotyped viruses bearing the putatively ancestral N519 also demonstrated significantly reduced infectivity in cells expressing the human ACE2 receptor compared to H519. ELISA data corroborated that H519 enhances Spike binding affinity to the human ACE2 receptor compared to the putatively ancestral N519. Collectively, these findings suggest that the evolutionary transition at position 519 of the Spike protein played a critical role in SARS-CoV-2 emergence and adaptation to the human host. Additionally, this residue presents as a potential drug target for designing small molecule inhibitors tailored to this site.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of coronavirus disease 2019 (COVID-19), which has claimed just under seven million lives and caused 770 million cases of the disease since its emergence in 2019[1]. SARS-CoV-2 is a positive sense, single-stranded RNA virus of the family *Coronaviridae*, genus *Betacoronavirus*, and subgenus *Sarbecovirus*[2]. The first human cases were detected in the city of Wuhan in the Hubei

province of China[3,4] and were associated with the Huanan wholesale seafood market[5,6], where the virus may have spilled over from a live animal source[7,8].

While not definitive, there is evidence of a bat origin of the progenitor to SARS-CoV-2[9]. Sarbecoviruses isolated from bats have high nucleotide sequence similarity to SARS-CoV-2 on a genome level and within the receptor-binding domain (RBD) of the Spike protein[4,10]. BANAL-20-52, isolated from a bat in Laos has the highest nucleotide similarity with SARS-CoV-2 to date at 96.8%[10]. Hypotheses posited the progenitor to SARS-CoV-2 evolved in bats and was pre-adapted to humans since positive selection was detected on deep branches of the nCoV clade and not terminal branches leading to SARS-CoV-2 on a phylogenetic tree[9,10]. However, the precise molecular mechanisms of adaptation of the progenitor to SARS-CoV-2 to humans have not been well characterized despite how these data would aid in the development of therapeutics and surveillance for pandemic prevention. Our previous study determined evolution towards an alanine at position 372 in the SARS-CoV-2 Spike protein was important for human adaptation by first identifying selective sweeps in the SARS-CoV-2 genome [11]. Selective sweeps occur when a favorable mutation is quickly fixed in the population[12]. Neighboring "hitch-hiker" alleles also increase in frequency as a result of the driver mutation of the selective sweep on the recombinant genomic region[13,14] Thus, parsing out truly adaptive mutations requires experimental validation and is critical to determining mechanisms of emergence.

Many studies investigating the emergence of SARS-CoV-2 have strictly involved computational analyses or the generation of pseudotyped viruses expressing Spike. These methods are useful but do not validate predictions or provide the most accurate quantitation of fitness of SARS-CoV-2 compared to examining phenotypes of replication-competent virus. Here, we gain new insights into the evolutionary pressures and molecular mechanisms behind the

emergence of SARS-CoV-2 in humans by generating both pseudotyped and replication-competent virus with a putative ancestral residue consistently mapping to a selective sweep region, as based on the analysis of approximately two million SARS-CoV-2 genomes. The mutation (Spike H519N) significantly decreased SARS-CoV-2 replication in human lung epithelial cells and reduced infectivity in pseudotyped virus assays and binding to human ACE2 in a biochemical assay, consistent with structural predictions. This study provides evidence that the evolution at Spike site 519 was important for the progenitor of SARS-CoV-2's adaptation to the human host. The Spike site also remains highly conserved among SARS-CoV-2 sequences and offers a suitable candidate for targeting with small molecule inhibitors.

Results

SARS-CoV-2 sequences hold evidence of a past, strong selective event in the receptor binding domain of Spike

To investigate which regions in SARS-CoV-2 have contributed to its efficient infection and circulation in humans during the initial stages of the pandemic, we analyzed sequences for signs of strong positive selection events known as selective sweeps. We acquired 1,912,191 human-isolated SARS-CoV-2 sequences from GISAID, covering the period up to August 2021 to constitute an early period of the pandemic. We then used a computational pipeline combining OmegaPlus[15] and RAiSD[16] to identify regions with a high probability of selective sweep events. To discern the selective sweep regions instrumental in driving mutations during the early stages of human infection, we categorized the sequences by their sample collection dates, arranging them into monthly cohorts (**Figure 1A**). Sweep regions containing Spike D614G and Spike T372A, two human-adaptive mutations[11,17], were examined to ensure the reliability and accuracy of our methods. Our analysis identified an average of 11 sweep regions per month, with the

numbers ranging from 4 to 16 (Supplementary file 1). While many sweep regions were identified, our attention was primarily drawn to the sweep regions in the receptor-binding domain (RBD) of the Spike protein (nucleotide position 22,878 to 23,332; amino acid positions 439-590), as mutations in the RBD are known modulators of host tropism and receptor binding[18]. The selective sweep regions included the A372 residue in the RBD, a critical factor we previously identified for the emergence of SARS-CoV-2 and its sustained transmission among humans[11]. These regions were evident in the early months of the COVID-19 outbreak (**Figure 1A**).

Since selective sweep regions indicate a past selective event driven by various adaptive mutations and their associated hitchhiker mutations, we sought to determine the site within the Spike selective sweep region where pivotal mutations occurred upon early infection of humans[19]. While the progenitor virus to SARS-CoV-2 is unknown, we inferred the progenitor Spike sequence could assume several amino acid identities found at the corresponding site in Spike from closely related sarbecoviruses infecting bats and pangolins. To identify the closest coronaviruses related to SARS-CoV-2, we constructed a phylogenetic tree based on the amino acid sequence of Spike with PhyML (**Figure 1B**)[20,21]. Using these phylogenetic relationships, we aligned the amino acid sequence of the selective sweep region in Spike from the Wuhan-Hu-1, lineage B strain of SARS-CoV-2 with the homologous region in closely related sarbecoviruses to reveal sites with differential amino acid identities (**Figure 1C**; see also **Supplementary Figure 1** for the full amino acid alignment of the sweep region in Spike). Similar to amino acid position 372 in Spike, position 519 holds a non-synonymous mutation (nucleotide substitution C23117A) that differentiates SARS-CoV-2 from the aligned bat- and pangolin-derived *Sarbecovirus* sequences[11]. This position was identified in the selective sweep regions during one of the early

months of the outbreak (January 2021; **Figure 1A**). SARS-CoV-2 has a histidine at 519, while the bat and pangolin-derived sequences bear an asparagine or lysine. We hypothesized the progenitor to SARS-CoV-2, potentially a virus infecting bats or pangolins, acquired a histidine at some point in the evolutionary timeline, thereby gaining an adaptive advantage in humans.

Since sites with low amino acid diversity may implicate a crucial role of the predominant amino acid in viral fitness, we sought to investigate the amino acid diversity at position 519 in Spike in SARS-CoV-2. Using data from 3,848 genomes sampled between December 2019 and October 2023 from the nCoV GISAID dataset displayed on Nextstrain, we determined Spike position 519 in SARS-CoV-2 has a normalized Shannon entropy of 0, suggesting little to no flexibility is allowable at this position in humans (**Supplementary Figure 2**)[22,23]. With this result, we hypothesized the predominant histidine at Spike 519 in SARS-CoV-2 may hold an important function for the virus in humans, while the ancestral residue, asparagine, would result in deleterious effects on fitness in human cells.

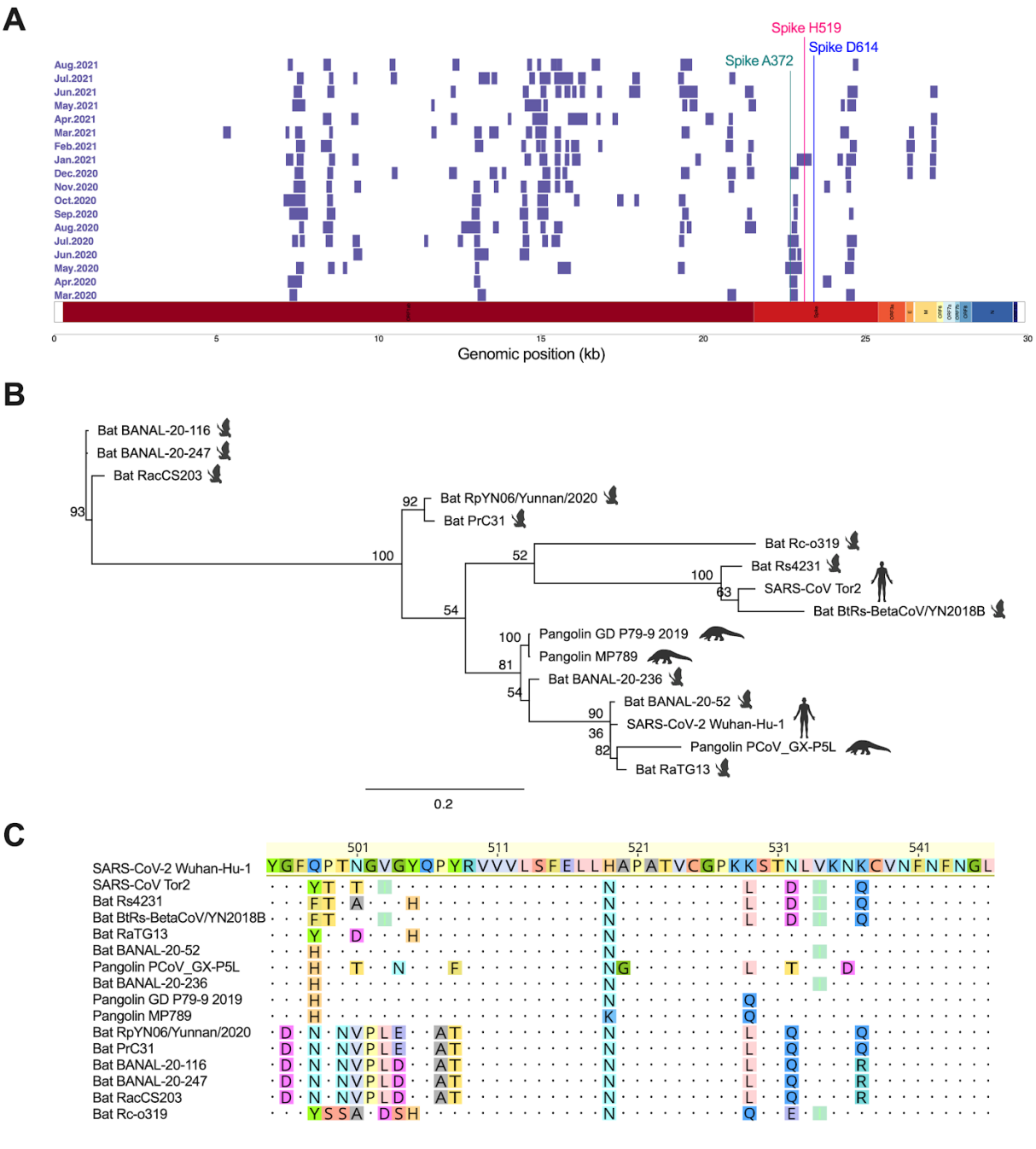


Figure 1. Analysis of millions of SARS-CoV-2 sequences identified selective sweep regions within Spike. A. Selective sweep regions in 2 million SARS-CoV-2 genomes were identified per month up to August 2021 to identify signatures of recent positive selection during the initial stages of the pandemic . Sweep regions are denoted in purple. Results are summarized for the 2

million sequences. Highlighted sites correspond to the site of interest and positive controls Spike T372A and Spike D614G. **B.** Phylogenetic tree of SARS-CoV-2 and SARS-related coronaviruses constructed from the Spike amino acid sequence using PhyML LG substitution model and 1000 bootstraps. The species from which the virus was isolated is depicted by an image of the animal to the right of the virus name. **C.** Amino acid alignment of sweep region in SARS-CoV-2 RBD with corresponding region in closely related sarbecoviruses from bats and pangolins and SARS-CoV. The alignment was performed using MAFFT.

SARS-CoV-2 Spike mutant bearing a residue of bat and pangolin *Sarbecovirus* origin has reduced replicative fitness and infectivity in human cells

Towards identifying the driving mutations of the putative selective event, we reasoned evolution from the asparagine to histidine at position 519 may have driven early adaptation of the progenitor virus to humans. Since position 519 is in the RBD of Spike, we first used a previously described pseudovirus system[24] to determine whether Spike H519N had any functional significance during infection of cells expressing human ACE2 (hACE2), the receptor for SARS-CoV-2[25]. First, we verified hACE2 protein levels by western blot expressed in human embryonic kidney cells (**Supplementary Figure 3**). We generated lentiviruses pseudotyped with full wild-type (WT) Spike, Spike H519N, Spike D614G, and Spike A372T, a RBD mutant bearing an ancestral threonine which we previously showed attenuates the virus in human lung epithelial cells[11]. Spike D614G is used as a known human-adaptive variant that emerged early in the pandemic and is now fixed in all SARS-CoV-2 variants[17]. The pseudoviruses express both luciferase and a green fluorescence protein (GFP), ZsGreen[24], allowing for sensitive detection of infection efficiency. We detected Spike protein levels from prepared virus stocks and observed a greater incorporation of G614 into the pseudoviruses as previously observed (**Figure 2A**)[26].

Ectopic expression of Spike was similar between WT Spike and Spike H519N. Next, we used pseudovirus particles to infect human embryonic kidney cells expressing hACE2 to determine whether infectivity through hACE2 is altered by Spike H519N. Spike D614G enhanced the infectivity of the pseudotyped viruses compared to WT Spike, in agreement with data extensively reported in literature (**Figure 2B**)[26]. Spike A372T significantly reduced the infectivity of SARS-CoV-2, consistent with our previous study (**Figure 2B**)[11]. Importantly, Spike H519N significantly reduced the infectivity of SARS-CoV-2 compared to Spike D614G and WT Spike (**Figure 2B**). When infected cells were quantified based on luciferase expression, we also observed significant decreases in infectivity for Spike H519N compared to WT Spike (**Figure 2C; p< 0.0001**).

Towards determining whether Spike N519 reduces the replicative fitness of SARS-CoV-2 in human cells, we generated a replication-competent, SARS-CoV-2 mutant bearing an asparagine at position 519 in Spike, the amino acid present in closely related sarbecoviruses. In human lung epithelial cells, the putatively ancestral SARS-CoV-2 Spike H519N mutant replicated to significantly lower titers than the WT virus and Spike D614G, demonstrating a nearly 2-log difference in viral titers (**Figure 2D; p<0.01 at all timepoints**). These data suggest that reverting the histidine at Spike 519 to the ancestral asparagine significantly reduces infection and replication in human lung epithelial cells.

