

RESEARCH ARTICLE

Evolution and spread of Venezuelan equine encephalitis complex alphavirus in the Americas

Naomi L. Forrester¹*, Joel O. Wertheim², Vivian G. Dugan³*, Albert J. Auguste¹, David Lin⁴, A. Paige Adams¹*, Rubing Chen¹, Rodion Gorchakov¹*, Grace Leal¹, Jose G. Estrada-Franco¹*, Jyotsna Pandya¹, Rebecca A. Halpin³, Kumar Hari⁴, Ravi Jain⁴, Timothy B. Stockwell⁵*, Suman R. Das⁵*, David E. Wentworth³*, Martin D. Smith⁶*, Sergei L. Kosakovsky Pond²*, Scott C. Weaver¹



1 Institute for Human Infections and Immunity, Department of Pathology and Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, Texas, United States of America, **2** Department of Medicine, University of California, San Diego, La Jolla, California, United States of America, **3** Virology Group J. Craig Venter Institute, Rockville, Maryland, United States of America, **4** cBio, Fremont, California, United States of America, **5** Informatics Group, J. Craig Venter Institute, Rockville, Maryland, United States of America, **6** Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, California, United States of America

OPEN ACCESS

Citation: Forrester NL, Wertheim JO, Dugan VG, Auguste AJ, Lin D, Adams AP, et al. (2017) Evolution and spread of Venezuelan equine encephalitis complex alphavirus in the Americas. *PLoS Negl Trop Dis* 11(8): e0005693. <https://doi.org/10.1371/journal.pntd.0005693>

Editor: Adalgisa Caccone, Yale University, UNITED STATES

Received: November 15, 2016

Accepted: June 8, 2017

Published: August 3, 2017

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All data are available through the NCBI database with the following accession numbers: KC344505, KC344483, KC344485, KC344430, KC344516, KR260376*, KU059753*, KU059754*, KU059755*, KU059756*, KU059757*, KC344528, KC344484, KC344525, KC344526, KC344486, KC344487, KC344488, KC344506, KC344523, KC344490, KC344502, KC344503, KC344504, KC344507, KC344508, KC344509, KC344510, KC344511, KC344512, KC344513, KC344514, KC344517, KC344518, KC344519, KC344520, KC344521,

* These authors contributed equally to this work.

*a Current address: Virology Surveillance and Diagnosis Branch Influenza Division, Centers for Disease Control, Atlanta, Georgia, United States of America

*b Current address: International Animal Health and Food Safety Institute, Kansas State University Olathe, Olathe, Kansas, United States of America

*c Current address: Department of Pediatrics, Section of Tropical Medicine, Baylor College of Medicine, Houston, Texas, United States of America

*d Current address: The College of the Southern Border (EcoSur), San Cristobal de las Casas, Chiapas, Mexico

*e Current address: Unidad San Cristóbal de Las Casas, Departamento de Salud, Periférico Sur s/n, María Auxiliadora, San Cristóbal de Las Casas, Mexico

*f Current address: Centers for Disease Control and Prevention CDC/National Center for Emerging Zoonotic Infectious Diseases (NCEZID)/Office of Infectious Diseases (OID), Edinburg, Texas, United States of America

*g Current address: Pacific Biosciences, Menlo Park, California, United States of America

*h Current address: Department of Biology, Temple University, Philadelphia, Pennsylvania, United States of America

*i Current address: National Biodefense Analysis and Countermeasures Center, Fort Detrick, Maryland, United States of America

*j Current address: Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America

* naforres@utmb.edu

Abstract

Venezuelan equine encephalitis (VEE) complex alphaviruses are important re-emerging arboviruses that cause life-threatening disease in equids during epizootics as well as spill-over human infections. We conducted a comprehensive analysis of VEE complex alphaviruses by sequencing the genomes of 94 strains and performing phylogenetic analyses of 130 isolates using complete open reading frames for the nonstructural and structural polyproteins. Our analyses confirmed purifying selection as a major mechanism influencing the evolution of these viruses as well as a confounding factor in molecular clock dating of ancestors. Times to most recent common ancestors (tMRCAs) could be robustly estimated only

KC344522, KC344524, KC344471, KC344472, KC344475, KC344476, KC344473, KC344474, KC344459, KC344461, KC344527, KC344468, KC344492, KC344515, KC344431, KC344480, KC344466, KC344467, KC344437, KC344440, KC344444, KC344446, KC344448, KC344455, KC344457, KC344458, KC344443, KC344432, KC344433, KC344434, KC344435, KC344436, KC344438, KC344439, KC344441, KC344442, KC344445, KC344447, KC344454, KC344456, KC344449, KC344451, KC344464, KC344482, KC344465, KC344450, KC344492, KC344478, KC344479, KC344481, KC344463, and KR260737*.

Funding: This work was supported in part with federal funds from the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Department of Health and Human Services (DHHS) under contract number HHSN272200900007C, and by the Defense Threat Reduction Agency (DTRA) under Interagency Agreement with Lawrence Livermore National Laboratory through contract "DTRA100271A-2359 – BASIC". Some of the genomes referenced in this publication were sequenced through funding provided by DHS S&T through Contract No. HSHQDC-13-C-B0009 "Capturing Global Biodiversity of Pathogens by Whole Genome Sequencing. NLF was supported by a grant from the NIAID through the Western Regional Center of Excellence (WRCE) for Biodefense and Emerging Infectious Disease Research, NIH grant U54 AI057156, and by a grant from the NIAID NIH R01-AI095753-01A1. AJA is supported by the James W. McLaughlin endowment fund. JOW, MDS, and SLKP were supported by a University of California Laboratory Fees Research Program Grant (12-LR-236617), and JOW was also supported by an NIH-NIAID Career Development Award (K01AI110181) and an NIH-NIAID R21 (AI115701). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: David Lin, Kumar Hari and Ravi Jain work for cBio.

for the more recently diverged subtypes; the tMRCA of the ID/IAB/IC/II and IE clades of VEE virus (VEEV) were estimated at ca. 149–973 years ago. Evolution of the IE subtype has been characterized by a significant evolutionary shift from the rest of the VEEV complex, with an increase in structural protein substitutions that are unique to this group, possibly reflecting adaptation to its unique enzootic mosquito vector *Culex (Melanoconion) taeniopus*. Our inferred tree topologies suggest that VEEV is maintained primarily *in situ*, with only occasional spread to neighboring countries, probably reflecting the limited mobility of rodent hosts and mosquito vectors.

Author summary

The Venezuelan equine encephalitis (VEE) complex comprises a broadly distributed group of alphaviruses in the Americas that have the potential to emerge and cause severe disease. Historically, VEE complex viruses have caused recurring outbreaks of human and equine encephalitis in Central and South America as well as Mexico, with at least one outbreak resulting in movement of the virus to the southern United States. We present the most comprehensive phylogenetic analysis of complete genomic sequences of the most prominent member of the VEE complex, VEE virus (VEEV). We were able to identify the major forces influencing VEEV evolution, and using the inferred phylogenies we determined that VEEV evolves in geographically segregated lineages with enzootic transmission between rodents and mosquitoes apparently limiting its spread.

Introduction

The Venezuelan equine encephalitis (VEE) antigenic complex comprises a group of alphaviruses that share similar genetic characteristics and can be defined by broad cross-reactivity antigenically [1] which defines them as a group within the *Alphaviridae*. The VEE complex alphaviruses are classified into six subtypes, designated I to VI, and consist of 9 species [2], of which subtype I contains the veterinary and medically important VEE virus (VEEV) (Table 1). Historically, the VEE complex subtypes I–VI were defined by serological analysis, and this nomenclature has persisted. However, the advent of sequencing resulted in several viruses now being reclassified as distinct species. For example, subtype IF is genetically distinct from the remainder of the subtype I viruses and is now classified as the species *Mosso das Pedras virus*. Most VEE complex viruses have enzootic cycles where they circulate between wild animals, generally rodents and mosquitoes, particularly *Culex (Melanoconion) spp.* mosquito vectors. With the exception of VEEV subtype II, VEE complex viruses are geographically distributed throughout Central and South America. VEEV subtype II is found only in Florida and is usually transmitted by *Culex cedei* mosquitoes. The recent appearance for the first time of *Culex (Melanoconion) species* in southern Florida [3] underscores the continuing threat of emergence, possibly enhanced by climate change, which also increases the potential for other VEEV subtypes to spread northwards and establish enzootic transmission cycles. Although many VEE complex viruses have not been implicated in human disease, those that are associated with human disease (VEEV) can cause acute, often severe febrile illness that may progress to encephalitis, causing severe human morbidity and mortality [4]. Patients who survive encephalitis are often left with permanent neurologic sequelae, and the cost for treatment and long-term care related to a single case can be several million dollars [5]. In addition to VEEV

Table 1. The subtypes of VEE complex alphaviruses and their transmission cycles.

Virus	Subtype	Abbreviation	First Isolation	Geographic Range	Vertebrate Host Range	Mosquito vector	Human Disease	Endemic/ Epidemic
Venezuelan equine encephalitis	IAB	VEEV	Venezuela 1938	Trinidad, Peru, Colombia, Guatemala-Mexico-Texas	Horses/humans	<i>Aedes</i> and <i>Psorophora</i> spp	Yes	Epidemic
	IC		Colombia, 1962	Colombia, Venezuela, Peru	Horses/humans	<i>Aedes</i> and <i>Psorophora</i> spp		Epidemic
	ID		Colombia, 1961	South and Central America	Rodents	<i>Culex (Melanoconion)</i> spp.		Endemic
	IE		Panama, 1961	Central America	Rodents, horses, humans	<i>Culex (Melanoconion) taeniopus</i> and <i>Psorophora</i> and <i>Aedes</i> spp.		Epidemic/ Endemic
Everglades	II	EVEV	Florida, USA, 1963	Florida	Birds	<i>Culex (Melanoconion spp.)</i>	No	Endemic
Mucambo	IIIA	MUCV	Trinidad, 1954	South America	Unknown	<i>Culex portesi</i>	yes	Endemic
Tonate	IIIB	TONV	French Guiana, 1973	South and Central America	Unknown	Unknown	No	Unknown
71D1252	IIIC			South America	Unknown	Unknown	No	Unknown
Pixuna	IV	PIXV	Brazil, 1964	South America	Unknown	Unknown	No	Unknown
Cabassou	V	CABV	French Guiana, 1968	French Guiana	Unknown	<i>Culex portesi</i>	No	Unknown
Rio Negro (AG80_663)	VI	RNV	Argentina, 1980	Argentina	Unknown	Unknown	No	Unknown
Mosso das Pedras (78V3531)	IF	MDPV	Brazil, 1978	Brazil	Unknown	<i>Culex (Melanoconion) sp.</i>	No	Unknown

<https://doi.org/10.1371/journal.pntd.0005693.t001>

(subtype I), which causes the majority of the encephalitis cases within the VEE subtype, subtype II Everglades virus (EVEV), which is found only in Florida, can cause neurologic disease in humans [6] and equids [7]. Subtype IIIA, Mucambo virus, also causes febrile disease in humans [8, 9].

VEEV is associated with human disease, and is further subdivided into subtypes IAB, IC, ID, IE. Subtypes ID and IE comprise enzootic/endemic strains that circulate continuously in forests and swamps of northern South America, Central America and Mexico and cause a large burden of endemic disease from direct spillover [10]. The remaining VEEV subtypes, IAB and IC, comprise epizootic/epidemic strains that are associated with periodic equine-amplified outbreaks that result in severe disease in equids and extensive spillover to humans [11]. These outbreaks can spread from South America as far north as the southern United States [12, 13], resulting in up to hundreds-of-thousands of cases over a period of months to a few years. Prior to the 1980s, VEE epizootics involving high case-fatality rates were frequently recorded. Because horses have been an important component of the local agricultural economies within many Latin American regions, VEE has often had a sizeable economic impact as well as a direct effect on public and veterinary health [14]. Recent outbreaks during the 1990s in Venezuela, Colombia and Mexico have demonstrated the potential for VEEV to re-emerge periodically from enzootic progenitors [15–18]. The emergence of VEEV into an epidemic/epizootic form has been associated with specific mutations that arise in the VEEV envelope glycoprotein 2 (E2) gene of enzootic subtype ID or IE strains. These mutations result in the

addition of positively charged amino acid changes on the surface of the virion spikes [19] that give rise to increased virulence and viremia in equids [20, 21] and sometimes enhanced infection of epidemic vector mosquitoes such as *Aedes (Ochlerotatus) taeniorhynchus* [22]. The higher viremia levels in equids can lead to infection of mosquitoes that are not normally involved in enzootic circulation [23], which can then result in spillover infections of humans and other domestic animals.

The phylogenetic characteristics of the VEE complex have been studied for several decades, recently focusing on structural protein gene sequences [24–27]. These studies support the hypothesis that the IAB and IC subtypes arise from mutations in enzootic ID strains that result in the acquisition of epizootic/epidemic characteristics [19, 20, 28]. Although some studies have addressed the evolution and continued circulation of VEEV ID and IE strains, the use of only partial genomic sequences has limited their resolution and accuracy.

To date, insufficient complete genomic sequences have been available to permit a detailed, global analysis of all VEE complex species/strains and obtain high-resolution phylogenetic results. Our goal was to determine more accurately the temporal origin of the VEE complex and patterns of historic spread. By increasing the number of sequenced VEEV strains from the ID and IE subtypes, we sought a more robust analysis of the evolution of these subtypes. To this end, a set of 130 complete genome sequences was prepared, of which 94 were determined in this study, and a comprehensive phylogenetic study was performed to determine the origin and evolutionary patterns of these important viruses.

Materials and methods

Virus isolates

Viruses were obtained from the World Reference Center for Emerging Viruses and Arboviruses and other collections at the University of Texas Medical Branch, and included 94 isolates that had not been previously sequenced or had only partial genome sequences available. These strains were geographically and temporally distributed across North and Central America, and over the past 80 years. The metadata for the 94 virus strains sequenced in this study and their accession numbers are found in [S1 Table](#).

Virus propagation and cDNA generation

Vero (African green monkey kidney) cells (CCL-81, ATCC) were grown in 150 cm² flasks to ~80% confluency and inoculated with VEE complex virus strains. [S1 Table](#) shows the strains included in this study and their associated metadata. Infected cells were maintained at 37°C until the development of cytopathic effects. Then, cell culture supernatants were clarified at 1,125 *x g* for 10 min and mixed with a 1/3 volume of 4X precipitation buffer (28% PEG 8000, 9.2% NaCl). After overnight incubation at 4°C, virus was pelleted at 2,880 *x g* for 30 min at 4°C and resuspended in 250 µl of TEN buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 0.1 M NaCl), which was then added to 750 µl of TRIzol LS Reagent (Invitrogen, Grand Island, NY). RNA was extracted following the manufacturer's protocol, then resuspended in 50 µl of H₂O and stored at -80°C.