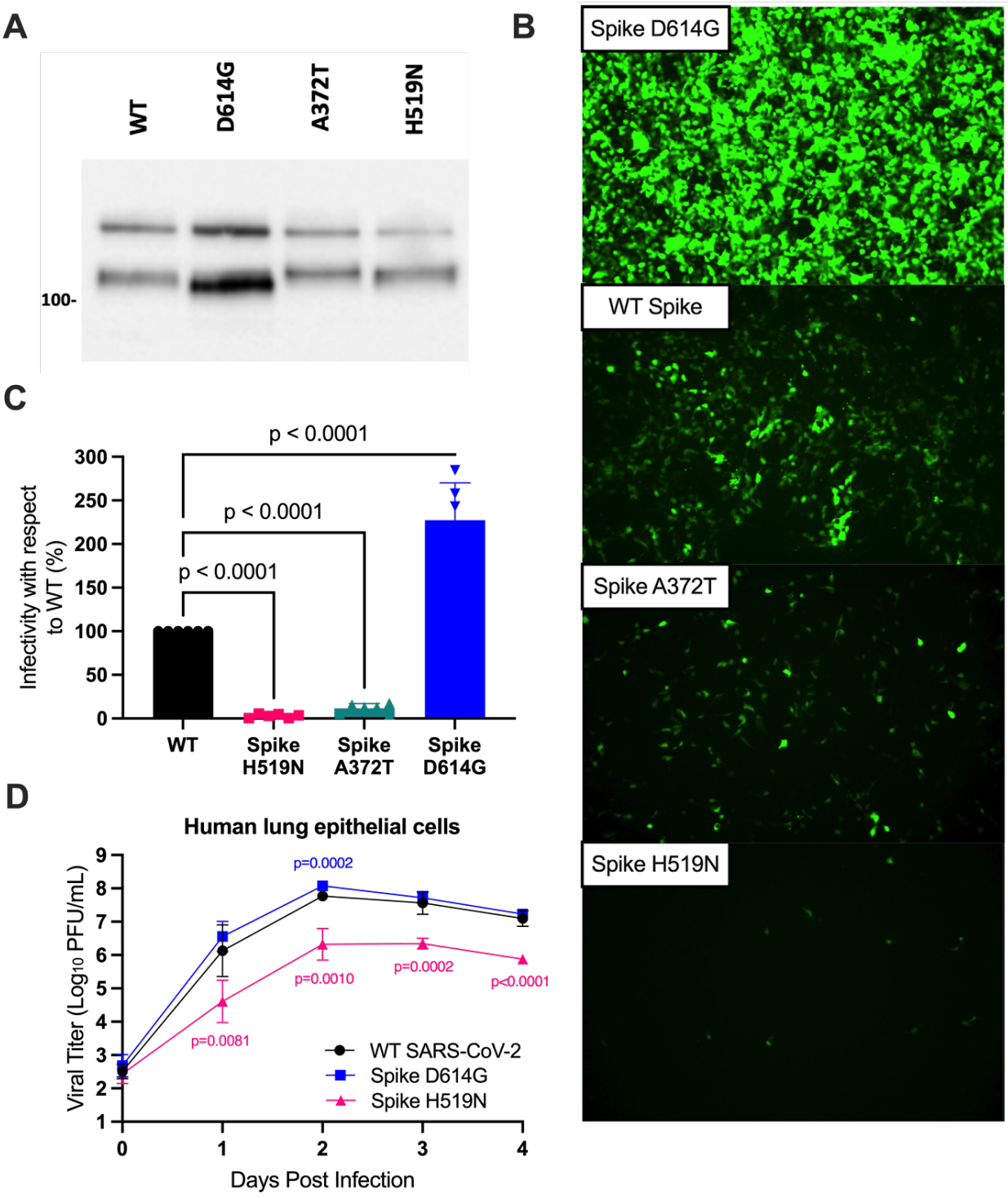


Figure 2. Decreased replicative fitness and infectivity of SARS-CoV-2 Spike H519N in human lung and kidney cells. A. Western blot analysis of Spike protein from prepared

pseudovirus stocks. An antibody was used to detect the S1 subunit of Spike. The higher band at 180 kilodaltons (kDa) corresponds to full Spike while the band near 100 kDa corresponds to furin-cleaved Spike. Units are represented in kDa. **B.** Representative images of Zsgreen expression in Spike pseudotyped virus infected HEK-293T-hACE2 cells observed by live-cell fluorescence microscopy. Cells infected with pseudoviruses expressing ZsGreen and luciferase pseudotyped with full Wuhan-Hu-1 Spike or mutants Spike D614G, A372T, or H519N. Images were taken 48h post-infection and are representative of one of two technical replicates from three independent experiments **C**. Infectivity was quantified using a luciferase assay 48h post-infection. Experiments were performed in duplicate for three biological experiments. Statistical analyses were performed using a One-Way ANOVA with Dunnett's correction for multiple comparisons. **D.** Growth curve of Spike H519N in human lung epithelial cells. Calu-3 cells were infected at a MOI of 0.1 with the WT Wuhan-Hu-1 strain of SARS-CoV-2 and mutants bearing either Spike H519N or D614G. Supernatant was collected each day post-infection and titered on Vero E6-TMPRSS2-T2A-hACE2 cells by plaque assay. Infections were performed in triplicate in two independent biological experiments. A two-way ANOVA with Dunnett's correction for multiple comparisons was performed on these data. Bars represent standard deviation.

Spike H519N potentially impacts ability to conformational transitions and reduces binding affinity to human ACE2

In elucidating the mechanism of attenuation of SARS-CoV-2 bearing the putatively ancestral asparagine at Spike position 519, we analyzed the interactions between Spike chains to identify how H519N is impacting infection (**Figure 3A**). Residue 519 is in the RBD but not within the receptor binding motif (RBM) to interact with ACE2 directly. Rather, residue 519 is positioned on a cleft at the interface between Spike chains that constitute the full trimer complex. Indeed,

residue 519 is within 4-5 Å of residues of adjacent chains of the Spike trimer, participates in polar interchain interactions, and conformational up/down movement can be impacted at this interface. Here, we sought to utilize this interaction interface to compute interchain interaction energy via MM/GBSA calculations analyzing up/down conformation favorability by probing interchain interaction energy in Spike. Results demonstrate that the H519 up conformation has similar interchain interaction energy to the N519 up conformation when unprotonated (-284.1 kcal/mol and -264.1 kcal/mol, respectively). The down conformation exhibits larger differences between the H519 and N519 structures, with H519 interchain interaction energy of -259.2 kcal/mol and N519 of -365.9 kcal/mol. Interestingly, when analyzing protonated H519, we observe a predicted interchain interaction energy for the up conformation H519 of -359.4 kcal/mol and -280.9 kcal/mol for the down confirmation of H519. This suggests that in a lower physiological pH, the H519 Spike samples the up conformation more favorably compared to the N519, which alternatively energetically favors the down conformation based on interchain interactions. Structural analysis of the neighboring residues reveals that 519 is surrounded by a pocket of polar and charged residues from a neighboring Spike chain (**Figure 3B-C**). Surface mapping of residue properties highlights that H519 exhibits a neutral surface area (**Figure 3D**) while N519 results in a more polar surface area (**Figure 3E**). These results suggest that H519 may result in a lower energy barrier to overcome when transitioning between down/up positioning to bind ACE2, potentially allowing for increased infection efficiency.

With the insight that H519N appears to impact conformational transitions in protonated SARS-CoV-2 Spike and the potential ability to bind ACE2 by sampling the up position more favorably, we further sought to determine whether Spike H519 would experimentally bind with higher affinity to hACE2. Enzyme-linked immunosorbent assay (ELISA) was performed with

hACE2 and several concentrations of RBD from either WT Spike, Spike N501Y, Spike A372T, or Spike H519N. Spike N501Y was used as a positive control with known increased binding to hACE2[27]. Consistent with reduced binding efficiency, the absorbance curve for Spike H519N and Spike A372T RBD are shifted right from WT RBD (**Figure 3F**). We observed significantly lower EC_{50} values for Spike N501Y RBD and significantly higher EC_{50} values for Spike H519N and Spike A372T compared to WT RBD (**Figure 3G**). This result suggests more molecules of Spike H519N RBD would be required to saturate hACE2 binding sites; therefore, this mutation may reduce the affinity to hACE2 leading to reductions in replication.

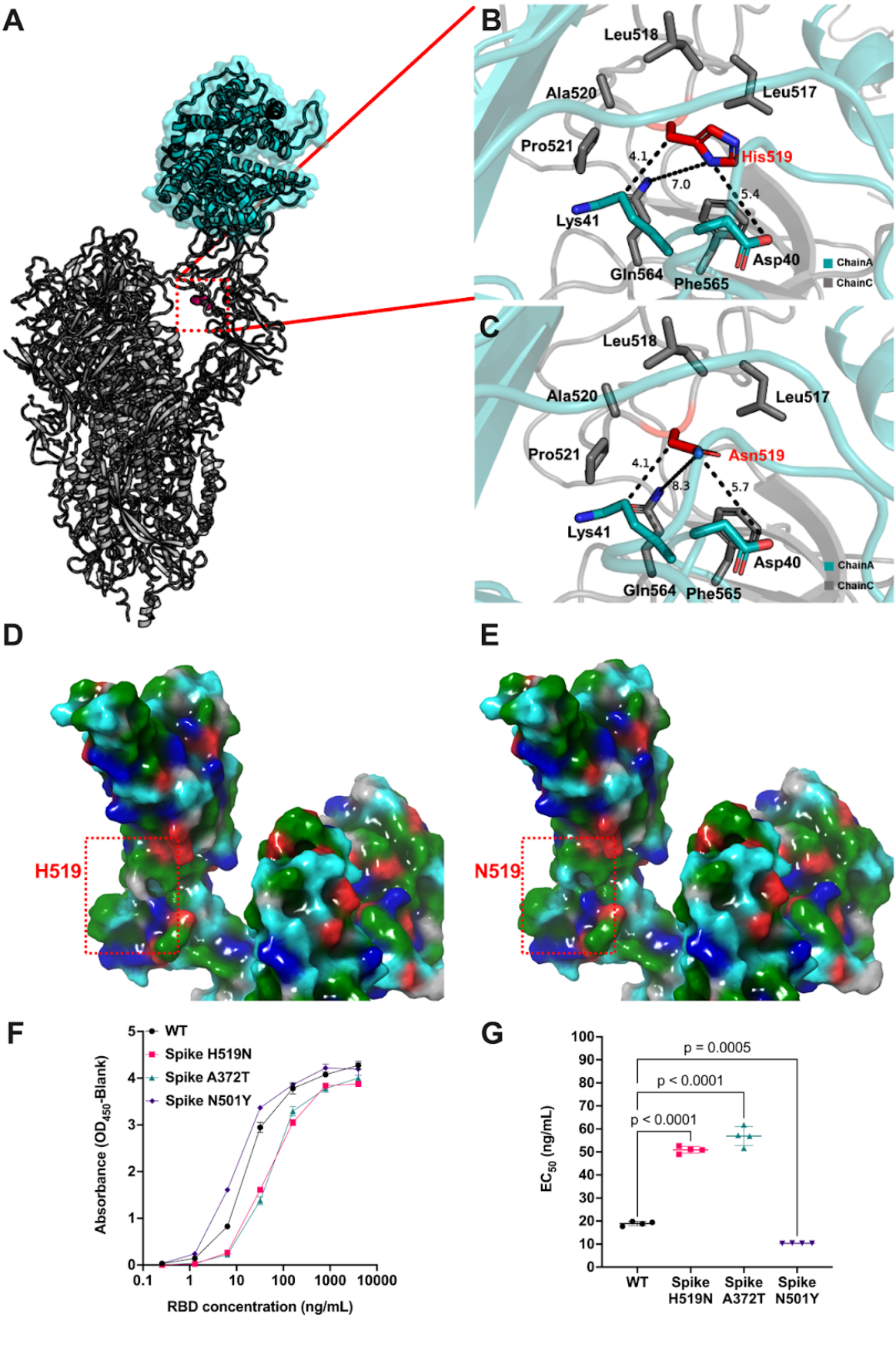

A
B
Leu518
Ala520
Leu517
Pro521
His519
4.1
7.0
5.4
Lys41
Asp40
Gln564
Phe565
ChainA
ChainC
C
Leu518
Ala520
Leu517
Pro521
Asn519
4.1
8.3
5.7
Lys41
Asp40
Gln564
Phe565
ChainA
ChainC
D
H519
E
N519
F
WT
Spike H519N
Spike A372T
Spike N501Y
Absorbance (OD450-Blank)
RBD concentration (ng/mL)
G
p = 0.0005
p < 0.0001
p < 0.0001
EC50 (ng/mL)
WT
Spike H519N
Spike A372T
Spike N501Y

Figure 3. Structural and biochemical analyses of Spike H519N reveal decreased affinity to human ACE2. A. Structure of Spike bound to hACE2 (PDB: 7KNB). 519 location is represented in pink spheres**. B.** Visualization of H519 and **C.** N519 structural location. 519 is highlighted in red, with residues within the same chain in gray and a neighboring chain in teal. **D.** Surface mapping of Spike H519 and **E.** N519 colored by residue sidechain properties. Colors represent gray for neutral, teal for polar uncharged, blue for positive charge, red for negative charge, and green for hydrophobic. **F.** ELISA saturation curves. Human ACE2-mFc was coated at 2 ug/mL, and various concentrations of Spike RBD were detected with an antibody. Absorbance values were measured at OD_{450}nm, and the values normalized to the blank are depicted. Curves intersect the mean of four binding events. **G.** EC_{50} concentrations from ELISA curves extrapolated with a nonlinear regression with a least squares fit. A One-Way ANOVA with Dunnett's Correction was used to test for statistical significance.

Discussion

During zoonotic spillover events, selective pressures exist for pathogens to adapt to the cellular components and immune responses of new host species in the pursuit of optimal fitness. RNA viruses like SARS-CoV-2 are skilled at navigating the fitness landscape in new hosts, taking advantage of both recombination and the error-prone RNA-dependent RNA polymerase to source new genotypes that can lead to fitness advantages[28]. Our study provides evidence of historical selection acting on SARS-CoV-2 via selective sweeps and assesses the contribution of an amino acid site within a selective sweep region in the RBD of Spike that likely facilitated adaptation to humans. We provide evidence that the putatively ancestral asparagine at Spike 519 in closely related sarbecoviruses reduces the fitness of SARS-CoV-2 in cells expressing hACE2

due to reduced affinity of Spike RBD N519 to hACE2 and higher binding free energy in the up conformation.

Several studies have gathered strong evidence for a natural origin of SARS-CoV-2, but there is disagreement on how SARS-CoV-2 might have evolved to sustain productive infection of and transmission between humans. Some studies have suggested the progenitor to SARS-CoV-2 evolved in bats and required little to no adaptation to humans before spillover and emergence[10]. While we do not refute evolution of the progenitor to SARS-CoV-2 occurred in bats, our study provides evidence of a mutation that enhanced human infection but does not predict in which host this mutation arose. It is possible that genetic drift led to mutations in the reservoir host which enabled human adaptation prior to human exposure[9]. Furthermore, our study indicates the transition at Spike 519 from N to H in SARS-CoV-2 evolution led to a gain of fitness in humans, but a fitness tradeoff in putative ancestral host species is not necessarily implicated nor required for adaptation of Spike H519 to humans to remain true. For example, transitions in RBD residues of SARS-CoV isolated from a palm civet to the homologous residue in human-isolated SARS-CoV sequences resulted in a gain of function with infection via hACE2 but no corresponding fitness loss with infection via palm civet ACE2[18]. Therefore, viral adaptation to one species does not require a fitness loss in another host. This study is the second of ours that supports the theory that the progenitor to the Wuhan-Hu-1 strain of SARS-CoV-2 was subject to a strong selective pressure resulting in adaptation to humans at a site in the RBD of Spike[11].

Zoonotic spillover events are often associated with the selection of mutations in viral receptor binding proteins that afford the use of orthologous receptors in the human host[29]. For coronaviruses in particular, mutations in the Spike protein, which binds the host ACE2 receptor,

are sufficient to expand the host range of the virus even though a variety of other factors are important for tropism[18]. Another zoonotic coronavirus, severe acute respiratory syndrome coronavirus (SARS-CoV), which caused an epidemic from 2002-2003, emerged likely due to a set of mutations both in the receptor binding domain (RBD) and other parts of Spike and underwent adaptive evolution from its reconstructed ancestral sequence[18,30]. Our results are in agreement with other studies that mutations outside of the RBM of the coronavirus Spike protein, including both subunits of Spike, can affect receptor binding and are also associated with host-range expansion[11,31,32].

Amino acid position 519 in Spike is in the RBD but not within the receptor binding motif (RBM). The putatively ancestral asparagine has more favorable co-association of Spike in the down conformation, which impacts the potential to transition to an up conformation and bind with hACE2. This can be observed with the reduced binding affinity of Spike H519N RBD to hACE2 via ELISA and interchain interaction energy calculations highlighting the H519 up conformation favorability in an acidic environment. While these results provide insight into potential structural mechanisms related to the fitness of H519, more robust molecular dynamics (MD) simulations could reveal more major morphological changes related to 519 fitness in transiting between up/down conformations. Our results are not surprising since mutations outside of the RBM have been shown to increase binding affinity to hACE2 during a deep mutational scan of the Spike RBD[33], and our previous work indicates Spike 372 influences hACE2 binding[11] . Furthermore, Spike 614, which lies outside of the RBD, affects ACE2 binding[34]. Previous work has indicated the structural region surrounding residue 519 is involved in pH sensing[35] and structural transition states[36] which can also ultimately affect infectivity. Our results are biologically relevant given that the human nasal cavity pH is 6.6 which might provide a more

optimal environment for the protonated SARS-CoV-2 Spike H519 to associate more with hACE2 than Spike N519 in this acidic environment[37]. Since transmission of coronaviruses between bats is suspected to primarily occur via the fecal-oral route[38,39], the microenvironment of the bat gastrointestinal tract rather than the bat nasal cavity likely has more influence on coronavirus transmission between bats. More research is needed on this microenvironment and how it may shape bat coronavirus evolution. Taken in an evolutionary context, it is possible that Spike N519H arose in the progenitor virus due to stochastic or deterministic processes and that this mutation increased association with hACE2 when the virus initially encountered humans, leading to enhanced infection and transmissibility.