The SuperScript III First-Strand Synthesis System (Invitrogen, Grand Island, NY) was used to produce cDNA following the manufacturer's recommendations. Three 20 µl-reactions were performed, each with 6 µl of extracted RNA and 1 µl of one of the following cDNA primers, 50 ng/µl of random hexamers, 50 µM oligo (dT)₂₀, and 10 µM of reverse primers designed to anneal approximately 500 nt downstream of the 5' end of the viral genome. Samples were treated with RNase H, and the resulting cDNA generated from random hexamers and oligo (dT)₂₀ primers were then mixed together. The cDNA samples were stored at -80°C until

further processing, and efficient reverse transcription was confirmed by PCR using 0.5 μ l of each sample and primer pairs designed to anneal near the 5' and 3' portions of the genome.

Sequencing of VEEV genomes

Sequences were assembled using sequence-independent single primer amplification (SISPA) to barcode random primed cDNAs [29, 30] from individual cDNA samples. SISPA products were normalized and pooled into a single reaction that was purified using a PCR purification kit (Qiagen, Valencia, CA). Samples were subsequently gel purified to select for products ranging from 300-500bp in size for sequencing with the Illumina Genome Analyzer II or 500-800bp in size for Roche 454 Titanium (GS-FLX) sequencing [31], or were sequenced on the Illumina HiSeq using the following protocol; cDNA (0.05–1.7 μ g) was fragmented by incubation at 94°C for eight (8) minutes in 19.5 μ l of fragmentation buffer (Illumina 15016648). Samples were tracked using the “index tags” incorporated into the adapters as defined by the manufacturer. Cluster formation of the library DNA templates was performed using the TruSeq PE Cluster Kit v3 (Illumina) and the Illumina cBot workstation using conditions recommended by the manufacturer. Paired end 50 base sequencing by synthesis was performed using TruSeq SBS kit v3 (Illumina) on an Illumina HiSeq 1500 using protocols defined by the manufacturer.

Assembly of VEEV genomes

Next generation sequencing (NGS) reads from Roche 454 GS-FLX were sorted based on SISPA barcode matches, trimmed, and searched by TBLASTX against a custom reference nucleotide database of full-length VEE complex genomes downloaded from GenBank. Any chimeric VEEV sequences or non-VEEV sequences amplified during the random hexamer-primed amplification were removed. For each sample, the filtered GS-FLX reads were then *de novo* assembled using CLC Bio's *clc_novo_assemble* program. The consensus sequence of the *de novo* assembly was used to identify the best full-length VEEV genome downloaded from GenBank to use as a mapping reference sequence. Both GS-FLX and Illumina reads were then mapped to the selected reference VEEV genome using CLC Bio's *clc_ref_assemble_long* program. At loci where both GS-FLX and Illumina sequence data agreed on a variant compared to the reference sequence, the latter was updated to reflect the difference. A final mapping of all sequences to the updated reference sequences was then performed, resulting in the final assembled genome. Upon review of NGS assemblies, there were circumstances that required the use of RT-PCR followed by Sanger capillary sequencing to fill gaps in genomic regions with low coverage. These cases included finishing/closure tasks to increase the sequencing coverage of genome regions inadequately covered by NGS.

Sequences identified with an asterisk in [S1 Table](#) were part of a different sequencing project and were therefore analyzed using a different platform. These sequences were first subjected to a blast analysis to determine the number of viral reads and the percentage of contaminating reads from hosts, or other sources. Then the contamination reads from the Vero cells were filtered out using the African Green Monkey (*Chlorocebus sabeus*) genome as a template. The remaining Fastq files were additionally processed using trimmomatic to remove low quality sequence. Finally, assembly was performed using iMetAMOS [32] using the standard parameters.

Data sets

Nucleotide sequences were manually aligned with similar GenBank sequences using MUSCLE implemented in SeaView [33]. Untranslated regions (UTRs) of alphavirus genomes show

limited conservation and are often difficult to align reliably. Therefore, the sequences were edited and trimmed to remove untranslated regions, resulting in concatenated ORFs for 130 (126 VEEV and 4 EEEV) sequences. Sequences were then re-aligned as protein sequences before being reverse translated to nucleotides to maintain codon alignments. Within the VEEV genome, there are two regions that have proved historically difficult to align, the 3' end of the nsP3 and the 5' end of the capsid gene [34]. Thus these were removed from the alignment and the resulting alignment was used for subsequent analyses and this alignment is available upon request. Eastern equine encephalitis virus, the sister virus to VEEV in the alphavirus genus [35], was used as an outgroup in all analyses except molecular dating and selection. Sequences were analyzed for percent identity at the nucleotide and amino acid level using BioEdit [36].

Phylogenetic trees

A maximum likelihood (ML) phylogeny was inferred using PAUP* 4.0 [37] and the General Time Reversible (GTR + Γ_4 +I) model, selected by Modeltest [38]. To assess the robustness of tree topologies, bootstrapping was performed using 1,000 replicate neighbor-joining trees. Bayesian phylogenetic inference was performed using the GTR + Γ_4 +I model in MrBayes v3.1 [39, 40]. Analyses were run for one million iterations until they reached convergence.

Estimated evolutionary rates and dates of divergence

Substitution rates and times to most recent common ancestor (tMRCAs) were estimated using Bayesian evolutionary analysis by sampling trees (BEAST) v1.7.1 [41, 42]. BEAST employs a Bayesian Markov chain Monte Carlo (MCMC) approach to infer demographic histories, evolutionary rates and dates of divergence from serially (dated) sampled sequence data. Statistical uncertainty in the data is reflected in the 95% highest posterior density (HPD) values. Analyses were typically performed using the Bayesian Skyline Plot (BSP) model of population growth, which does not use a pre-specified demographic model [43]. We used this model because the VEE complex would not all show the same patterns of demographic change and this did not impose constraints on the different subtypes of VEEV. To estimate substitution rate variation among lineages we used the models implemented in the BEAST program [44] including the uncorrelated lognormal (UCLN) model for the VEEV IE strains and the uncorrelated exponential (UCEX) model for the VEEV ID/II/IAB/IC strains as these were determined to be the most accurate model using the stepping stone algorithm implemented in BEAST [45]. The MCMC chain was 100 million generations long, thinned to include every 5000th generation in the final sample. The program Tracer version 1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>) was used to confirm convergence and mixing. The software TreeAnnotator version 1.7.1 (<http://beast.bio.ed.ac.uk/software/TreeAnnotator>) was used to summarize the data output from BEAST. The maximum clade credibility (MCC) tree was estimated using mean node heights after discarding the initial 10% of generations as burn-in.

Selection analysis

To investigate the nature of selective pressures acting on VEE complex viruses, we estimated the average number of nonsynonymous (d_N) and synonymous (d_S) nucleotide substitution per site (d_N/d_S ratio), using a counting method SLAC [46], as well as the numbers and locations of sites experiencing episodic positive selection using MEME [47].

In the presence of strong purifying selection, standard models of molecular evolution (e.g. GTR + Γ_4 +I) tend to underestimate branch lengths in RNA and DNA virus phylogenies [48–50]. Therefore we re-estimated branch lengths using a Branch-Site REL (BRSEL) [51] model

that accounts for variation in selection pressure across the genome and across different branches in the phylogeny, using the HyPhy software package (50). The original formation of the BSREL model allowed three ω (d_N/d_S) rate classes on each branch, each representing a proportion of sites in the alignment. To prevent over-fitting, we implemented a step-up procedure via an adaptive BSREL model (aBSREL; [52]), starting with a single d_N/d_S class on each branch and testing the fit of an additional d_N/d_S class using small-sample corrected Akaike information criteria (c-AIC).

Recombination analyses

All sequences were analyzed for recombination using the program RDP3 [53]. Recombination events were determined using RDP, GENECONV, MaxChi, Chimaera and 3Seq, and were only considered robust if they were identified using more than one method.