We used Nextstrain to examine the diversity within Spike and identified low normalized Shannon entropy approaching 0 at position 519 in Spike of SARS-CoV-2 sequences from 2019-2023 from GISAID. Since we have shown the histidine at position 519 in Spike is important for replication in human lung cells and that it is highly conserved regardless of lineage and variant of concern, position 519 might constitute a suitable drug target. Indeed, structural inspection of the surface area maps and structural residue region of 519 indicates that it is within a solvent accessible pocket, surrounded by polar and charged residues from neighboring chains. Suitable pockets for small molecule inhibitors may also correlate with regions of low diversity in the genome, however further investigation is needed. Efforts to design small molecule inhibitors of SARS-CoV-2 have resulted in targeting genes such as the main protease of SARS-CoV-2[40]. Recent surveillance studies identified multiple mutations in SARS-CoV-2 that confer resistance to nirmatrelvir[41]. Another protease inhibitor ensitrelvir is being used for emergency use authorization in Japan, but mutations conferring resistance to this inhibitor have been identified as well[41]. This highlights the importance of finding new components of the virus to target, and the

RBD at position 519 in Spike might be suitable since other amino acids at this position have not emerged to any appreciable extent.

To conclude, a mutation altering the amino acid identity at position 519 in the Spike protein of SARS-CoV-2 is likely to have partly mediated adaptation of the progenitor virus to the human host. We demonstrated the conserved histidine at position 519 in Spike affords fitness advantages through increased replication in human cells via increased hACE2 binding affinity. Position 519 in Spike holds promise as a site to which small molecular inhibitors could be designed amidst the concern of the development of drug resistance.

Methods

Selective sweep analysis of SARS-CoV-2 genomes

A total of 1,914,191 human-derived, complete SARS-CoV-2 genomes up to Aug. 2021 were downloaded from the GISAID EpiCov database (https://www.gisaid.org/). Sequences that had low coverage or displayed over 5% nucleotide ambiguities (designated as 'N') were excluded from the analysis to ensure analytical rigor. For further analysis, sequences were categorized monthly based on their sample collection dates. OmegaPlus and RAiSD were used with these data for the determination of selective sweep regions. OmegaPlus[15] is an implementation of the ω statistic[42] for whole-genome data, which takes into account linkage disequilibrium, whereas RAiSD takes into account three signatures of selective sweeps[16]. Selective sweep identification was conducted within these monthly cohorts using previously established data curation and methodologies[11].

Construction of phylogenetic tree

The following coronavirus sequences were downloaded from GenBank: Bat coronavirus isolate BANAL-20-52/Laos/2020 MZ937000.1, Bat coronavirus isolate BANAL-20-103/Laos/2020, complete genome MZ937001.1, Bat coronavirus isolate BANAL-20-116/Laos/2020, complete genome MZ937002.1, Bat coronavirus isolate BANAL-20-236/Laos/2020, complete genome MZ937003.2, Bat coronavirus isolate BANAL-20-247/Laos/2020, complete genome MZ937004.1, Bat coronavirus RacCS203, complete genome MW251308.1, Bat coronavirus RaTG13, complete genome MN996532, Bat coronavirus strain BetaCoV/Rm/Yunnan/YN02/2019 spike protein (S) gene, partial cds MW201982.1, Bat sarbecovirus sp. isolate PrC31, complete genome MW703458.1, Bat SARS-like coronavirus isolate Rs4231 KY417146.1, Bat Severe acute respiratory syndrome-related coronavirus Rc-o319 RNA, complete genome LC556375.1, Betacoronavirus sp. RpYN06 strain bat/Yunnan/RpYN06/2020, complete genome MZ081381.1, Coronavirus BtRs-BetaCoV/YN2018B MK211376.1, MAG Pangolin coronavirus isolate GD P79-9 2019, complete genomeOQ297708.1, Pangolin coronavirus isolate MP789, complete genome MT121216.1, Pangolin coronavirus isolate PCoV_GX-P4L, complete genome MT040333.1, Pangolin coronavirus isolate PCoV_GX-P5L MT040335.1, SARS coronavirus Tor2 AY274119.3, Severe_acute_respiratory_syndrome_coronavirus_2_isolate_Wuhan-Hu-1_NC_045512.2. Geneious (version 2023.1 created by Biomatters. Available from https://www.geneious.com) was used to construct a phylogenetic tree based on the amino acid sequence of Spike using PhyML 3.3.20180214 with a LG substitution model and 1000 bootstraps[20]. An amino acid alignment of the selective sweep region in Spike identified in January 2021 was performed using Geneious with the same viruses as above to identify non-synonymous mutations compared to the Wuhan-Hu-1 strain of SARS-CoV-2.

Diversity analysis in Spike

Nextstrain was accessed on November 1, 2023 displaying 3,848 nCov genomes sampled between December 2019 and October 2023 from the GISAID nCoV dataset. Amino acid diversity data was obtained by downloading the full Spike diversity dataset from the site. Nextstrain calculates normalized Shannon entropy by summing the products of the amino acid frequency at a given position and the natural log of the frequency for each amino acid observed. The frequency is normalized by the number of tips in the phylogenetic tree on Nextstrain[22].

Cell lines and plasmids

Human embryonic kidney cells expressing human ACE2 (HEK-293T-hACE2; NR-52511) and African green monkey kidney epithelial cells expressing TMPRSS2 and human ACE2 (Vero E6-TMPRSS2-T2A-hACE2; NR-54970) were obtained from BEI Resources. Human lung epithelial cells (Calu-3 HTB-55) and human embryonic kidney cells (HEK-293T) were acquired from ATCC. Cells were grown in a humidified atmosphere with carbon dioxide supplied at 5% at 37°C. HEK-293T and HEK-293T-hACE2 cells were grown in Dulbecco's Modified Eagle Medium (DMEM; Corning™ 10013CV) supplemented with 10% Fetal Bovine Serum, 100 U/mL penicillin, 100 U/mL streptomycin, 2mM L-glutamine. Vero E6-TMPRSS2-T2A-hACE2 and Calu-3 cells were grown in Dulbecco's Modified Eagle Medium supplemented with gentamicin sulfate (0.1%), non-essential amino acids (1X), HEPES (25mM), and either 5% (Vero E6-TMPRSS2-T2A-hACE2) or 20% (Calu-3) fetal bovine serum. Vero E6-TMPRSS2-T2A-hACE2 also required the addition of 0.01 mg/mL puromycin. SARS-Related Coronavirus 2, Wuhan-Hu-1 Spike-Pseudotyped Lentiviral Kit V2 was obtained from BEI Resources (NR-53816).

Generation of Spike pseudotyped virus mutants

Spike mutations A372T and H519N were introduced to the human codon-optimized Spike gene on vector pHDM SARS-Related Coronavirus 2 Wuhan-Hu-1 Spike Glycoprotein (BEI; NR-53742). We used the BioLabs Q5 Site-Directed Mutagenesis Kit (NEB #E0554) as described from the manufacturer's protocol with mutagenesis primers Forward (CGAATTGCTCAACGCTCCAGC) and Reverse (AAACTCAAGACCACAACTCTG) for mutant H519N, and Forward (GTATAATAGTACAAGCTTTAGCACATTC) and Reverse (AATACTGAGTAGTCCGCC) for mutant A372T. Mutations were confirmed by Sanger sequencing. Vector HDM SARS2-Spike-del121 D614G was purchased from Addgene (cat #158762). Pseudotyped lentiviral particles were generated by transfecting HEK-293T cells using PolyJet In Vitro DNA Transfection reagent (SignaGen cat #SL100688) following the manufacturer's recommendations for the six-well plate, with the adjusted amount of 3µg for each of the following BEI plasmids: pRC-CMV-Rev1b (NR-52519), HDM-tat1b (NR-52518), HDM-Hgpm2 (NR-52518), Luciferase-IRES-ZsGreen (NR52516), and 3µg of the pHDM vector encoding the Wuhan-Hu-1 Spike Glycoprotein (WT), or D614G, A372T, or H519N spike mutations, respectively. We monitored the transfection efficiency by detecting ZsGreen signal via live cell fluorescence microscopy using a Leica DMI4000B inverted microscope. Supernatants with lentiviral particles were harvested 48- and 72-hours post-transfection. Pseudoviruses were precipitated overnight at 4◦C with 40% PEG-8000/1.2 M NaCl pH 7.2 at a 1 to 3 ratio, centrifuged at 1200rpm for 1 hour at 4◦C, and obtained pellet resuspended in 1/10 of the original collection volume.

Western blot analysis of pseudotyped virus Spike incorporation

We verified the presence of SARS-CoV-2 Spike protein on WT and mutated pseudotyped lentiviral particles by Western blot analysis. Viral protein lysates were denatured at 95◦C, resolved on a 10% SDS-PAGE gel, and transferred to Biorad Immun-Blot PVDF membrane (cat #1620177). After 1 hour Blocking in 5% milk in TBS-T buffer (50 mM Tris [pH 7.5], 150 mM NaCl containing 0.1% Tween 20), the membrane was incubated overnight with Invitrogen Anti-SARS-CoV2-Spike protein S1 primary antibody (cat #PA5-114528). After three washes with TBS-T buffer, blot was probed for 1 hour with Peroxidase-conjugated AffiniPure Goat Anti-Rabbit IgG (H+L) secondary antibody (Jackson ImmunoResearch Laboratories cat #115-035-003) in 5% milk/TBS-T buffer. The membrane was washed three times with TBS-T, and the signal was detected with Biorad Clarity Western ECL substrate cat (#170-5061) and imaged using Biorad ChemiDoc MP Imaging System.

Quantification of SARS-CoV-2 plasmid pseudogenome via real-time PCR

For the quantification of pseudovirus particles, we extracted the DNA from lentiviral stocks using Qiagen DNeasy Blood & Tissue Kit (cat #69504) according to the manufacturer's protocol. We quantified a ZsGreen amplicon by qPCR via Applied Biosystems PowerSYBR Green PCR Master Mix (cat #4376659) using Forward (GACCATGAAGTACCGCATG) and Reverse (CCGCCCTCCACCACGCAC) ZsGreen-specific primers. The DNA Standard curve of 10 ng/µL, 1 ng/µL, 0.1 ng/µL and 0.01 ng/µL serial dilutions of pHAGE-CMV-Luc2-ZsGreen plasmid served as our reference. The relative amounts of pseudoviral genomes of all mutants were adjusted to WT lentivirus before infections.

Luciferase Assay

The luciferase activity of HEK-293T-hACE2 cells infected with pseudovirus was measured using the Promega Luciferase Assay System according to the manufacturer's protocol (cat #E1500). Cells were washed 2 times with PBS and lysed in 1X Lysis Buffer per well in a 12-well plate. 100 µL of the 1X Luciferase Assay per 20 µL of lysate was used for detection of luciferase activity in each well of a 96-well black microplate. The efficiency of infection was determined by measuring the intensity of luminescence calculated as relative luciferase units using the Agilent BioTek Synergy HTX Multi-Mode Microplate Reader.

Generation of SARS-CoV-2 mutants

Live SARS-CoV-2 mutants were generated from a previously constructed full-length infectious clone of the Wuhan-Hu-1 strain of SARS-CoV-2 (GenBank accession no. NC_045512.2). Mutagenic primers were used to introduce mutations into the backbone by PCR with Invitrogen Platinum SuperFi II PCR master mix (12368010) or Quantabio repliQa HiFi ToughMix (95200-025). Fragments were purified from agarose gels using the Machery-Nagel nucleospin gel and pcr clean-up kit (740609.250). Purified fragments were mixed at an equimolar ratio and assembled and amplified by replication cycle reaction using the OriCiro Genomics Cell-Free Cloning System. Virus was recovered by transfecting a 1:1 ratio of replication cycle reaction product and pUC19[43] into BHK-21 cells using Polyplus jetOPTIMUS DNA transfection reagent (101000051). Three days post-transfection, a blind passage of transfection supernatant on Vero E6-TMPRSS2-T2A-hACE2 cells was performed. After observing cytopathic effect, the virus was harvested, titered by plaque assay on Vero E6-TMPRSS2-T2A-hACE2 cells, and 400bp libraries were prepared in-house using IDT Artic V4.1 NCOV-2019 Panel (10011442) and sequenced by Genewiz NGS Amplicon-EZ Illumina sequencing.

Growth curve of Spike H519N in human lung epithelial cells

Human lung epithelial cells (Calu-3) were infected at 80% confluency with a MOI of 0.1 of SARS-CoV-2 Wuhan-Hu-1 (wild-type), SARS-CoV-2 Spike D614G mutant, and SARS-CoV-2 Spike H519N mutant diluted in Roswell Park Memorial Institute 1640 (RPMI-1640) medium supplemented with fetal bovine serum (2%) and HEPES (10 mM). Supernatant was harvested every 24 hours post-infection and titered by plaque assays on Vero E6-TMPRSS2-T2A-hACE2 cells.

Quantitation of infectious virus

Infectious virus was quantified by plaque assay by infecting 90% confluent monolayers of Vero E6-TMPRSS2-T2A-hACE2 with serial dilutions of virus made in RPMI-1640 medium supplemented with fetal bovine serum (2%) and HEPES (10 mM). After one hour of adsorption at 37°C, a 1:1 mixture of 2X media (2X EMEM, 4X L-glutamine, 0.735% sodium bicarbonate, 0.2 mg/mL gentamicin sulfate, 4% FBS, 20 mM HEPES) and 3% methylcellulose was added to the cells. Infection proceeded for three days at 37°C before fixation in 10% buffered formalin and staining with a 0.1% crystal violet solution containing 20% ethanol to visualize plaques.

Structural Analyses

Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) was used to predict interchain interaction energy with the Schrodinger- Maestro software v. 2020-2 (Schrödinger, LLC, New York, NY, 2020). Spike protein (PDB ID: 7KNB) was split into individual chains and modified to reflect D614G in the Wuhan-Hu-1 strain[44]. Predicted interaction energy was calculated between Spike chains first on one chain in an up-RBD conformation and one chain in a down-RBD conformation to approximate up conformation free energy and two down-RBD

conformation chains for down conformation free energy on H519 (neutral), H519 (protonated), and N519 structures. Protonated and unprotonated states were simulated at a pH of 5 and 7, respectively for the H519. Analysis using IPC 2.0[45] of the Spike sequence indicated that the region around H519 is mostly neutral with acidic propensity, suggesting the increased potential of an acidic environment and lower pH to promote a protonated H519. Surface mapping was performed using Schrodinger- Maestro software v. 2020-2.

Biochemical characterization of Spike with human ACE2

A sandwich ELISA was performed using recombinant, purified RBD of His-tagged wild-type Wuhan-Hu-1 SARS-CoV-2, SARS-CoV-2 Spike N501Y, SARS-CoV-2 Spike H519N, or SARS-CoV-2 Spike A372T and hACE2. A plate was coated with ACE2-mFc (Cat:10108-H05H) at 2 μg/mL and incubated overnight at 4°C. The following day, RBD of each aforementioned genotype was added at the following concentrations: 0.256, 1.28, 6.4, 32, 160, 800, and 4000 ng/mL. The antibody anti-His-HRP (Sino A5327) was used to detect the His-tagged RBDs. Absorbance was measured at 450nm, and values were normalized to the blank.

Statistical Analyses

All statistical analyses for figures 2-3 were performed using Prism 9 (GraphPad). Growth curve data was analyzed using a Two-Way ANOVA with Dunnett's correction for multiple comparisons. All other data were analyzed with a One-Way ANOVA with Dunnett's correction for multiple comparisons.

Data availability statement

Sequencing files of the viruses constructed in the current study are available from the corresponding author on reasonable request. All remaining data generated or analyzed during this study are included in this published article (and its supplementary information files).

Acknowledgements

We extend our thanks to the scientists who performed surveillance and sequencing of sarbecoviruses and SARS-CoV-2 and made the data accessible via GenBank and GISAID and to the team that built and maintains Nextstrain. We would like to acknowledge Sino Biological for performing ELISAs.

Competing interests declaration

The authors of this manuscript do not declare any competing interests.

Author contributions

Conceptualization: P.M., L.K., J.W.L., A.B., S.D.

Data curation: C.C., K.M., A.S., P.M., L.K., J.W.L., A.B., S.D.

Formal analysis: C.C., K.M., A.S., P.M., L.K., J.W.L., A.B., S.D.

Investigation: C.C., K.M., A.S., P.M., L.K., J.W.L., A.B., S.D.

Methodology: P.M., L.K., J.W.L., A.B., S.D.

Validation: C.C., K.M.

Writing – original draft: C.C.

Writing – review & editing: C.C., K.M., A.S., S.D., L.K., A.B., J.W.L., P.M.

Funding acquisition: J.W.L., P.M.

Project administration: J.W.L., P.M.

Resources: J.W.L., P.M., S.D.

Supervision: P.M., J.W.L., A.B., S.D.

All authors read and approved the final manuscript.

Biosecurity

All research protocols were approved by the Virginia Tech Institutional Biosafety Committee prior to beginning experiments. Pseudotyped virus work was conducted at biosafety level 2, while live virus manipulation of SARS-CoV-2 was performed in a biosafety level 3 (BSL3) laboratory at Virginia Tech with appropriate fitting N95 respirators and other standard PPE. SARS-CoV-2 was appropriately inactivated at 65C for 10 minutes after the addition of a viral lysis buffer, as described previously, before removal from the BSL3 for library preparation[46].

References

1. WHO Coronavirus (COVID-19) dashboard. https://covid19.who.int/.
2. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* **5**, 536–544 (2020).
3. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
4. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
5. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

6. Organization, W. H. & Others. WHO-convened global study of origins of SARS-CoV-2: China Part. (2021).
7. Worobey, M. *et al.* The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science* **377**, 951–959 (2022).
8. Holmes, E. C. *et al.* The origins of SARS-CoV-2: A critical review. *Cell* **184**, 4848–4856 (2021).
9. MacLean, O. A. *et al.* Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. *PLoS Biol.* **19**, e3001115 (2021).
10. Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330–336 (2022).
11. Kang, L. *et al.* A selective sweep in the Spike gene has driven SARS-CoV-2 human adaptation. *Cell* **184**, 4392-4400.e4 (2021).
12. Durrett, R. & Schweinsberg, J. Approximating selective sweeps. *Theor. Popul. Biol.* **66**, 129–138 (2004).
13. Nordborg, M., Charlesworth, B. & Charlesworth, D. The effect of recombination on background selection. *Genet. Res.* **67**, 159–174 (1996).
14. Lai, M. M. & Cavanagh, D. The molecular biology of coronaviruses. *Adv. Virus Res.* **48**, 1–100 (1997).
15. Alachiotis, N., Stamatakis, A. & Pavlidis, P. OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics* **28**, 2274–2275 (2012).
16. Alachiotis, N. & Pavlidis, P. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Commun Biol* **1**, 79 (2018).
17. Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).

18. Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* **24**, 1634–1643 (2005).
19. Stephan, W. Selective Sweeps. *Genetics* **211**, 5–13 (2019).
20. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
21. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
22. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
23. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol* **4**, vex042 (2018).
24. Dadonaite, B. *et al.* A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. *Cell* **186**, 1263-1278.e20 (2023).
25. Jackson, C. B., Farzan, M., Chen, B. & Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* **23**, 3–20 (2022).
26. Zhang, L. *et al.* SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* **11**, 6013 (2020).
27. Liu, Y. *et al.* The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature* **602**, 294–299 (2022).
28. Sun, F. *et al.* SARS-CoV-2 Quasispecies Provides an Advantage Mutation Pool for the Epidemic Variants. *Microbiol Spectr* **9**, e0026121 (2021).
29. Warren, C. J. & Sawyer, S. L. How host genetics dictates successful viral zoonosis. *PLoS Biol.* **17**, e3000217 (2019).

30. Yuan, Z., Nan, Z., Pei, H. & Yang, Z. Reconstruction of the most recent common ancestor sequences of SARS-Cov S gene and detection of adaptive evolution in the spike protein. *Chin. Sci. Bull.* **49**, 1311–1313 (2004).

31. Becker, M. M. *et al.* Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19944–19949 (2008).

32. McRoy, W. C. & Baric, R. S. Amino acid substitutions in the S2 subunit of mouse hepatitis virus variant V51 encode determinants of host range expansion. *J. Virol.* **82**, 1414–1424 (2008).

33. Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310.e20 (2020).

34. Ozono, S. *et al.* SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat. Commun.* **12**, 848 (2021).

35. Lobo, V. R. & Warwicker, J. Predicted pH-dependent stability of SARS-CoV-2 spike protein trimer from interfacial acidic groups. *Comput. Struct. Biotechnol. J.* **19**, 5140–5148 (2021).

36. Qu, K. *et al.* Engineered disulfide reveals structural dynamics of locked SARS-CoV-2 spike. *PLoS Pathog.* **18**, e1010583 (2022).

37. Kreutzberger, A. J. B. *et al.* SARS-CoV-2 requires acidic pH to infect cells. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2209514119 (2022).

38. Hu, B. *et al.* Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* **13**, e1006698 (2017).

39. Poon, L. L. M. *et al.* Identification of a novel coronavirus in bats. *J. Virol.* **79**, 2001–2009 (2005).

40. Liu, H. *et al.* Development of optimized drug-like small molecule inhibitors of the SARS-CoV-2 3CL protease for treatment of COVID-19. *Nat. Commun.* **13**, 1891 (2022).

41. Moghadasi, S. A. *et al.* Transmissible SARS-CoV-2 variants with resistance to clinical protease inhibitors. *Sci Adv* **9**, eade8778 (2023).

42. Kim, Y. & Nielsen, R. Linkage disequilibrium as a signature of selective sweeps. *Genetics* **167**, 1513–1524 (2004).

43. Søndergaard, J. N. *et al.* Successful delivery of large-size CRISPR/Cas9 vectors in hard-to-transfect human cells using small plasmids. *Commun Biol* **3**, 319 (2020).

44. Zhou, T. *et al.* Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* **28**, 867-879.e5 (2020).

45. Kozlowski, L. P. IPC 2.0: prediction of isoelectric point and pKa dissociation constants. *Nucleic Acids Res.* **49**, W285–W292 (2021).

46. Castellanos-Gonzalez, A. *et al.* Direct RT-PCR amplification of SARS-CoV-2 from clinical samples using a concentrated viral lysis-amplification buffer prepared with IGEPAL-630. *Sci. Rep.* **11**, 14204 (2021).

Chapter 4: Recurring mutation in SARS-CoV-2 and its impact on infection of human and animal cells

The recurrent Spike A222V mutation in SARS-CoV-2 enhances replication in primary deer lung cells

Chelsea Cereghino[1,2], Kateland Tiller[1,2], Lin Kang[3,4,5], Pawel Michalak[3,5,6], James Weger-Lucarelli[1,2*]

1. Department of Biomedical Sciences and Pathobiology, Virginia Tech, Blacksburg, Virginia, USA
2. Center for Emerging, Zoonotic, and Vector-borne Pathogens, Virginia Tech, Blacksburg, Virginia, USA
3. Department of Biomedical Research, Edward Via College of Osteopathic Medicine, Monroe, LA, USA
4. College of Pharmacy, University of Louisiana Monroe, Monroe, LA, USA
5. Center for One Health Research, VA-MD College of Veterinary Medicine, Blacksburg, VA, USA
6. Institute of Evolution, University of Haifa, Haifa, Israel

*corresponding author

Contact information of corresponding author: Center for One Health Research, 1410 Prices Fork Road, Building 441 Room 113, Blacksburg, Virginia, 24060; weger@vt.edu

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infects humans and animals and is therefore a pathogen of grave concern within a One Health framework. Identifying animal-adaptive mutations is critical to preserving One Health, as these mutations could also lead

to the persistence of SARS-CoV-2 in animal reservoirs with continual spillover to humans. Therefore, we sought to pair experimental evolution and epidemiological data to identify putative human- and animal-adaptive viral residues and determine their impact on replication-competent SARS-CoV-2 in both human and animal cells. We passaged SARS-CoV-2 in cells expressing human, dog, cat, mink, and white-tailed deer ACE2 and sequenced the passaged populations. In addition, we searched SARS-CoV-2 sequences for mutations following patterns of convergent evolution that were common to both human and animal-derived SARS-CoV-2 sequences. We identified the epidemiologically relevant Spike A222V mutation from our passaging experiment in cells expressing cat ACE2, a mutation that has also arisen independently across eight lineages of SARS-CoV-2 from human- and animal-derived sequences. To assess its impact on replication in human and animals cells, we constructed SARS-CoV-2 Spike A222V in the Wuhan-Hu-1 backbone with Spike D614G; this virus replicated similarly to the WT SARS-CoV-2 in human lung epithelial cells. In contrast, SARS-CoV-2 Spike A222V demonstrated an advantage in replication in primary deer lung cells, which was not mediated by the deer ACE2 receptor. Infection via the human, dog, cat, and mink ACE2 receptor resulted in reduced replication of SARS-CoV-2 Spike A222V. Our experiments identified Spike A222V as a putatively deer-adaptive mutation. Future studies should assess Spike A222V's relevance to transmission within deer and to other animal species in contact with deer.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), remains a public health concern five years after the start of the pandemic. Cases of COVID-19 mortality in the United States in 2024 exceeded 47,000, and, as of February 2025, cumulative deaths in the United States since the beginning of the

pandemic exceed 1.22 million[1]. Morbidity is not restricted to humans, as SARS-CoV-2 infects and causes disease in a variety of animals[2]. A total of 775 outbreaks have been reported in animals to date[3] with at least 2,000 animal-derived SARS-CoV-2 genomes sequenced[4]. A notable outbreak of SARS-CoV-2 in animals occurred in 2020 on mink farms in Denmark and the Netherlands. Detection of SARS-CoV-2 in mink was facilitated by observed respiratory symptoms in mink[5], and ultimately led to the culling of large numbers of farmed mink[6]. COVID-19 symptoms have been reported in other species like domesticated cats[7], lions[8], tigers[9], and hamsters[10], including neurodegenerative pathology in canines[11]. Spillback from animals to humans has been reported for mink[12], deer[13], and hamsters[14]; thus SARS-CoV-2 presents serious "One Health" implications.

The broad species tropism of SARS-CoV-2 is attributed primarily to the ability of SARS-CoV-2 to use receptors from multiple species, notably angiotensin-converting enzyme 2 (ACE2). Documented ACE2 usage has been reported for SARS-CoV-2 for various species: rhesus macaques, Mexican free-tailed bats, Chinese horseshoe bats, palm civets, raccoon dogs, ferret badgers, hog badgers, white-tailed deer, mule deer, elk, dogs, cats, mink, rabbit, and pangolins[15–18]. SARS-CoV-2 variants harboring Spike N501Y, including B.1.1.7, B.1.351, P.1, and P.3, also have an ability to infect cells via mouse and rat ACE2[19].

Infection of cells is mediated by the interaction of the viral Spike glycoprotein and ACE2 on cells. Spike, which is exposed on the surface of the virion and composed of S1 and S2 subunits[20], binds to ACE2 via the receptor binding motif within the receptor binding domain of the S1 subunit. Cleavage by host proteases transmembrane serine protease 2 (TMPRSS2) or cathepsins potentiates Spike rearrangement and fusion of the viral and host membranes either in

the endosome or at low pH, at the surface of the cell[21]. The positive-sense viral RNA is released into the cell to initiate translation of the viral genome and subsequent replication.

Replication in animals has been well-characterized for cats: SARS-CoV-2 replicates in the lungs, upper respiratory tract, soft palates, tonsils, distal trachea, tracheobronchial lymph node, small intestines, and kidneys[22,23]. Shedding of virus in respiratory droplets leads to transmission between cats[23]. SARS-CoV-2 is also transmitted between white-tailed deer[17,24]. In experimental infections of deer, viral RNA has been detected in the nose, mouth, rectum, tonsils, lymphs nodes, spleen, liver, kidney, bone marrow, stomach, ileocecal junction, olfactory bulb, heart, brain, CSF, and large and small intestine. Infectious virus was detected in the trachea, bronchi, nasal wash, and bronchoalveolar lavage fluid[24].

Several animals susceptible to SARS-CoV-2 and to intra-species transmission frequently interact with humans. Studies suggesting that the frequency of SARS-CoV-2 spillover from humans to deer was previously underestimated after examining phylogenetic relationships between human- and deer-derived SARS-CoV-2 sequences[25], highlighting the inherent risk of transmission in both directions across the animal-human interface. Mutations in receptor binding proteins often mediate host tropism and can be responsible for increased transmission[26–29]. In fact, several mutations in Spike have facilitated increased adaptation to mink[30], and there is some evidence for adaptation to deer with a mutation in the ORF3a gene[25]. Therefore, we sought to identify mutations that adapt SARS-CoV-2 to animals, as these could lead to the persistence of SARS-CoV-2 in animal populations. We serially passaged SARS-CoV-2 in cells expressing dog, cat, mink, and white-tailed deer ACE2 and searched for mutations congruent with epidemiologically relevant residues. We identified Spike A222V, which arose in cells expressing cat ACE2, and is a recurrent mutation across eight independent SARS-CoV-2 lineages in human-

derived sequences and occurs in animal sequences. Since Spike A222V is also present in animal-derived SARS-CoV-2 sequences, we hypothesized Spike A222V could be both human and animal adaptive. Thus, we sought to determine the impact of Spike A222V on replication of SARS-CoV-2 in a relevant human cell line and primary deer lung cells, and infection via the ACE2 receptor of various relevant animal species.

Materials and Methods

Phylogeny

The phylogenetic tree of SARS-CoV-2 sequences was built by NextStrain[31]. The nCoV dataset from GISAID on the NextStrain app was accessed on June 19, 2024. The interactive display was used to highlight the amino acid identity at Spike 222 in approximately 4000 global sequences with dates from December 2019 to June 2024. 4000 sequences were selectively displayed by NextStrain for maximum build performance and legibility in the app with automatic, representative subsampling. Wuhan-Hu-1/201 is used as a reference for site numbering, and the tree is rooted using early sequences from Wuhan. An assumed rate of 8×10^{-4} nucleotide substitutions per site per year was used.

Variant analysis from SARS-CoV-2 sequences

Complete SARS-CoV-2 genomes from multiple hosts (including human, dog, cat, mink, and deer) were downloaded from the GISAID EpiCov database (www.gisaid.org) as of June 20, 2024. Genomes with low coverage (defined as having more than 5% ambiguous bases, N's) were excluded from the dataset. Additionally, 80 SARS-CoV-2 genomes from deer, made publicly available by McBride et al.[25], were obtained from GenBank for inclusion in the analysis. All sequences were aligned to SARS-CoV2 reference (NCBI Reference Sequence/NC_045512.2)

using Minimap2[32] with default settings except for the use of the parameter '-ax asm5'. Sequences with aligned lengths less than 20,000 base pairs were excluded from the analysis to ensure adequate coverage for variant calling. Variant extraction was performed using the 'mpileup' command in Samtools[33].

Cell lines and plasmids

Baby hamster kidney cells (BHK-21 [C-13]) and human lung epithelial cells (Calu-3 HTB-55) were acquired from ATCC. African green monkey kidney epithelial cells expressing TMPRSS2 and human ACE2 (Vero E6-TMPRSS2-T2A-hACE2; NR-54970) were acquired from BEI Resources. White-tailed deer primary lung cells were provided by the USDA. All cells were grown in carbon dioxide (5%) incubator at 37 °C with a humidified atmosphere. Primary deer lung cells were grown in Medium 199 (Gibco 12-340-030) with 10% Fetal Bovine Serum. All other cells were grown in Dulbecco's modified Eagle medium (DMEM; Corning™ 10013CV) supplemented with gentamicin sulfate (0.1%), non-essential amino acids (1X), HEPES (25 mM), and either 5% (Vero E6-TMPRSS2-T2A-hACE2 and BHK-21) or 20% (Calu-3) fetal bovine serum. Vero E6-TMPRSS2-T2A-hACE2 also required the addition of 0.01 mg/mL puromycin. Expression plasmids encoding the human (pGL113), dog (pGL114), cat (pGL116), mink (pGL271), or white-tailed deer ACE2 (pGL369) sequence were shared by B. Zhou and G. Larson at the CDC.