Results

Phylogenetic analysis of the VEE complex strains

The unrestricted Bayesian (MrBayes) and maximum likelihood trees (PAUP*) inferred the expected topology with Cabassou virus (CABV, subtype V), Rio Negro virus (RNV, subtype VI), and Mosso das Pedras virus (MDPV, subtype IF) falling basal to the remainder of the subtypes, as has been shown in previous analyses (Fig 1) [25, 34]. Unexpectedly, the MCC tree, inferred using BEAST (S1 Fig) did not resolve the placement of CABV, RNV, and MDPV, with this grouping showing low posterior support.

Mutations that were unique to each subtype or group of subtypes were identified, as well as those synapomorphies that defined lineages within subtypes; uninformative sites were not counted. For VEEV subtype IE, there were significantly more synapomorphic mutations in the structural protein genes than would be expected if mutations were randomly distributed across the genome ($p < 0.05$, Fisher's exact test). For all other subtypes and groups of subtypes, the distribution of mutations was not significantly different from expected given the length of the ORF's. For the VEE complex subtypes III-VI, the number of mutations was identified on each branch (See S1 Fig). There was no correlation between branch length and the number of amino acid substitutions unique to each of these subtypes.

Correcting for the effect of purifying selection when estimating tMRCAs and evolutionary rates

The predominance of purifying selection in alphavirus evolution has been widely reported [54, 55], and this type of selection has been postulated to bias tMRCA estimations for ancient divergence events in RNA viruses [48, 49]. To test our hypothesis that, like other ancient RNA viruses [48–50], dating estimates may be biased in the VEE complex due to the presence of strong purifying selection, we employed a branch-site random effects likelihood (BSREL) approach to determine the extent to which branch lengths (and tMRCAs) were underestimated by standard models of nucleotide evolution. By moving up through the VEE complex phylogeny in an iterative process, we removed the most basal lineages and compared total tree length optimized under BSREL and GTR+ Γ_4 , thereby identifying subtrees whose branch lengths did not expand under BSREL compared with a standard nucleotide model (GTR+ Γ_4) (Figs 1 & 2). We assert that it is these subtrees, which were not underestimated by GTR+ Γ_4 , whose tMRCA can be reliably inferred by standard Bayesian molecular clock dating (e.g. BEAST) (Fig 2).

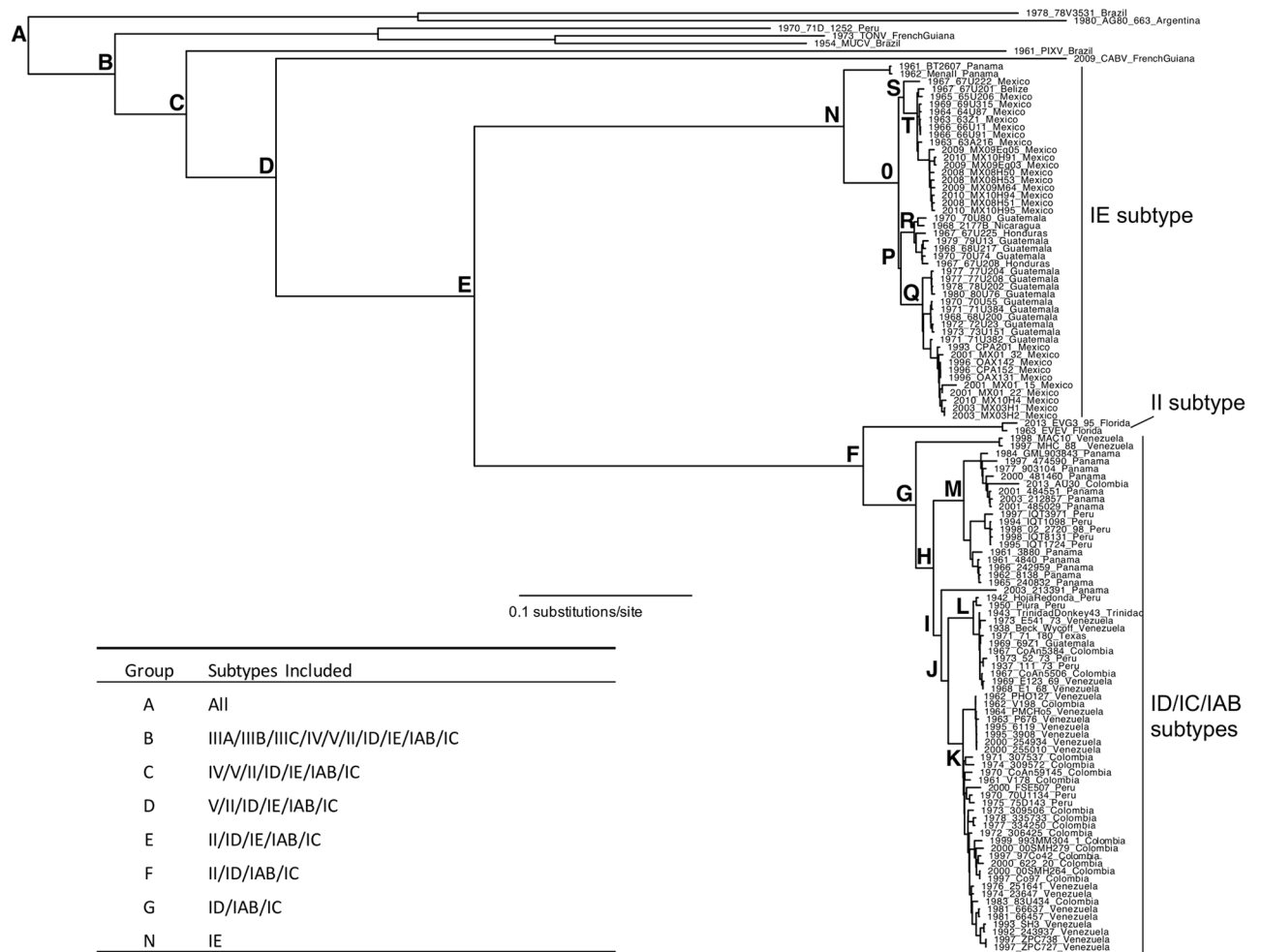


Fig 1. Maximum likelihood phylogeny for VEEV complex. Relevant internal nodes are identified.

<https://doi.org/10.1371/journal.pntd.0005693.g001>

As we could not generate a fully resolved MCC tree for the entire VEE complex, two trees were generated for subtypes IAB/IC/ID and for subtypes IE. Removing the other branch lengths removed some of the uncertainty regarding the substitution rates as well as the lack of samples for many of the other subtypes, which can skew the data. The subtype IAB/IC/ID VEEV strains fell into two main groups, a Panamanian lineage and a Venezuelan/Colombian lineage. In addition, we sequenced two additional strains MAC10 and MHC88 from Venezuela, which are outliers from the rest of the ID/IAB/IC strains (Fig 3). These strains showed that VEEV has been circulating in its current form for around 253 years, although the date of this node could not be accurately determined as determined by the BSREL analysis. In fact only group K (Figs 1 and 3), which fell within the Colombian/Venezuelan lineage could be reliably dated at 1934 (1904–1950) (without BSREL correction). The Venezuelan samples MAC10 and MHC88 were basal to the rest of the strains. Within the Venezuelan/Colombian and Panamanian lineages, the Peruvian strains were present in both lineages and, given the shape of the tree and the distribution of these viruses, it is most likely that they represent recent introductions. Although synapomorphic amino acids were identified for the major lineages of the ID subtype (see S2 Fig) the distribution of the amino acid substitutions between the structural

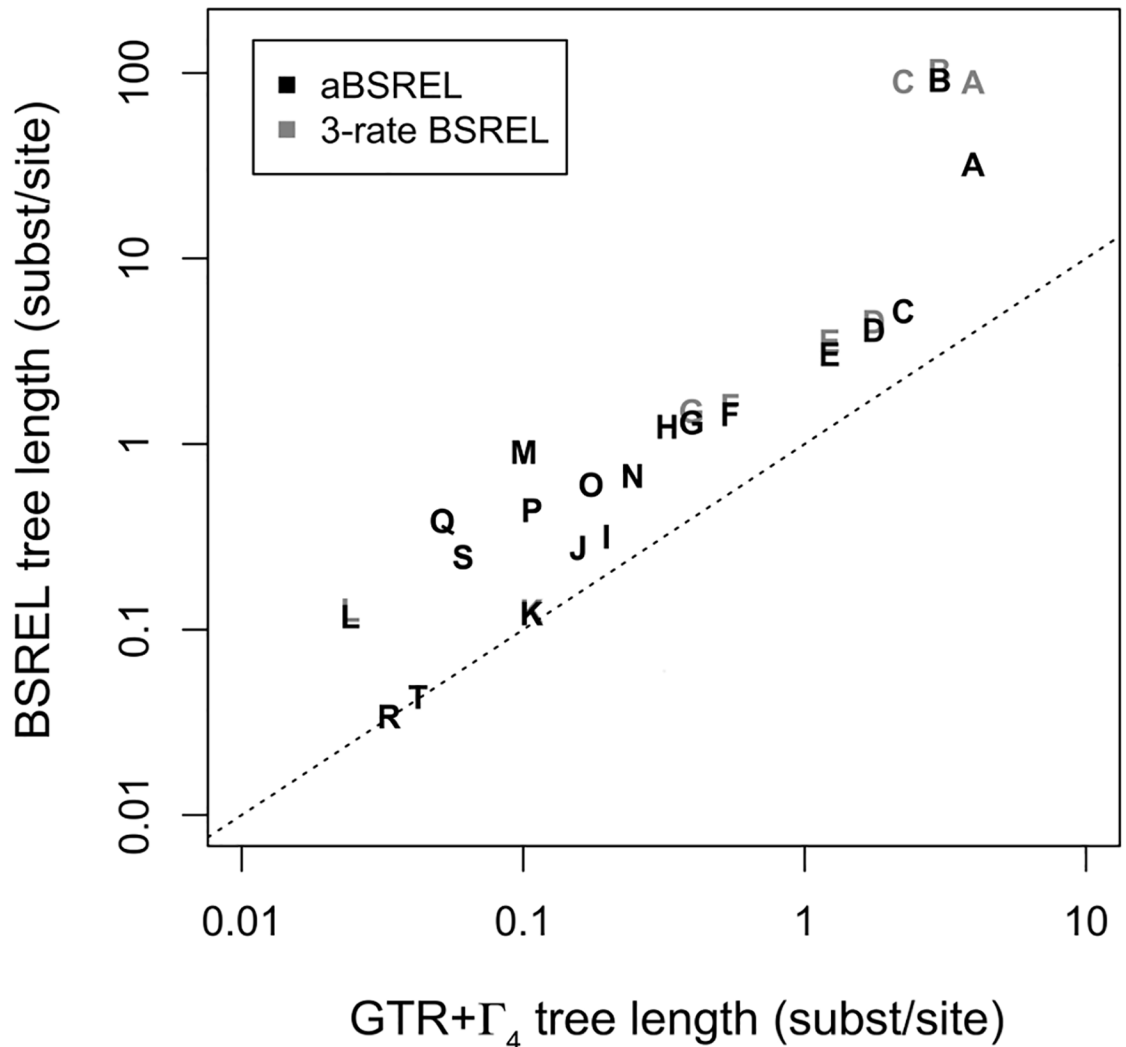


Fig 2. Evidence of branch length underestimation in the VEE complex shown via comparison of tree lengths inferred under Branch-Site REL (BSREL) and GTR+ Γ_4 substitution models. Letters correspond to nodes in maximum likelihood tree. The aBSREL analysis performed with an optimized number of rate classes is shown in black. BSREL analysis performed with a fixed number of three rate classes is shown in gray. Instances in which only a black letter is shown indicate that aBSREL and BSREL produced identical results. The dashed line depicts $x = y$, an unbiased analysis.

<https://doi.org/10.1371/journal.pntd.0005693.g002>

and non-structural proteins for each lineage was not significantly different from expected based on a random distribution across the genome.

All of the subtype IAB strains occupied a single clade (see Fig 3), with the Hoja Redonda (1942.HojaRedonda.Peru) and Piura (1950.Piura.Peru) strains forming the most basal branch. Previous analyses suggested that the IAB outbreaks after 1943 were caused by incompletely-inactivated vaccines derived from early IAB isolates [27], consistent with our results. These vaccines were used in South and Central America until the early 1970s, when the TC-83 live-attenuated strain replaced them after its demonstrated efficacy during the 1971 Texas outbreak and in experimental studies [56]. The mechanism of the original IAB emergence is believed to be mutations in the E2 protein gene of enzootic ID strains [19]. Unfortunately, the ID progenitor strains for this group have not been identified.