Passaging experiments and sequencing

BHK-21 cells were transfected with 0.5 ug of either pUC19 or pGL113 (human), pGL114 (dog), pGL116 (cat), pGL271 (mink), or pGL369 (white-tailed deer) expression plasmids for ACE2. Cells were infected in quadruplicate at a MOI of 0.01 24h later with SARS-CoV-2 Wuhan-Hu-1

(GenBank accession NC_045512.2) bearing the Spike D614G mutation generated from an infectious clone detailed below. Supernatants were harvested from each replicate at peak titer and quantified by plaque assay on Vero E6 hACE2-TMPRSS2 cells. Supernatants were passaged independently for a total of ten passages. RNA was extracted on the unpassaged virus and from passage 5 and 10 viruses using the Zymo Research *Quick*-RNA Viral kit (R1035). Library preparation was performed using an Illumina COVIDseq assay kit with ARTIC V3 primers. Paired-end 2x150bp sequencing was performed on the Illumina HiSeq platform. Sequencing files were trimmed with bbduk to remove low quality reads and ARTIC primer sequences.

Generation of live SARS-CoV-2 mutants

Mutant viruses were generated using a full-length infectious cDNA clone of the Wuhan-Hu-1 strain of SARS-CoV-2 (GenBank accession no. NC_045512.2). Mutagenic PCRs were performed to generate overlapping PCR fragments from the clone using Invitrogen Platinum SuperFi II PCR master mix (12368010) or Quantabio repliQa HiFi ToughMix (95200-025). PCR fragments were digested to cut residual PCR template and gel purified using the Machery-Nagel nucleospin gel and PCR clean-up kit (740609.250). Purified fragments were assembled using the OriCiro Genomics 2X RA master mix. Unassembled product was digested, and the assembled plasmid containing the OriC cassette was amplified by replication cycle reaction using the OriCiro Genomics 10X RE mix. Virus was rescued in a BSL-3 laboratory by DNA transfection of BHK-21 cells using a 1:1 mix of amplified replication cycle reaction product and pUC19 with the Polyplus jetOPTIMUS DNA transfection reagent (101000051). At two and three days post-transfection, supernatant was blind passaged on Vero E6-TMPRSS2-T2A-ACE2 cells, and virus was harvested at 25% CPE. Viral genomic sequences were validated by preparing libraries with

IDT Artic V4.1 NCOV-2019 Panel (10011442) and were sequenced using Genewiz NGS Amplicon-EZ services with 2x250bp reads.

Growth curves and infections with transient ACE2 expression

All live virus manipulation was performed at BSL-3. Human lung epithelial cells (Calu-3) and baby hamster kidney cells (BHK-21) were infected at 80% confluency with a MOI of 0.1. White-tailed deer primary lung cells were infected at a MOI of 1. All cells were infected with either SARS-CoV-2 Wuhan-Hu-1 with the Spike D614G mutation (wild-type) or SARS-CoV-2 Spike A222V-D614G diluted in Roswell Park Memorial Institute 1640 (RPMI-1640) medium supplemented with fetal bovine serum (2%) and HEPES (10 mM). Supernatant was harvested every 24 h post-infection and titered by plaque assays on Vero E6-TMPRSS2-T2A-hACE2 cells.

Titration of infectious virus

Infectious virus was quantified by plaque assay by infecting 90% confluent monolayers of Vero E6-TMPRSS2-T2A-hACE2 with serial dilutions of virus made in RPMI-1640 medium supplemented with fetal bovine serum (2%) and HEPES (10 mM). Adsorption was performed for 1h at 37 °C. Following adsorption, a 1:1 mixture of 2X media (2X EMEM, 4X L-glutamine, 0.735% sodium bicarbonate, 0.2 mg/mL gentamicin sulfate, 4% FBS, 20 mM HEPES) and 3% methylcellulose was added to the cells. Cells were fixed after three days of infection using a 10% buffered formalin solution. Cells were stained with a 0.1% crystal violet solution containing 20% ethanol to visualize plaques.

Results

Passaging of SARS-CoV-2 in cells expressing animal ACE2 results in minimal high frequency Spike mutations

SARS-CoV-2 is described as a generalist virus[34] because of its ability to use several species as hosts[35,36]. Despite this generalist nature, several animal-adaptive mutations have been identified[30], raising concerns about how these changes may affect the transmission dynamics, virulence, and immune evasion of SARS-CoV-2. In addition, the evolutionary rates of SARS-CoV-2 are elevated in animals like farmed mink[37], white-tailed deer[25], and cats[38], suggesting animal-adaptive mutations could arise in these animals and lead to the establishment of new animal reservoirs. To determine whether animal-adaptive mutations would arise after repeated exposure to animal ACE2, we serially passaged SARS-CoV-2 in a cell culture model of transient animal-ACE2 expression. Baby hamster kidney cells (BHK-21), a permissive but non-susceptible cell line (**Supplementary Figure 1**), transiently expressing either dog, cat, mink, deer, or human ACE2 were infected with the Wuhan-Hu-1 strain of SARS-CoV-2 bearing the Spike D614G, a likely human-adaptive mutation present in nearly all sequences after late 2020[39]. We performed 10 passages at a MOI of 0.01 and sequenced the input virus and the populations following 5 (p5) and 10 (p10) passages (**Figure 1A**).

We observed consensus-level, non-synonymous mutations within ORF1a, ORF1b, Spike, ORF3a, ORF7a, and Nucleocapsid of the passaged populations in all ACE2 environments (**Figure 1B-F**); however, we only discuss those found in Spike given the well-characterized interaction between Spike and ACE2. All mutations, including those outside of Spike, are detailed in supplementary file 1. Spike R685H was observed at p10 in one replicate (1/2) of cells expressing human ACE2 and in p5 and p10 for one replicate (1/4) of cells expressing cat ACE2 in addition to mink ACE2 (1/4) (**Figure 1B, 1D, and 1E**). Spike R682Q arose in both p5 and

p10 for all replicates of dog ACE2 (4/4), three of cat ACE2 (3/4), three of mink ACE2 (3/4), and all of deer ACE2 (4/4) but not human ACE2 (0/2) (**Figure 1B-F**). Previous studies have shown Spike R682Q and R685H arise during passaging in TMPRSS2-deficient Vero cells[40,41], and, consequently, are likely adaptations towards the BHK-21 cells we used. Spike S943N arose at both p5 and p10 in one replicate (1/4) of cells expressing dog ACE2, three replicates (3/4) of cells expressing cat ACE2, one replicate (1/4) of cells expressing mink ACE2, and two replicates (2/4) of cells expressing deer ACE2 but not in cells expressing human ACE2 (**Figure 1B-F**). Finally, Spike A222V, which lies in the receptor binding domain (RBD), arose in both p5 and p10 in one replicate (1/4) of cells expressing cat ACE2 (**Figure 1D**). This mutation was characteristic of the 20E (EU1) variant and later the Delta subvariant AY.4.2[42]. Given that Spike A222V and Spike S943N arose in both p5 and p10 of cells expressing animal ACE2, we posited these mutations may confer an adaptive advantage for SARS-CoV-2 through infection via animal ACE2.

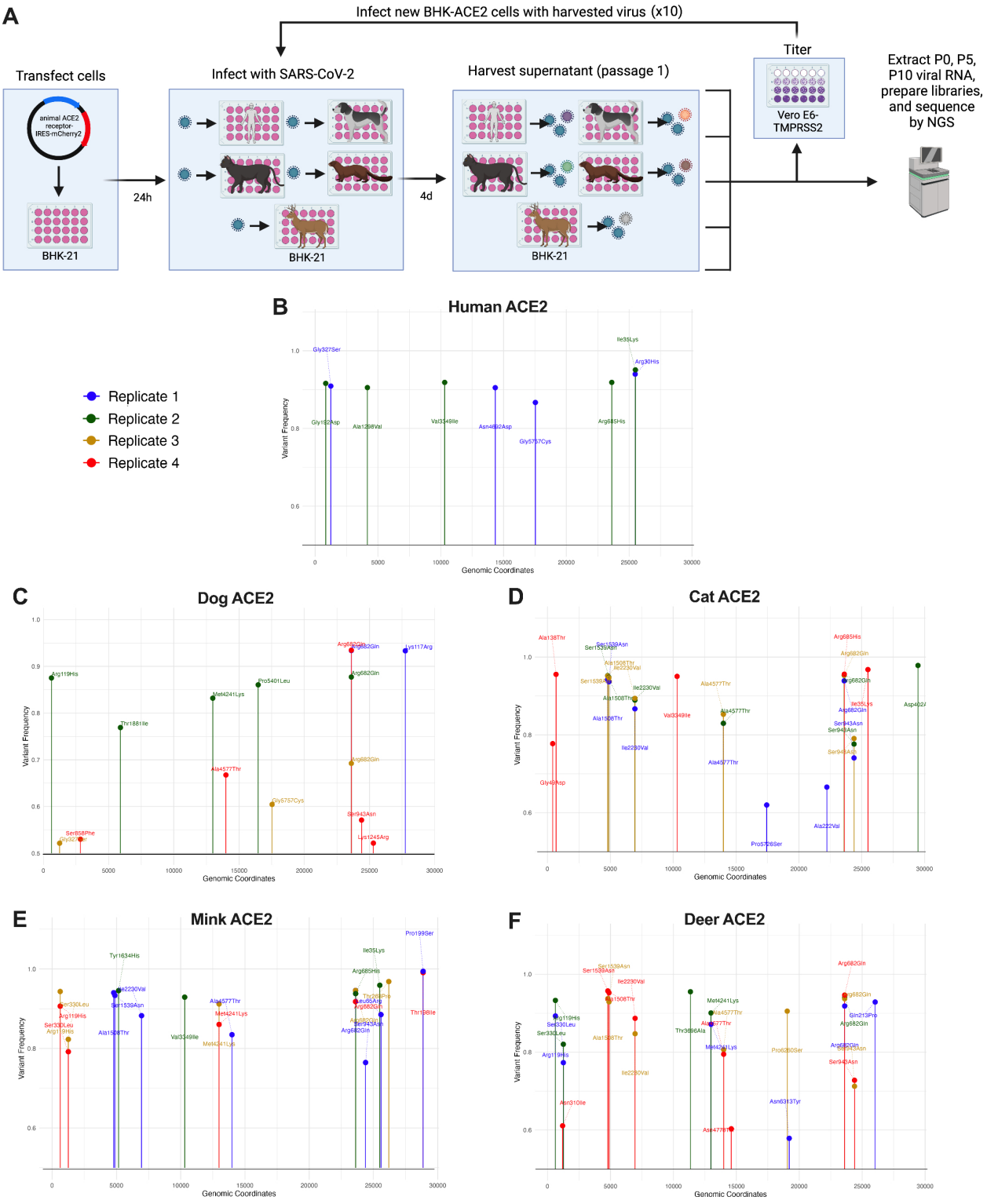


Figure 1. Mutations in SARS-CoV-2 following passaging in cells expressing animal ACE2.

The Wuhan-Hu-1 strain of SARS-CoV-2 with the Spike D614G mutation was passaged ten

times at a MOI of 0.01 in BHK-21 cells expressing either human, dog, cat, mink, or white-tailed deer ACE2 receptor. All passaging was performed in four replicates, except for human ACE2, which had two replicates. Passaged populations were sequenced via next generation sequencing, and mutations are reported across the length of the genome. Variants were called with a p-value threshold of 0.01 after Bonferroni correction, and only those which exceeded 0.5 variant frequency are shown. **A.** Schematic of passaging experiment. **B-F**: SARS-CoV-2 mutations after ten passages in BHK-21 cells expressing human **(B)**, dog **(C)**, cat **(D)**, mink **(E)**, or white-tailed deer **(F)** ACE2.

Prevalence of mutations from passaging experiments in SARS-CoV-2 sequences from humans and animals across phylogeny and time

Animal-adaptive SARS-CoV-2 mutations have been associated with outbreaks and mortality in farmed animals and have spilled back to humans, thus classifying such mutations as critical One Health threats. To identify animal-adaptive mutations in SARS-CoV-2 that constitute important targets in a One Health framework, we searched for Spike mutations A222V and S943N identified from our passaging experiments among other human and animal-derived SARS-CoV-2 sequences. While Spike S943N arose in 7 passaging replicates, this mutation was not detected in any circulating SARS-CoV-2 sequences and thus is likely not relevant to infection of a natural host. In contrast, Spike A222V was a recurring mutation in human-derived SARS-CoV-2 sequences from distinct clades as previously reported[42] (**Figure 2A**). The natural reoccurrence of Spike A222V in human-derived SARS-CoV-2 sequences marked its relevance to public health, and as such, we investigated this mutation further to determine why it may recur. Spike A222V is a mutation in the N-terminal domain of SARS-CoV-2 that shows signals of positive selection and was associated with both the 20EU and Delta variants[42]. Spike A222V is

detected sporadically and to very low frequencies in more recent lineages, some of which include Omicon (21M), JN.1, XBB.2.3, and EG.5. (**Figure 2A**). We detected Spike A222V in 642,857 of 16,192,873 (3.97%) human-derived SARS-CoV-2 sequences from GISAID (**Figure 2B**). We next asked whether dog, cat, mink, or white-tailed deer SARS-CoV-2 sequences also harbor Spike A222V since these animals closely interact with humans and support SARS-CoV-2 transmission[12,25,43,44]. Spike A222V is present in 6.5% (8/123) of dog-derived sequences, 7.69% (13/169) of cat-derived sequences, 13.19% (178/1,349) of mink-derived sequences, and 0.68% (4/588) of white-tailed deer-derived sequences (**Figure 2C-F**). Based on the presence of Spike A222V in these sequences, we hypothesized Spike A222V could be adaptive across a wide range of species, including humans.