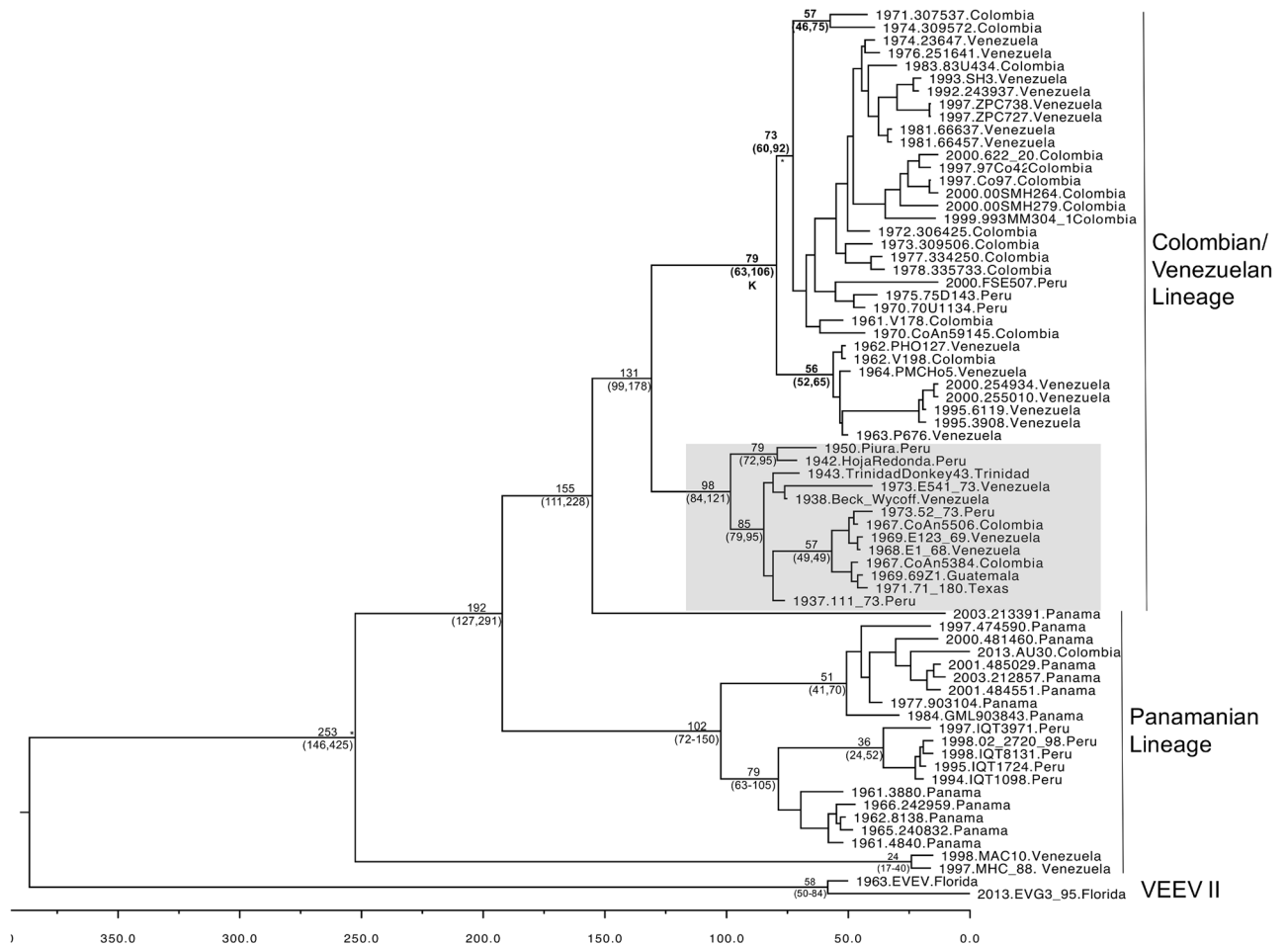


Fig 3. The genetic relationships of the II/ID/IAB/IC subtypes of the VEEV complex. MCC tree as determined by coalescent analysis for the ID/II strains with the IAB and IC strains included. All branches had posterior probabilities of >0.97 except for the branches marked with a *, which did not have significant support. The time to most recent common ancestors (tMRCA) dates for individual branches as identified from the MCC tree are indicated on significant branches with the highest posterior density (HPD) in brackets, dates of clades supported by BSREL were emboldened. Subtype IAB strains are highlighted by the grey box.

<https://doi.org/10.1371/journal.pntd.0005693.g003>

For the subtype IE, a separate coalescent analysis was performed (Fig 4). The BT2607 and MenaII isolates were basal in this tree as expected given previous analyses [16]. These viruses were isolated in the 1960s and no further strains are available from this lineage. The remaining subtype IE strains fell into two major groups with the tMRCA of 91 (68–124) years before the present. The Pacific Coast strains contained samples from the entire known geographic range of VEEV subtype IE, except Panama (Guatemala, Honduras, Nicaragua and Mexico) and the Gulf Coast strains included isolates from Mexico only, and a single strain from Belize. The most recent VEEV strains from Mexico were found in this second group. However, the paucity of recent isolates from Central America suggests that some of the temporal groupings in our trees may reflect sampling bias. The Gulf Coast IE strains diverged after the Pacific Coast strains 76 (95% HPD; 58–102) years prior to 2010, and the Pacific Coast strains diverged 85 (95% HPD; 63–116). However, only the clades designated R and T (Figs 1 & 4) could be reliably dated.

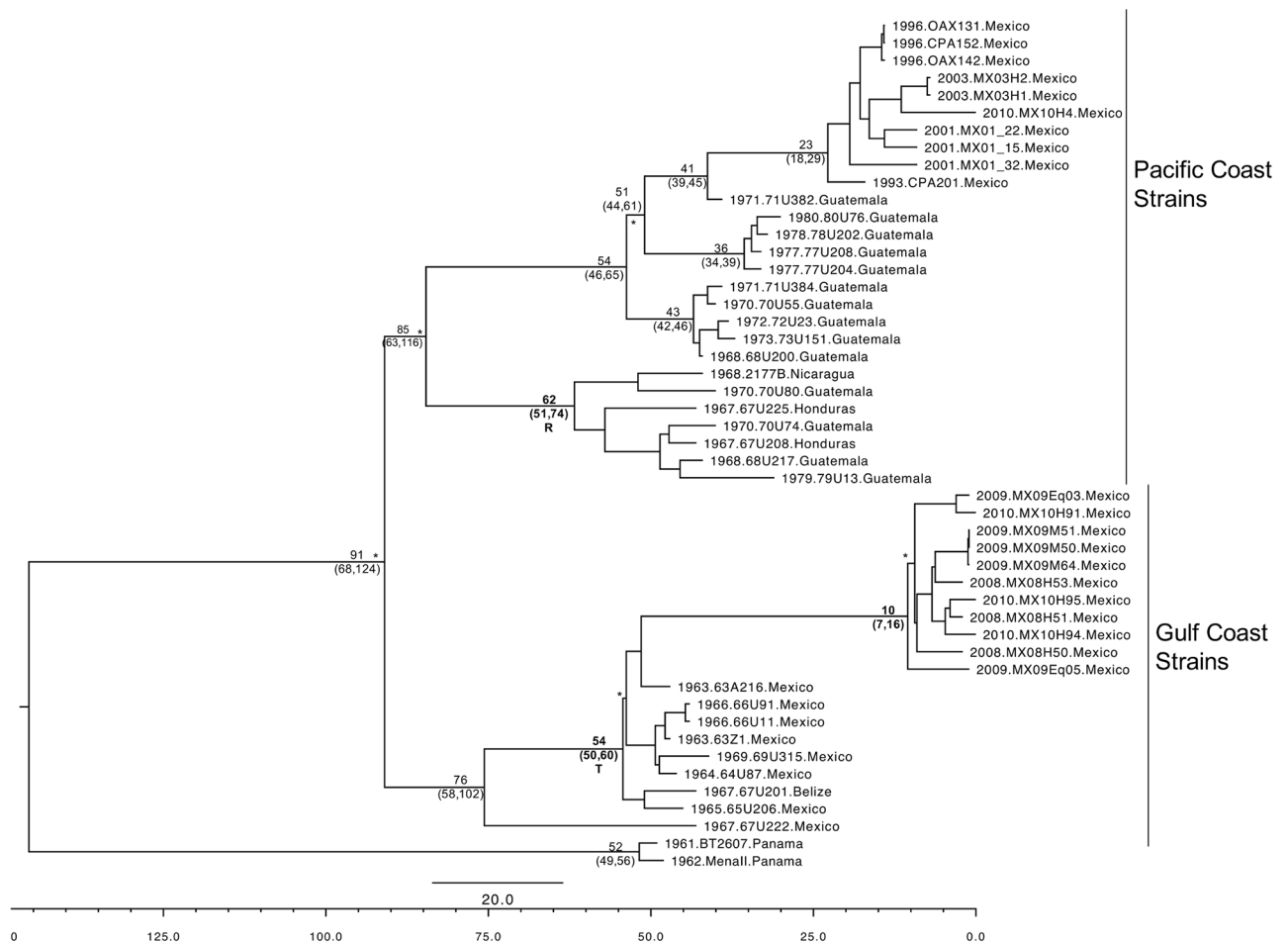


Fig 4. The genetic relationships of the IE subtype of the VEEV complex. MCC tree as determined by coalescent analysis for the IE strains. All branches had posterior probabilities of >0.97 except for the branches marked with a *, which did not have significant support. Time to most recent common ancestor (tMRCA) as determined by the MCC tree were added to major branches with the HPD in brackets. If the clades were supported by BSREL the tMRCA and HPD were emboldened.

<https://doi.org/10.1371/journal.pntd.0005693.g004>

Recombination analysis

There was evidence of VEEV recombination between nucleotides 4800–5830 observed using the program RDP. However, this region corresponds to the nsP3 gene, which is highly variable with frequent insertions and deletions, even within subtypes. Aligning the insertions and deletions is problematic, so this recombination result cannot be confirmed.

Selection analysis

Using SLAC, we estimated a global d_N/d_S ratio of 0.057, suggesting that purifying selection is the predominant evolutionary force acting on the VEEV genome. A total of 2900 negatively selected sites were detected using the SLAC algorithm (at $p < 0.1$). Nonetheless, we found evidence for episodic diversifying positive selection at 23 codons using MEME (at $p < 0.05$; identified relative to their position within each protein relative to prototype VEEV strain 3880 in Table 2). Several positively selected codons were detected in important protein genes including the nsP4 RNA-dependent RNA polymerase, Capsid, E2 and E1 glycoproteins. However, these sites did not include substitutions previously demonstrated experimentally to mediate

Table 2. Codon positions predicted to be under episodic positive selection, showing the amino acid changes and the proteins altered. Amino acid position is defined using the Genome sequence 3908 (Accession number [U55350](https://www.ncbi.nlm.nih.gov/nuccore/U55350) in the NCBI database) and numbered against the non-structural and structural polyproteins.

Protein	Amino acid position (ORF)	p-value	Amino acid change	Protein function
nsP1	3	0.021	K→T	NA
nsP1	130	0.016	T→Q/R	Viral methyltransferase
nsP1	444	0.002	N→P/S/D	NA
nsP1	487	0.049	I→V/L	NA
nsP2	884	0.007	P→S	Viral Helicase
nsP3	1344	0.045	Y→M/I/E/S	Appr -1 -p processing enzyme
nsP4	1921	0.031	S→V/A	RNA polymerase
nsP4	1991	0.043	L→E	RNA polymerase
nsP4	2162	0.014	M→L/A	RNA polymerase
nsP4	2397	<0.001	C→S/R	RNA polymerase
Capsid	10	0.014	M→T	Capsid
Capsid	66	0.007	P→K/Q/R/S	Capsid
Capsid	92	0.025	K→G/P/Q/R/S	Capsid
Capsid¹	100	0.028	A→N/T/PV/H/G	Capsid
E3	303	0.006	A→S/V	E3 glycoprotein
E2	451	0.046	D→E/G/N/K	E2 glycoprotein
E2	527	0.044	G→R	E2 glycoprotein
E2	533	0.016	E→D/K	E2 glycoprotein
E2	547	0.001	T→K/S/R/Q	E2 glycoprotein
E2	647	0.003	E→H/N/K	E2 glycoprotein
E1	819	0.038	T→S	E1 glycoprotein
E1	976	0.020	A→V	E1 glycoprotein
E1	1153	0.025	K→R	E1 glycoprotein

¹Protein change associated with epidemic strains of IE in 1996 and 2003 [58].

<https://doi.org/10.1371/journal.pntd.0005693.t002>

adaptation for equine amplification or bridge vector infection involved in epizootic VEE emergence [21, 57], underscoring the limitations of d_N/d_S -based selection analyses.

Discussion

VEEV continues to be a major public health problem in Latin America and understanding the evolution and spread of this virus in the New World is integral to implementing surveillance strategies and prevention programs. Although the genetic mechanisms of emergence of epidemic strains is relatively well understood [19, 20, 28, 59], the evolution of the group as a whole has not been studied in a comprehensive manner. Using Bayesian molecular dating techniques, we investigated the evolutionary dynamics of the VEE complex while taking into consideration the effect of purifying selection on the accuracy of dating ancient nodes within this group of viruses. The genomic sequencing of 94 VEEV isolates from all known endemic countries, the inclusion of most isolates from epidemic/epizootic events, and incorporation of the known temporal distribution of available isolates facilitated an extensive investigation into the origin and evolutionary history of this important group of pathogens.

We estimated dates of divergence for several subtypes in the VEE complex. However, the presence of purifying selection can bias such analyses, and our attempts to date the entire VEE complex were limited by this bias. Analysis of ID/IAB/IC strains revealed two main lineages. One circulates primarily in Panama, with a few Peruvian strains included. The origin of this

group was estimated at around 102 years ago, although confidence limits were broad. From the phylogeny we were able to infer that circulation of VEEV may have been present initially in Panama, with a subsequent introduction into Peru. The exact mechanism of introduction into Peru is not understood; the enzootic cycle of VEEV involves rodents and mosquitoes that have limited geographic range, suggesting other mechanisms of transport such as birds may be occasionally be involved in virus movement. The presence of Peruvian sequences in two distinct lineages suggests that there have been at least two independent introductions. Exclusion of VEEV subtypes III-VI and the lack of isolates from countries intermediate between Panama and Peru will therefore limit the possibility of fully resolving the ancestral dispersal of these viruses.

As in previous analyses, our results confirm that the IAB strains group together with strain Hoja Redonda, isolated in Peru in 1942, which is basal in the group. Previous partial sequencing and analysis of these strains suggested that some of the IAB outbreaks were a result of incompletely inactivated vaccines [27]. We were able to corroborate this finding using full genome strains that showed tight genetic distance (96.08–99.23%), and no identified ID progenitors (i.e. strains that are genetically related but lack the defining E2 mutations that are associated with the IAB subtypes), although this could be due to inaccurate sampling. However, compared to the spacing of the strains comprising the IC subtype, which are not known to have been used for vaccine production and are interspersed throughout the phylogeny, all the IAB strains appear to have a single origin, and previous analyses have shown that these cluster with vaccine strains used around the time these viruses were circulating [27].

The IC subtype, which has been characterized both phylogenetically and experimentally using reverse genetics, was determined to be a result of amino acid changes in the E2 protein of enzootic subtype ID strains [28, 59]. The IC strains are found within 2 distinct clades in the VEEV phylogeny, particularly in the Colombian/Venezuelan lineage. However, our data suggest that no IC epizootic/epidemic strains have arisen from the Panamanian lineage. It is possible that epistatic interactions limit the emergence of IC epidemic mutations from these Panamanian ID strains, as has been described for chikungunya virus emergence [60, 61]. Despite strong experimental evidence that substitutions in the E2 protein mediate adaptation for equine amplification or bridge vector infection involved in epizootic VEE emergence [21, 57], our sequence analyses failed to detect positive selection on the corresponding codons (Table 2). This finding underscores the limitations of current d_N/d_S methods for identifying unique adaptive substitutions occurring during virus evolution, as they typically require repeated substitution events at a given site to detect adaptive evolution.

The IE VEEV subtype is confined to Central America and Mexico; in fact there appears to be a demarcation between the ID subtype and the IE subtype at the Panamanian/Costa Rican border with the exception of the BT2607 and Menall strains isolated in western Panama in 1961 and 1962, respectively (Fig 1). More surveillance in Panama is required to determine if these VEEV IE subtypes are still circulating. The main VEEV IE lineage is subdivided into two clades: a widely dispersed group with strains from Guatemala, Honduras, Nicaragua and Mexico (Pacific Coast lineage) and a group with more limited dispersal circulating mainly in Mexico (Gulf Coast lineage). As is expected of a rodent-hosted arbovirus, there is very little spread among countries and the dispersal within the IE subtype appears to occur between neighboring countries, presumably reflecting the limited mobility of rodents and mosquitoes (Fig 4). Given the limited number of countries represented in the IE phylogeny and the lack of recent sampling in most countries, it is hard to draw further conclusions about the historical spread of this subtype within Central America. However, recent detailed studies of VEEV circulation in Mexico demonstrate geographic stability of independently evolving lineages in that country [62].