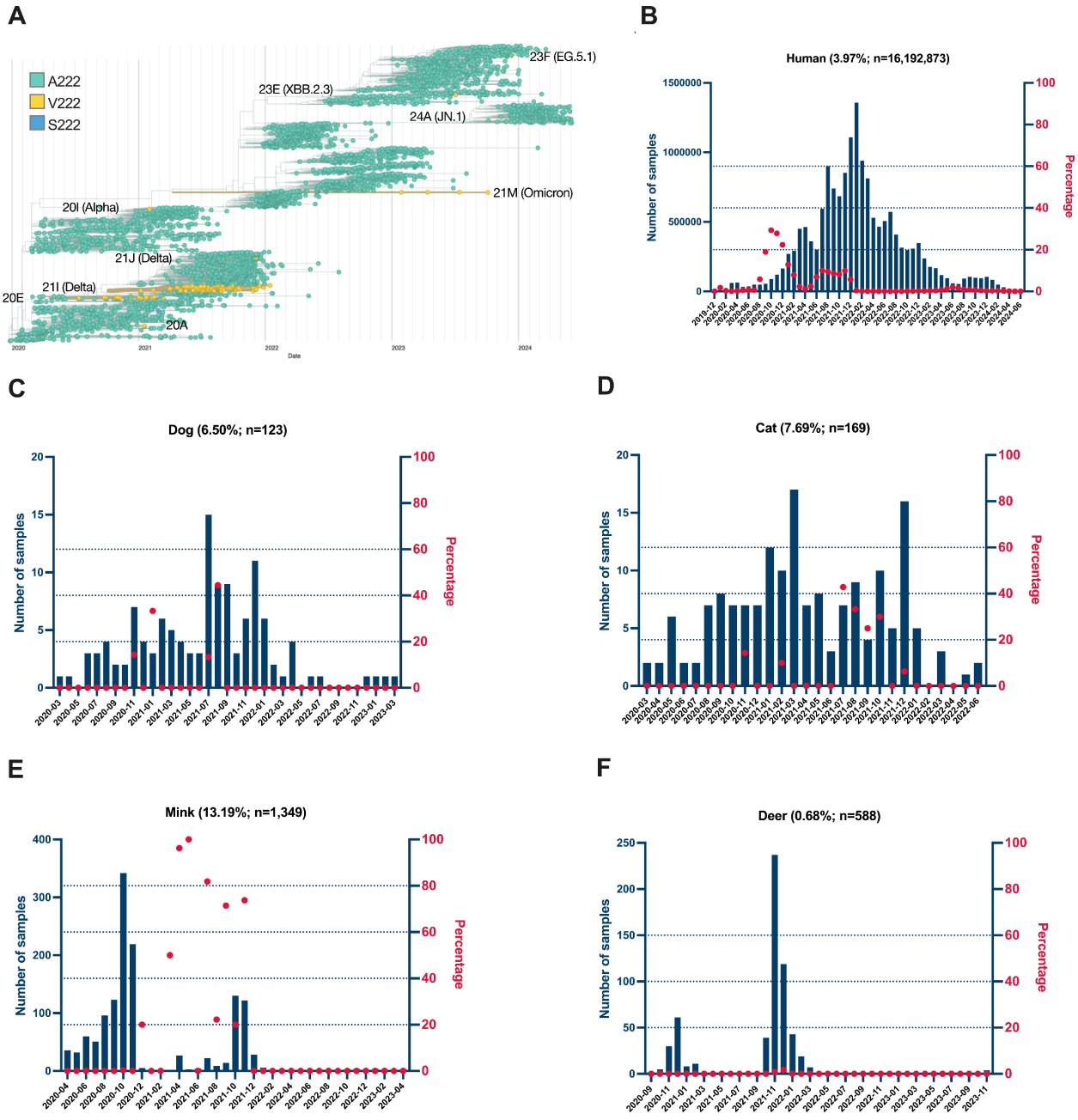


Figure 2. Spike A222V reoccurs in human and animal sequences across time and clade. A. Phylogenetic tree of SARS-CoV-2 sequences from all hosts on GISAID. Phylogeny represents ~4,000 genomes selected by NextStrain. The amino acid identity Spike V222 is highlighted in the tree (yellow). Clades are annotated according to sequences containing Spike V222. **B-F.** Frequency of the Spike A222V mutation in SARS-CoV-2 sequences found in different hosts

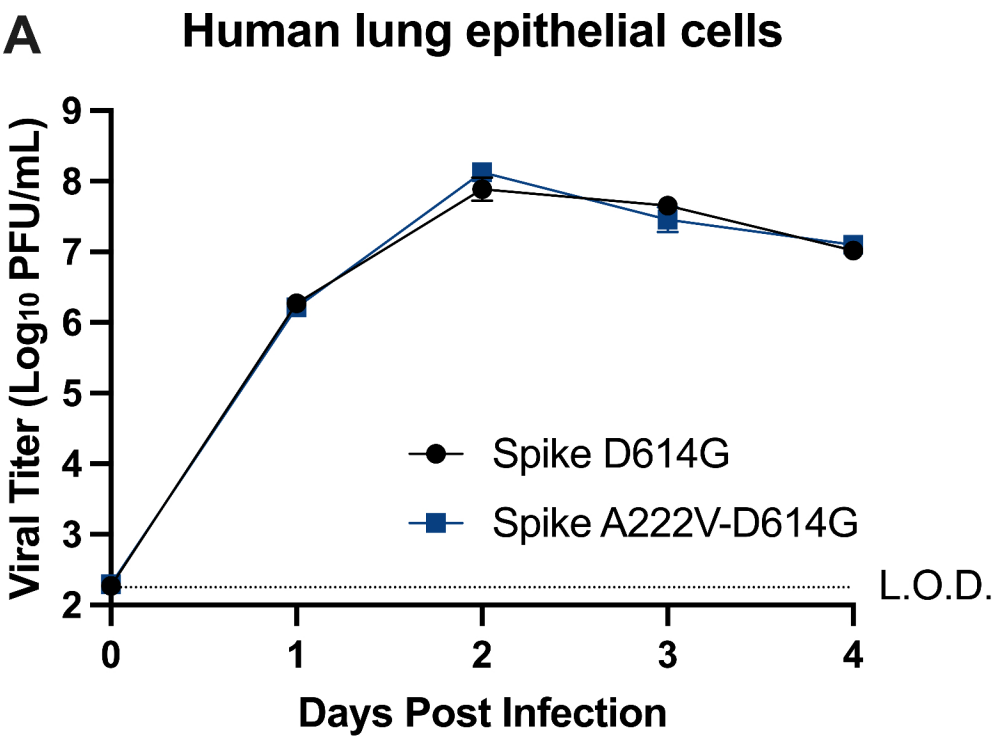
over time. Frequencies are shown for humans **(B)**, dog **(C)**, cat **(D)**, mink **(E)**, and white-tailed deer **(F)**. The blue bars represent the number of samples collected in each month. The red dots indicate the percentage of samples with the A222V mutation in a given month. The percentage within parentheses following the name of each host for each panel (e.g., Human (3.97%)) denotes the overall frequency of the A222V mutation observed in that host.

SARS-CoV-2 Spike A222V replicates to higher levels in primary deer lung epithelial cells but has no effect on replication in human lung epithelial cells

Due to the prevalence of Spike A222V in both human and animal-derived SARS-CoV-2 sequences, we first asked whether Spike A222V confers an advantage for SARS-CoV-2 during replication in human cells. Previous biochemical studies examining the binding affinities of Spike A222V-D614G and Spike D614G to human ACE2 (hACE2) determined that Spike A222V-D614G binds to hACE2 with a higher affinity[42], suggesting this mutation could provide an advantage during entry into human cells. To determine if there are differences in replication kinetics of SARS-CoV-2 bearing Spike A222V in human cells, we constructed a live, Spike A222V virus in the Wuhan-Hu-1 background with Spike D614G. We infected human lung epithelial cells (Calu-3) with SARS-CoV-2 Spike D614G and Spike A222V-D614G at a MOI of 0.1. Notably, SARS-CoV-2 Spike A222V-D614G did not replicate differently to SARS-CoV-2 Spike D614G alone (**Figure 3A**).

While Spike A222V has a neutral effect on the replication of SARS-CoV-2 in human lung cells, we sought to test the hypothesis that Spike A222V is animal-adaptive. While we could not obtain primary cat, dog, or mink cells to infect with SARS-CoV-2 despite the higher frequency of Spike A222V in SARS-CoV-2 sequences from these host species, we instead infected primary deer cells. Deer are a highly relevant species given their ability to transmit

SARS-CoV-2 to other deer in nature with potential spillover to humans[13], and they show evidence of evolution of SARS-CoV-2[25] and constitute an important One Health host for their interaction with many species, including humans. We posited evolution of SARS-CoV-2 in deer and their maintenance of the virus could be likely and thus sought to test whether Spike A222V could increase SARS-CoV-2 replication in primary deer cells if it were to arise *de novo* in deer or spillover from other species. We infected white-tailed deer primary lung cells with SARS-CoV-2 Spike A222V-D614G and Spike D614G at a MOI of 1. SARS-CoV-2 Spike A222V-D614G replicated to significantly higher titers earlier than Spike D614G (**Figure 3B**). Thus, Spike A222V increases fitness of SARS-CoV-2 in primary deer lung epithelial cells.


A
Human lung epithelial cells
Viral Titer (Log10 PFU/mL)
9
8
7
6
5
4
3
2
0
1
2
3
4
Days Post Infection
Spike D614G
Spike A222V-D614G
L.O.D.

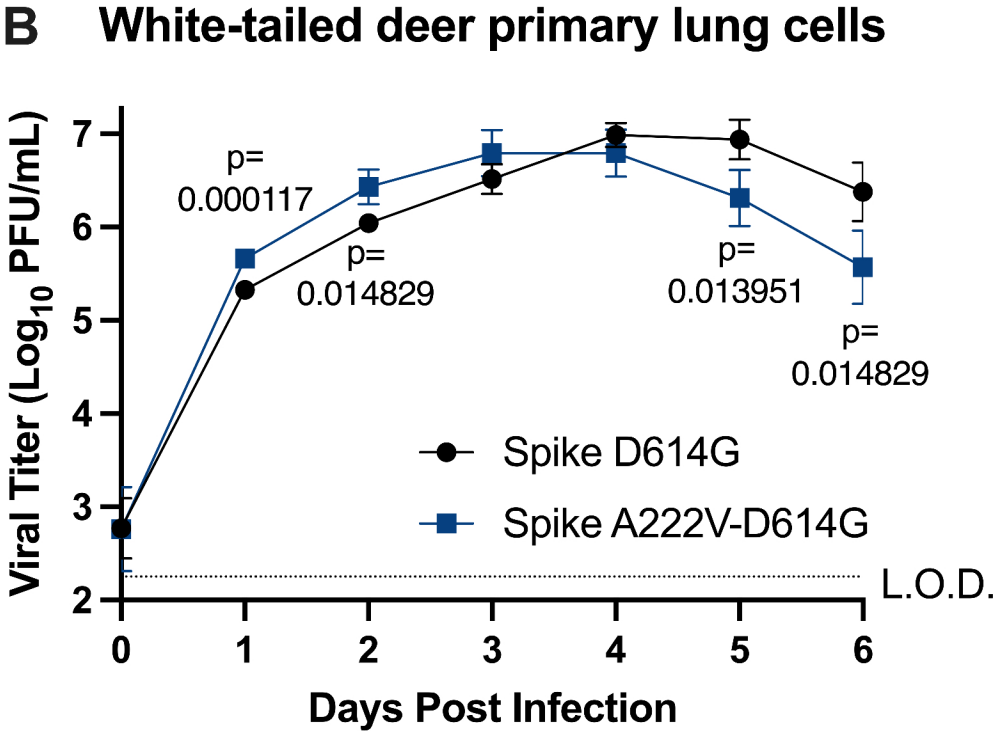

B
White-tailed deer primary lung cells
Viral Titer (Log10 PFU/mL)
7
6
5
4
3
2
0
1
2
3
4
5
6
Days Post Infection
p= 0.000117
p= 0.014829
p= 0.013951
p= 0.014829
Spike D614G
Spike A222V-D614G
L.O.D.

Figure 3. Replication of SARS-CoV-2 Spike A222V in human and deer lung cells. A. Replication of SARS-CoV-2 Spike A222V-D614G in human lung epithelial cells (Calu-3). Cells were infected with a MOI of 0.1, and supernatant was collected each day post infection. Infectious virus was titrated on VeroE6-human ACE2-TMPRSS2 cells. Means represent the average of 6 technical replicates performed in 2 independent biological replicates. **B.** Replication of SARS-CoV-2 Spike A222V-D614G in white-tailed deer primary lung cells. Cells were infected with a MOI of 1, and supernatant was collected each day post infection. Infectious virus was titrated on VeroE6-hACE2-TMPRSS2 cells. Means represent the average of 6 technical replicates. Growth curve data (panels A and B) are representative of two independent biological replicates. Statistical comparisons were made using multiple unpaired t-tests with Welch correction and Holm-Šídák's correction for multiple comparisons where $\alpha=0.05$. Error bars represent standard deviation. L.O.D.=limit of detection.

Spike A222V does not enhance SARS-CoV-2 replication in deer cells via the ACE2 receptor

Since Spike A222V is near the receptor binding domain of Spike, we hypothesized that increased replication in deer cells was mediated by increased entry via deer ACE2. To test this, we inoculated cells transiently expressing different animal ACE2 receptors with either SARS-CoV-2 Spike A222V-D614G or Spike D614G at a MOI of 0.1. Spike A222V-D614G displayed reduced titers compared to Spike D614G in BHK-21 cells expressing deer ACE2 (**Figure 4A**). Replication was also diminished via human, dog, cat, and mink ACE2 (**Figure 4B-E**) at one day post infection. Replication of SARS-CoV-2 Spike A222V-D614G returned to titers similar to Spike D614G by two days post infection in all cells. These data suggest SARS-CoV-2 replication is attenuated by Spike A222V at the step of entry, though only when entry occurs via

ACE2. In combination with the positive impact on replication in primary deer lung cells, these data reveal the complex role Spike A222V plays in entry via ACE2-dependent and independent host factors.

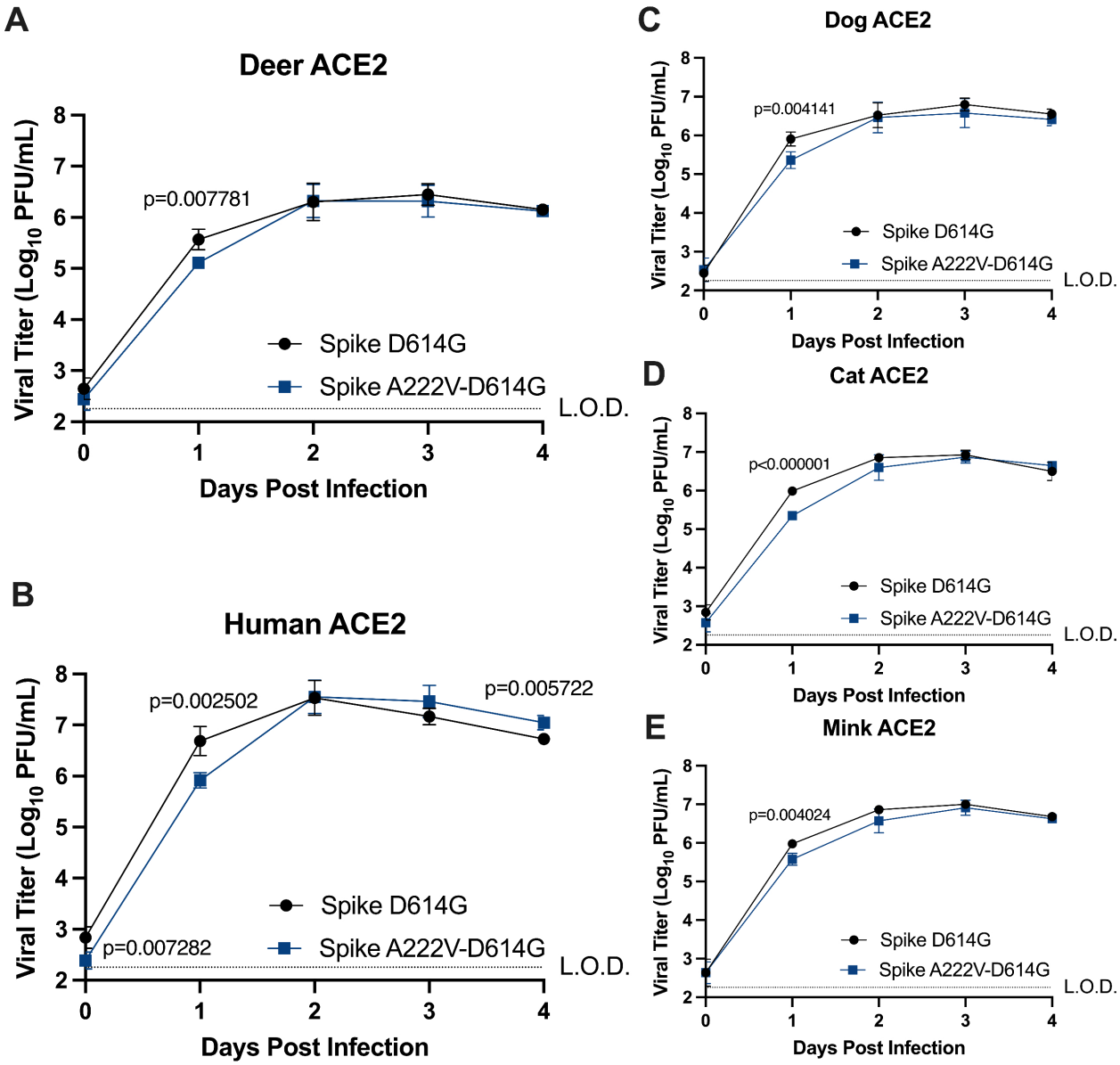


Figure 4. Replicative fitness of Spike mutants with cells expressing deer, human, dog, cat, and mink ACE2. A-E. Replication of SARS-CoV-2 Spike A222V-D614G in BHK-21 cells expressing animal ACE2 receptors. BHK-21 cells were transfected with expression constructs for deer (**A**), human (**B**) dog (**C**), cat (**D**), or mink (**E**) ACE2. Cells were infected 24h later at a

MOI of 0.1, and supernatant was collected each day post infection. Infectious virus was titrated on VeroE6-human ACE2-TMPRSS2 cells. Data comprise two independent biological replicates, and means represent the average of 6 technical replicates. Statistical comparisons were made using multiple unpaired t-tests with Welch correction and Holm-Šídák's correction for multiple comparisons where α=0.05. Bars represent the standard deviation of the mean. L.O.D.= limit of detection.

Discussion

SARS-CoV-2 mutations that confer replication advantages in animals have resulted in large animal outbreaks resulting in animal mortality[5], economic losses to farming industries[45], and spillover events to humans[12]. With an aim to identify key animal-adaptive mutations, we experimentally evolved SARS-CoV-2 in the presence of cells expressing animal ACE2. We identified a mutation, Spike A222V, from the passaging experiment that recurred in human-derived sequences and was also present in dog, cat, mink, and deer-derived SARS-CoV-2 sequences. We assessed the replicative fitness of Spike A222V in the Wuhan-Hu-1 background with Spike D614G and determined it replicates to higher levels in primary deer lung epithelial cells.

The Spike A222V mutation in SARS-CoV-2 recurred in at least eight different lineages of human-derived SARS-CoV-2 sequences and was present in animal sequences. Biochemical and structural simulations have characterized the effect of the Spike A222V mutation on Spike stability and potential human ACE2 receptor binding[42], but our study is the first to examine the effect of Spike A222V on replication of authentic SARS-CoV-2 *in vitro* and to test replication in human and animal cells. Based on an increase in affinity of Spike A222V for human ACE2[42], we hypothesized Spike A222V would increase the replication of SARS-CoV-2 in human cells via

enhancing interaction with the ACE2 receptor. We infected cells transiently expressing human ACE2 that were otherwise not susceptible to infection and surprisingly observed decreased replication of the mutant bearing Spike A222V. Furthermore, we found no difference in the replication of Spike A222V-D614G and Spike D614G viruses in the Wuhan-Hu-1 background in human lung epithelial cells (Calu-3). Since Spike A222V first arose in the 20E (EU1) variant background and not the Wuhan strain that we use here, it is possible that other mutations characteristic of the 20E (EU1) variant (ORF10 V30L, N A220V and ORF14 L67F) work synergistically with Spike A222V. The importance of epistatic interactions with Spike A222V has been demonstrated by others and could explain the recurring nature of Spike A222V in human-derived SARS-CoV-2 sequences where the mutation enhances the fitness of the virus only in certain backgrounds. Spike A222V also negatively impacted replication in cells expressing cat ACE2, though this mutation arose in this same environment during passaging. Since our passaging experiments did not impose a significant bottleneck on the viral populations (1 virus particle per 100 cells), it is most likely that Spike A222V confers an advantage in replication via cat ACE2 only in the context of the other Spike mutations we identified: R682Q and S943N. While we did not construct and examine a mutant bearing all of these Spike mutations, we can speculate that these may have been required for Spike A222V to arise to such a high frequency. While we did not determine that Spike A222V enhances the fitness of SARS-CoV-2 in cells that express human or cat ACE2, this may not be true with other variant backgrounds. Future studies should examine this interplay of mutations *in vitro*.

SARS-CoV-2 is classified as a generalist virus that has seldom evolved by positive selection in animals. Instead, the majority of deterministic processes of evolution have been caused by purifying selection and stochastic processes[25,46]. However, a few instances of adaptation

to animals via positive selection have been documented. The cluster 5 variant of SARS-CoV-2, which caused an outbreak on mink farms in 2020, harbored the Spike Y453F mutation that enhanced interaction of Spike with mink ACE2[16]. Experimental infection of ferrets with SARS-CoV-2 also resulted in mutation to Spike 453F and Spike 501T which were associated with increased viral shedding in ferrets[30]. Spike F485L, which is has arisen multiple times in mink-derived SARS-CoV-2 sequences, adapts Spike to enhanced entry via ferret ACE2[30]. At least 58 sites under positive selection have been detected in SARS-CoV-2 from spillover events from humans to deer[13], and Spike H655Y was rapidly fixed in experimental inoculation of cats[46] though none of these mutations have been validated experimentally. In the present study, we identified a mutation in Spike at amino acid 222 that is a putative deer-adaptive mutation. While less than one percent of deer-derived SARS-CoV-2 sequences harbor this mutation, it is possible that the mutation is underrepresented as no deer sequences were available beyond the beginning of 2022. Live transmission studies in deer could elucidate whether this mutation arises *de novo* and whether it impacts deer-to-deer transmission or pathogenesis.

Spike A222V was represented at the highest frequency in mink-derived sequences (13%), but, surprisingly, showed decreased replication in cells expressing mink ACE2. We saw a similar result with human, dog, cat, and deer ACE2. Spike A222V attenuated SARS-CoV-2 replication via deer ACE2, yet in primary lung cells from white-tailed deer, replicative fitness of SARS-CoV-2 was enhanced by Spike A222V. A similar result was obtained in Calu-3 cells, where the Spike A222V showed equivalent replication to the WT virus but in BHK-21 cells expressing human ACE2 the mutant was attenuated. This suggests that other entry or attachment factors present on primary deer lung cells or Calu-3 cells could be interacting differentially with Spike V222 compared to Spike A222. TMPRSS2 is highly expressed in Calu-3 cells[47] and is known to

increase the infectivity of SARS-CoV-2[48]. While studies have not validated biochemically whether BHK-21 cells express this protease, the lungs of white-tailed deer express TMPRSS2[49]. The presence or absence of TMPRSS2 and differential dependence of Spike on TMPRSS2 could be responsible for altering the replication of the Spike A222V mutant in Calu-3 and deer lung cells versus BHK-21 cells expressing the ACE2 receptor. It would be useful to determine if Spike A222V has a positive impact on replication and which host factors are responsible for this in other susceptible animal cell lines, tissues, or during live transmission studies since we were limited by cell availability and animals for transmission experiments.

Our approach of serial passaging identified Spike A222V as a naturally recurring mutation that enhances replication in deer lung cells. While further examination of the infection, pathogenicity, and transmissibility of SARS-CoV-2 bearing Spike A222V to and between deer should be performed, surveillance efforts in animals should continue to monitor for this mutation and other putatively animal-adaptive mutations which could be under positive selection.

Acknowledgments

We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. We thank Bin Zhou and Gloria Larson from the CDC for ACE2 expression plasmids. The primary deer lung epithelial cells were graciously shared by Dane Jasperson from the USDA-ARS.

Data availability

The sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA1291747.

References

1. CDC. COVID data tracker. *Centers for Disease Control and Prevention* https://covid.cdc.gov/covid-data-tracker/ (2020).
2. Cui, S. *et al.* An Updated Review on SARS-CoV-2 Infection in Animals. *Viruses* **14**, (2022).
3. SARS-CoV-2. *WOAH - World Organisation for Animal Health* https://www.woah.org/en/disease/sars-cov-2/ (2023).
4. Khare, S. *et al.* GISAID's role in pandemic response. *China CDC Wkly.* **3**, 1049–1051 (2021).
5. Oreshkova, N. *et al.* SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. *Euro Surveill.* **25**, (2020).
6. van Volksgezondheid, M. & Sport, W. en. Advies OMT-Z nertsen en SARS-CoV-2. https://www.rijksoverheid.nl/documenten/brieven/2020/06/03/advies-omt-z-nertsen-en-sars-cov-2 (2020).
7. Garigliany, M. *et al.* SARS-CoV-2 natural transmission from human to cat, Belgium, March 2020. *Emerg. Infect. Dis.* **26**, 3069–3071 (2020).
8. McAloose, D. *et al.* From people to Panthera: Natural SARS-CoV-2 infection in tigers and lions at the Bronx Zoo. *bioRxiv* (2020) doi:10.1101/2020.07.22.213959.
9. Grome, H. N. *et al.* SARS-CoV-2 outbreak among Malayan tigers and humans, Tennessee, USA, 2020. *Emerg. Infect. Dis.* **28**, 833–836 (2022).

10. Chan, J. F.-W. *et al.* Simulation of the Clinical and Pathological Manifestations of Coronavirus Disease 2019 (COVID-19) in a Golden Syrian Hamster Model: Implications for Disease Pathogenesis and Transmissibility. *Clin. Infect. Dis.* **71**, 2428–2446 (2020).
11. Kim, D.-H. *et al.* Neurologic Effects of SARS-CoV-2 Transmitted among Dogs. *Emerg. Infect. Dis.* **29**, 2275–2284 (2023).
12. Oude Munnink, B. B. *et al.* Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **371**, 172–177 (2021).
13. Feng, A. *et al.* Transmission of SARS-CoV-2 in free-ranging white-tailed deer in the United States. *Nat. Commun.* **14**, 4078 (2023).
14. Yen, H.-L. *et al.* Transmission of SARS-CoV-2 delta variant (AY.127) from pet hamsters to humans, leading to onward human-to-human transmission: a case study. *Lancet* **399**, 1070–1078 (2022).
15. Zhao, X. *et al.* Broad and differential animal angiotensin-converting enzyme 2 receptor usage by SARS-CoV-2. *J. Virol.* **94**, (2020).
16. Ren, W. *et al.* Mutation Y453F in the spike protein of SARS-CoV-2 enhances interaction with the mink ACE2 receptor for host adaption. *PLoS Pathog.* **17**, e1010053 (2021).
17. Palmer, M. V. *et al.* Susceptibility of white-tailed deer (Odocoileus virginianus) to SARS-CoV-2. *J. Virol.* **95**, (2021).
18. Porter, S. M. *et al.* Experimental SARS-CoV-2 infection of elk and mule deer. *Emerg. Infect. Dis.* **30**, 354–357 (2024).
19. Shuai, H. *et al.* Emerging SARS-CoV-2 variants expand species tropism to murines. *EBioMedicine* **73**, 103643 (2021).

20. Kreutzberger, A. J. B. *et al.* SARS-CoV-2 requires acidic pH to infect cells. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2209514119 (2022).
21. Wang, J. *et al.* Computational insights into the membrane fusion mechanism of SARS-CoV-2 at the cellular level. *Comput. Struct. Biotechnol. J.* **19**, 5019–5028 (2021).
22. Rudd, J. M. *et al.* Clinical and Histopathologic Features of a Feline SARS-CoV-2 Infection Model Are Analogous to Acute COVID-19 in Humans. *Viruses* **13**, (2021).
23. Shi, J. *et al.* Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* **368**, 1016–1020 (2020).
24. Cool, K. *et al.* Infection and transmission of ancestral SARS-CoV-2 and its alpha variant in pregnant white-tailed deer. *Emerg. Microbes Infect.* **11**, 95–112 (2022).
25. McBride, D. S. *et al.* Accelerated evolution of SARS-CoV-2 in free-ranging white-tailed deer. *Nat. Commun.* **14**, 5105 (2023).
26. Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* **24**, 1634–1643 (2005).
27. Hou, Y. J. *et al.* SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020).
28. Kang, L. *et al.* A selective sweep in the Spike gene has driven SARS-CoV-2 human adaptation. *Cell* **184**, 4392–4400.e4 (2021).
29. Cereghino, C. *et al.* Evolution at Spike protein position 519 in SARS-CoV-2 facilitated adaptation to humans. *npj Viruses* **2**, (2024).
30. Zhou, J. *et al.* Mutations that adapt SARS-CoV-2 to mink or ferret do not increase fitness in the human airway. *Cell Rep.* **38**, 110344 (2022).

31. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).

32. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

33. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

34. Tan, C. C. S. *et al.* Transmission of SARS-CoV-2 from humans to animals and potential host adaptation. *Nat. Commun.* **13**, 2988 (2022).

35. Abdel-Moneim, A. S. & Abdelwhab, E. M. Evidence for SARS-CoV-2 Infection of Animal Hosts. *Pathogens* **9**, (2020).

36. Conceicao, C. *et al.* The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. *PLoS Biol.* **18**, e3001016 (2020).

37. Porter, A. F., Purcell, D. F. J., Howden, B. P. & Duchene, S. Evolutionary rate of SARS-CoV-2 increases during zoonotic infection of farmed mink. *Virus Evol* **9**, vead002 (2023).

38. Bashor, L., Gagne, R. B., Bosco-Lauth, A., Stenglein, M. & VandeWoude, S. Rapid evolution of SARS-CoV-2 in domestic cats. *Virus Evol* **8**, veac092 (2022).

39. Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).

40. Ogando, N. S. *et al.* SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *J. Gen. Virol.* **101**, 925–940 (2020).

41. Sasaki, M. *et al.* SARS-CoV-2 variants with mutations at the S1/S2 cleavage site are generated in vitro during propagation in TMPRSS2-deficient cells. *PLoS Pathog.* **17**, e1009233 (2021).

42. Ginex, T. *et al.* The structural role of SARS-CoV-2 genetic background in the emergence and success of spike mutations: The case of the spike A222V mutation. *PLoS Pathog.* **18**, e1010631 (2022).

43. Lyoo, K.-S. *et al.* Experimental Infection and Transmission of SARS-CoV-2 Delta and Omicron Variants among Beagle Dogs. *Emerg. Infect. Dis.* **29**, 782–785 (2023).

44. Bosco-Lauth, A. M. *et al.* Experimental infection of domestic dogs and cats with SARS-CoV-2: Pathogenesis, transmission, and response to reexposure in cats. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 26382–26388 (2020).

45. Wang, L., Didelot, X., Bi, Y. & Gao, G. F. Assessing the extent of community spread caused by mink-derived SARS-CoV-2 variants. *Innovation (Camb.)* **2**, 100128 (2021).

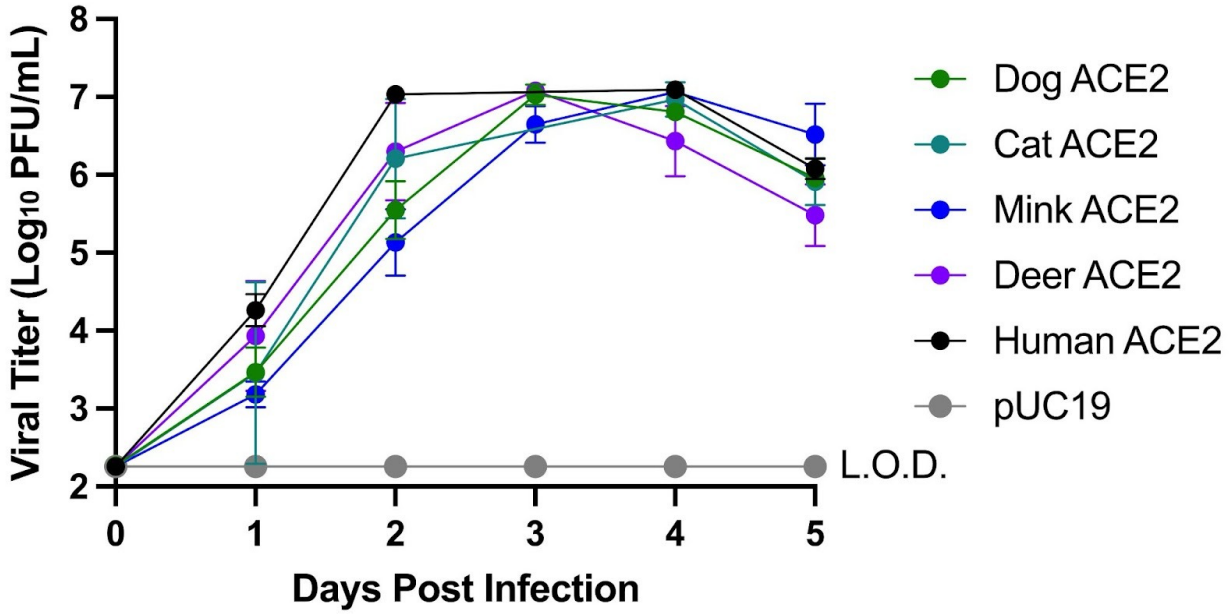
46. Braun, K. M. *et al.* Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck. *PLoS Pathog.* **17**, e1009373 (2021).

47. Koch, J. *et al.* TMPRSS2 expression dictates the entry route used by SARS-CoV-2 to infect host cells. *EMBO J.* **40**, e107821 (2021).

48. Zang, R. *et al.* TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. *Sci. Immunol.* **5**, eabc3582 (2020).

49. Martins, M. *et al.* From Deer-to-Deer: SARS-CoV-2 is efficiently transmitted and presents broad tissue tropism and replication sites in white-tailed deer. *PLoS Pathog.* **18**, e1010197 (2022).

Supplementary Figures



Supplementary Figure 1. Growth curve of SARS-CoV-2 Spike D614G in BHK-21 cells expressing dog, cat, mink, deer, or human ACE2. BHK-21 cells were transfected with expression constructs for human, dog, cat, mink, or deer ACE2 with pUC19 as a control. Cells were infected at a MOI of 0.1, and supernatant was harvested every 24h and titered by plaque assay. Data represent four technical replicates. L.O.D.=limit of detection.

Chapter 5: Conclusion

Emerging RNA viruses like Mayaro virus (MAYV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have caused outbreaks ranging in magnitude from over 70[1] to 778 million cases[2], respectively. SARS-CoV-2 is still a public health issue in 2025, as is MAYV with the most recent cases associated with travel to Suriname in 2024[3]. Progress toward generating more effective vaccines against COVID-19 has been made with recent advancements that generate mucosal immunity at the initial site of replication of SARS-CoV-2, potentially blocking infection and transmission[4,5]. Yet, apart from these vaccines approved for regional use in China and India that have the potential to block transmission, those in wide distribution do not prevent the transmission of SARS-CoV-2[4]. Additionally, since immune escape is always a concern, a more effective vaccine strategy is required to diminish transmission even further. Even if herd immunity were achieved via vaccine-induced and/or natural immunity, without a way to extinguish transmission between animal reservoirs and humans, SARS-CoV-2 would continue to cause human cases of disease after potential evolution in animal reservoirs, resulting in potentially more transmissible or virulent phenotypes. Methods of combating human-to-human transmission will be an important One Health strategy going forward.

As for the future prevalence of MAYV, this virus is likely to cause more frequent, larger outbreaks due to factors such as climate change, increased travel and trade, urbanization, and deforestation that all impact the overlap of the vector and human host. A recent survey indicates that *Aedes aegypti* are increasingly more common in rural communities along Amazon river trading routes, with some rural villages showing higher proportions of housing with *Aedes aegypti* than those of nearby cities[6]. This finding suggests that emerging arboviruses such as MAYV may bypass intermediate transmission cycles and instead emerge directly into an urban cycle involving *Aedes aegypti* localized around rural Amazonian communities. As MAYV is

likely vertically transmitted in *Aedes aegypti*[7], the spread of infected eggs along trading routes into more urban areas could seed larger outbreaks. Should adaptive evolution of MAYV occur either toward the human or urban vector, the consequences could be serious and match outbreaks at scales comparable to the closely related CHIKV. Celone et al. has estimated 58.9 million people in Central and South America are at risk of acquiring MAYV infection based on models of the potential distribution of the virus[8]. Strategies that mitigate MAYV transmission at the interface between sylvatic and urban cycles will become increasingly more important as the ecology of the virus changes since there is an increased trend in MAYV cases in Amazonian regions[9].

The experiments comprising this dissertation have defined single mutations in MAYV and SARS-CoV-2 genomes that enhance infection or virus spread within hosts, increasing their fitness and potential for transmission. For MAYV, a mutation in the structural gene E2, T179N, was identified from passaging the virus in *Aedes aegypti* cells and enhances transmission efficiency in the urban vector *Aedes aegypti* yet must overcome an initial infection barrier. For SARS-CoV-2, the acquisition of the Spike H519 mutation in the history of SARS-CoV-2 evolution increased virus entry and replication in human lung cells through increased binding to the cellular receptor ACE2. In addition, a recurring mutation in sequences of SARS-CoV-2 from humans that also occurs in animal sequences, Spike A222V, does not increase replication of the virus in human lung cells but rather in primary deer lung cells. The mechanism by which this occurs is not through ACE2 but likely other attachment factors on deer cells. The conclusion from these findings is that one non-synonymous mutation in a viral protein can have a significant effect on the replication of viruses in host cells, though sometimes at the cost of fitness in another host. This finding is consistent with numerous other studies that support the so-called

"trade-off hypothesis". For example, passaging Venezuelan equine encephalitis virus (VEEV) in mosquitoes alone generates a virus with higher fitness only in the mosquito host. Similarly, passaging VEEV in rodents generates a virus that is fitter in rodents alone. The mutations acquired by VEEV during these passaging experiments were not evaluated systematically but are presumed responsible for these increased fitness effects of the passaged viruses in a vertebrate or mosquito host alone[10]. However, the trade-off hypothesis is not always supported, as is the case with the E1-A226V mutation in CHIKV. E1-A226V increases the infectivity of CHIKV to *Aedes albopictus* and transmission by *Aedes albopictus* to suckling mice. This increase in fitness in the mosquito did not change the fitness in the mouse since an equal mixture of WT CHIKV and CHIKV E1-A226V inoculated into suckling mice did not result in a change in the ratio of CHIKV E1-226V to WT CHIKV in the blood[11]. Therefore, the trade-off hypothesis is not always supported and should be tested carefully before assumptions are made about the fitness of a mutant virus in another host, which can have implications on transmission cycles. We did not evaluate the effect of Spike H519N on SARS-CoV-2 replication in bat cells due to limitations in acquiring these specimens. However, given the bottlenecks involved in transmission from host to host[12], it is likely the Spike H519N mutation would have had at least a neutral impact on the fitness of the SARS-CoV-2 progenitor in bats or an intermediate host in order to be present in the viral population encountering a susceptible human host. Future experiments should test this hypothesis to determine whether human-adaptive mutations in SARS-CoV-2 are broadly beneficial across host species.

Single mutations can have dramatic impacts on the fitness of emerging RNA viruses, but the likelihood of beneficial mutations arising in the population of viruses transmitted to the next host is rare. Most mutations generated by RNA viruses are deleterious[13,14]. In addition to this,

population bottlenecks during spread in the host can reduce viral fitness significantly. Arboviruses encounter a severe bottleneck upon infection of the mosquito midgut. As few as 28 midgut epithelial cells are infected[15], and for some virus-vector pairings, an estimate of 5 virions is known to found the population of viruses entering the midgut epithelium[16]. Considering the non-natural mutation identified from our MAYV passaging experiments, the likelihood of E2-T179N arising during a transmission cycle is minimal without the acquisition of additional mutations with positive epistatic effects. E2-T179N did not arise in mosquito saliva after 1 passage of MAYV in *Aedes aegypti*, showing either its weaker fitness effect *in vivo* or, simultaneously, the strength of mosquito bottlenecks. Replication of viruses in mosquitoes results in faster diversification[17]; however, the adaptation rate is slower compared to in mammalian hosts[10]. Given this, MAYV E2-T179N may not be likely to fix in a strain of MAYV circulating in an urban transmission cycle because of mosquito bottlenecks, its attenuation in mice, and the strength of purifying selection in vertebrates[18]. We did not sequence the virus in the saliva of mosquitoes that were fed a bloodmeal containing MAYV E2-T179N. Sequencing the saliva of mosquitoes infected with MAYV E2-T179N would help determine its robustness in replicating through different mosquito tissues and whether acquisition of additional mutations are associated with the increased transmission efficiency. Natural transmission studies from *Aedes aegypti* to mice and sequencing the virus in mice could also determine whether this mutation is tolerated in the vertebrate.

For SARS-CoV-2, the mutations we analyzed are present in current or prior circulating viruses. Spike H519 is present in all sequenced SARS-CoV-2 derived from human hosts, and this pattern tends to indicate an importance for maintaining the fitness of a virus in the host. For flaviviruses, conserved residues in the structural protein E are critical for maintaining membrane

fusion function[19]. Rather than a conservation at the genus level, the Spike H519 residue is conserved only at the species level. SARS-CoV, a coronavirus similar to SARS-CoV-2 that also emerged in humans, does not contain a histidine at site 519 in Spike. A more expansive examination of other coronaviruses that infect humans, such as 229E, HKU1, NL63, OC43, and MERS-CoV, reveals that no other human coronavirus species contain a histidine at site 519 of Spike. This suggests H519 has a unique importance for SARS-CoV-2 that could be exploited through targeted antivirals. The other SARS-CoV-2 mutation we examined, Spike A222V, is present in a fewer percentage of overall SARS-CoV-2 sequences but exhibits patterns of homoplasy, which is when a mutation arises in multiple independent lineages of a species. As such, Spike A222V is likely to be associated with future variants. However, given that Spike A222V does not confer neutralization escape, it is not an important residue to consider in the design of vaccines[20].

Viral genetic determinants of host adaptation can be used for the careful design of epitope-based therapeutics or small molecule inhibitors. Mapping K417N, G446S, E484A and Q493R mutations in the Spike Omicron variant of SARS-CoV-2 to neutralizing sensitivity has led to the development of updated vaccines that help prevent severe disease[21]. The Spike H519 mutation examined in this dissertation is essential for optimal viral replication in human lung cells and is under purifying selection in human-derived SARS-CoV-2 sequences. Both of these considerations provide support for Spike H519 as a target for therapeutics. Our structural studies indicate that 519 of Spike, while in the RBD, is not surface exposed. It is located in a region where Spike trimers interface. Mapping of epitopes for SARS-CoV-2-directed antibodies shows that Spike 519 is not a target recognized by neutralizing antibodies[22]. T cell receptor epitope mapping from Lu et al., 2021 show that Spike 519 is also not a target of T cells, supporting the

idea that this site is inaccessible for epitope-based therapeutics[23]. These data suggest that small molecules, rather than epitope-based therapeutics, could target the interaction of H519 with neighboring residues in Spike to disrupt the structure and reduce infectivity.

Pinpointing the viral genetic determinants of adaptation to hosts can also assist with the development of viral countermeasures via public health interventions. In the context of the MAYV E2-T179N mutation, an identification of this mutation in field isolates could be indicative of an evolutionary trajectory of MAYV towards urban vector adaptation. A practical implementation toward identifying this mutation requires sequencing expertise, however, which can be lacking in the regions where MAYV cases are typically identified. Independent researchers or robust public health programs and laboratories are necessary to quickly identify and inform public health officials who can warn citizens early about the dangers of potential urban vectors of MAYV and its clinical signs. Sampling of the mutation is most likely to occur first in the vector. As previously mentioned, given that this mutation results in decreased fitness in the vertebrate mouse model, sampling in humans or animals is unlikely to identify this mutation. Yet, the genetic background of alphaviruses is known to have epistatic effects on mutations, which may increase the need for surveying this mutation in vertebrates in a different genetic context. For example, the positive effect of the E1-A226V mutation on CHIKV transmission by *Aedes albopictus* is negated by the presence of another E1 mutation, 98T, found in Asian strains[24]. Extended to MAYV, it is possible that future mutations in E2 or other genes may reverse the negative effect on fitness in vertebrates, allowing this mutation to be better tolerated in a vertebrate host. It is important to consider that since E2-T179N has an initial negative effect on the infection of *Aedes aegypti*, the virus bearing this mutation would likely

need to acquire further mutations to compensate for this negative contribution toward its fitness in the mosquito.

In summary, this dissertation has defined a genetic determinant of MAYV adaptation to the urban vector *Aedes aegypti* and of SARS-CoV-2 adaptation to humans and the deer reservoir. We show that MAYV has potential for increased transmission by *Aedes aegypti*, an important vector to target not only for MAYV but numerous other emerging arboviruses. In addition, we have uncovered a potential genetic pathway of emergence of SARS-CoV-2 into humans through evolution from a naturally occurring sarbecovirus infecting either bats or pangolins. Finally, we have shown that a mutation in SARS-CoV-2 occurring in humans has an unexpected enhancement of virus replication in deer lung cells, indicating that evolution of the virus in humans may enhance spillover to animals. Converging at the interface between the human host and the vector or reservoir, our findings highlight the adaptability of viruses as they cross species barriers.

References

1. Auguste, A. J. *et al.* Evolutionary and ecological characterization of Mayaro virus strains isolated during an outbreak, Venezuela, 2010. *Emerg. Infect. Dis.* **21**, 1742–1750 (2015).
2. COVID-19 cases. *datadot* https://data.who.int/dashboards/covid19/cases.
3. Orth, H. M. *et al.* A cluster of Mayaro virus infections in a film team returning from Suriname, February 2024. *Euro Surveill.* **29**, (2024).
4. Zhang, R. *et al.* A phase 1, randomized, double-blinded, placebo-controlled and dose-escalation study to evaluate the safety and immunogenicity of the intranasal DelNS1-nCoV-RBD LAIV for COVID-19 in healthy adults. *Vaccines (Basel)* **11**, 723 (2023).

5. Singh, C. *et al.* Phase III Pivotal comparative clinical trial of intranasal (iNCOVACC) and intramuscular COVID 19 vaccine (Covaxin®). *NPJ Vaccines* **8**, 125 (2023).
6. Fikrig, K. *et al.* Changing dynamics of Aedes aegypti invasion and vector-borne disease risk for rural communities in the Peruvian Amazon. *PLoS Negl. Trop. Dis.* **19**, e0012506 (2025).
7. Maia, L. M. S. *et al.* Natural vertical infection by dengue virus serotype 4, Zika virus and Mayaro virus in Aedes (Stegomyia) aegypti and Aedes (Stegomyia) albopictus: Vertical transmission of arboviruses. *Med. Vet. Entomol.* **33**, 437–442 (2019).
8. Celone, M. *et al.* Understanding transmission risk and predicting environmental suitability for Mayaro Virus in Central and South America. *PLoS Negl. Trop. Dis.* **18**, e0011859 (2024).
9. Martins-Filho, P. R., Carvalho, T. A. & Dos Santos, C. A. Mayaro fever in Brazil from 2014 to 2024. *J. Travel Med.* **31**, taae105 (2024).
10. Coffey, L. L. *et al.* Arbovirus evolution in vivo is constrained by host alternation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 6970–6975 (2008).
11. Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E. & Higgs, S. A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog.* **3**, e201 (2007).
12. Wang, D. *et al.* Population bottlenecks and intra-host evolution during human-to-human transmission of SARS-CoV-2. *Front. Med. (Lausanne)* **8**, 585358 (2021).
13. Duffy, S. Why are RNA virus mutation rates so damn high? *PLoS Biol.* **16**, e3000003 (2018).
14. Sanjuán, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8396–8401 (2004).
15. Smith, D. R., Adams, A. P., Kenney, J. L., Wang, E. & Weaver, S. C. Venezuelan equine encephalitis virus in the mosquito vector Aedes taeniorhynchus: infection initiated by a small

number of susceptible epithelial cells and a population bottleneck. *Virology* **372**, 176–186 (2008).

16. Scherer, W. F., Cupp, E. W., Lok, J. B., Brenner, R. J. & Ordonez, J. V. Intestinal threshold of an enzootic strain of Venezuelan encephalitis virus in Culex (Melanoconion) taeniopus mosquitoes and its implications to vector competency and vertebrate amplifying hosts. *Am. J. Trop. Med. Hyg.* **30**, 862–869 (1981).
17. Jerzak, G. V. S., Bernard, K., Kramer, L. D., Shi, P.-Y. & Ebel, G. D. The West Nile virus mutant spectrum is host-dependant and a determinant of mortality in mice. *Virology* **360**, 469–476 (2007).
18. Deardorff, E. R. *et al.* West Nile virus experimental evolution in vivo and the trade-off hypothesis. *PLoS Pathog.* **7**, e1002335 (2011).
19. Allison, S. L., Schalich, J., Stiasny, K., Mandl, C. W. & Heinz, F. X. Mutational evidence for an internal fusion peptide in flavivirus envelope protein E. *J. Virol.* **75**, 4268–4275 (2001).
20. Ginex, T. *et al.* The structural role of SARS-CoV-2 genetic background in the emergence and success of spike mutations: The case of the spike A222V mutation. *PLoS Pathog.* **18**, e1010631 (2022).
21. Cao, Y. *et al.* Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**, 657–663 (2022).
22. Piccoli, L. *et al.* Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* **183**, 1024–1042.e21 (2020).
23. Lu, X. *et al.* Identification of conserved SARS-CoV-2 spike epitopes that expand public cTfh clonotypes in mild COVID-19 patients. *J. Exp. Med.* **218**, (2021).

24. Tsetsarkin, K. A. *et al.* Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7872–7877 (2011).