Previous phylogenetic analyses of the VEE complex viruses were performed using partial and complete structural protein gene sequences [25]. Using these expanded sequences we observed 63 synapomorphic mutations associated with the IE subtype (Group H). Of these, 39 were in the non-structural protein genes and 37 in the structural genes. This finding was significantly different from an expected, random distribution throughout the VEEV genome. It is possible that the preponderance of structural protein substitutions reflects adaptation to different mosquito vectors. Subtype ID strains are vectored by at least three mosquito species: *Cx. (Melanoconion) adamesi*, *Cx. (Mel.) vomerifer* and *Cx. (Mel.) pedroi* [63], whereas the subtype IE strains appear to be more specialized and vectored nearly exclusively by *Cx. (Mel.) taeniopus* [64]. Experimental infections with subtypes IAB, IC and ID have demonstrated poor infectivity at the level of midgut infection, suggesting specific subtype IE adaptation to this vector [64, 65]. Although the true distributions of *Cx. taeniopus* and *Cx. pedroi* are not completely certain because they were not distinguished until 1980 [66], recent collection records and revisions [63, 67–73] suggest that the former is restricted to Mexico, Central America and the Caribbean, while the latter occurs throughout northern South America and Central America, as well as in Mexico. Additionally, neither *Cx. vomerifer* or *Cx. adamesi*, both subtype ID vectors, are found north of Panama [74, 75], which may restrict the distribution of the ID subtype. Regardless of the original mechanism of spatial compartmentalization, between these subtypes, the IE strains appear to have adapted specifically to *Cx. taeniopus* while the other subtypes remain poorly infectious for this species [76].

In summary, our comprehensive analysis of all available full-length VEE complex genomes demonstrated that purifying selection as a confounding factor in coalescent analyses and only the more recently diverged subtypes could have their tMRCA reliably estimated. Purifying selection is an inherent determinant of arbovirus evolution because of the alternating vector-host transmission and could therefore introduce bias into coalescent analyses of similar viruses. By restricting this bias by analyzing only those clades for which the BSREL analysis indicated that standard nucleotide models would not introduce severe bias we were able to robustly estimate the tMRCA of several VEEV lineages. Evolution of the IE subtype appears to have been characterized by a significant evolutionary shift from the rest of the VEEV complex, with an increase in structural protein substitutions that may reflect adaptation to its mosquito vector. Additionally our inferred tree topologies, suggest that VEEV is maintained primarily in limited geographic foci with only occasional spread to neighboring countries, probably reflecting the limited mobility of rodent hosts and mosquito vectors. Additional strains of subtypes II–VI are needed to more completely characterize the evolution of the entire VEE complex of alphaviruses.

Supporting information

S1 Table. List of virus strains used in the study and known metadata.

(PDF)

S1 Fig. MCC tree as determined by coalescent analysis for all the VEEV subtypes. Numbers on the branch lengths show the number of unique amino acids that are associated with each particular subtypes.

(PDF)

S2 Fig. Unique amino acids associated with each subtype.

(PDF)

Acknowledgments

The authors wish to thank Robert Tesh and Hilda Guzman of the WRCEVA for providing many of the viruses used in this study. We thank Nadia Fedorova for reassembly and validation of sequences.

Author Contributions

Conceptualization: Naomi L. Forrester, Joel O. Wertheim, Martin D. Smith, Sergei L. Kosakovsky Pond, Scott C. Weaver.

Data curation: Suman R. Das.

Formal analysis: David Lin, Kumar Hari, Ravi Jain, Timothy B. Stockwell, David E. Wentworth, Martin D. Smith, Sergei L. Kosakovsky Pond.

Funding acquisition: Vivian G. Dugan, Scott C. Weaver.

Investigation: Naomi L. Forrester, Joel O. Wertheim, Vivian G. Dugan, Albert J. Auguste, David Lin, A. Paige Adams, Rubing Chen, Rodion Gorchakov, Grace Leal, Jyotsna Pandya, Rebecca A. Halpin, Kumar Hari, Ravi Jain, Timothy B. Stockwell, David E. Wentworth, Martin D. Smith, Sergei L. Kosakovsky Pond, Scott C. Weaver.

Methodology: Naomi L. Forrester, Joel O. Wertheim, Vivian G. Dugan, Scott C. Weaver.

Project administration: Vivian G. Dugan, Scott C. Weaver.

Resources: Jose G. Estrada-Franco.

Writing – original draft: Naomi L. Forrester, Joel O. Wertheim, Albert J. Auguste, Scott C. Weaver.

Writing – review & editing: Naomi L. Forrester, Joel O. Wertheim, Albert J. Auguste, David E. Wentworth, Scott C. Weaver.

References

1. Weaver SC, Winegar R, Manger ID, Forrester NL. Alphaviruses: Population genetics and determinants of emergence. *Antiviral Res.* 2012. Epub 2012/04/24. <https://doi.org/10.1016/j.antiviral.2012.04.002> PMID: [22522323](https://pubmed.ncbi.nlm.nih.gov/22522323/).
2. Powers AM, Huang HV, Roehrig JT, Strauss EG, Weaver SC. *Togaviridae*. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, editors. *Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses*. Oxford: Elsevier; 2011. p. 1103–10.
3. Blosser EM, Burkett-Cadena ND. *Culex (Melanoconion) panocossa* from peninsular Florida, USA. *Acta Trop.* 2017; 167:59–63. <https://doi.org/10.1016/j.actatropica.2016.12.024> PMID: [28012907](https://pubmed.ncbi.nlm.nih.gov/28012907/)
4. Zacks MA, Paessler S. Encephalitic alphaviruses. *Veterinary Microbiology.* 2010; 140:281–6. <https://doi.org/10.1016/j.vetmic.2009.08.023> PMID: [19775836](https://pubmed.ncbi.nlm.nih.gov/19775836/)
5. Armstrong PM, Andreadis TG. Eastern equine encephalitis virus—old enemy, new threat. *N Engl J Med.* 2013; 368(18):1670–3. Epub 2013/05/03. <https://doi.org/10.1056/NEJMp1213696> PMID: [23635048](https://pubmed.ncbi.nlm.nih.gov/23635048/).
6. Calisher CH, Murphy FA, France JK, Lazuick JS, Muth DJ, Steck F, et al. Everglades virus infection in man, 1975. *Southern Medical Journal.* 1980; 73(11):1548. PMID: [7444536](https://pubmed.ncbi.nlm.nih.gov/7444536/)
7. Bigler WJ, Ventura AK, Lewis AL, Wellings FM, Ehrenkranz NJ. Venezuelan equine encephalomyelitis in Florida: endemic virus circulation in native rodent populations of Everglades hammocks. *American Journal of Tropical Medicine and Hygiene.* 1974; 23:513–21. PMID: [4150911](https://pubmed.ncbi.nlm.nih.gov/4150911/)
8. Aguilar PV, Greene IP, Coffey LL, Medina G, Moncayo AC, Anishchenko M, et al. Endemic Venezuelan equine encephalitis in northern Peru. *Emerging infectious diseases.* 2004; 10(5):880–8. <https://doi.org/10.3201/eid1005.030634> PMID: [15200823](https://pubmed.ncbi.nlm.nih.gov/15200823/).
9. Demucha Macias J, S'Anchez Spindola I. Two Human Cases of Laboratory Infection with Mucambo Virus. *The American journal of tropical medicine and hygiene.* 1965; 14:475–8. PMID: [14292756](https://pubmed.ncbi.nlm.nih.gov/14292756/).

10. Aguilar PV, Estrada-Franco JG, Navarro-Lopez R, Ferro C, Haddow AD, Weaver SC. Endemic Venezuelan equine encephalitis in the Americas: hidden under the Dengue umbrella. *Future Virology*. 2011; 6:721–40. PMID: [21765860](#)
11. Aguilar PV, Estrada-Franco JG, Navarro-Lopez R, Ferro C, Haddow AD, Weaver SC. Endemic Venezuelan equine encephalitis in the Americas: hidden under the dengue umbrella. *Future Virol*. 2011; 6(6):721–40. Epub 2011/07/19. PMID: [21765860](#); PMC3134406.
12. Lord RD. History and geographic distribution of Venezuelan equine encephalitis. *Bulletin of the Pan American Health Organisation*. 1974; 8:100–10.
13. Sudia WD, Newhouse VF, Beadle LD, Miller DL, Johnston JG Jr, Young R, et al. Epidemic Venezuelan equine encephalitis in North America in 1971: Vector studies. *American Journal of Epidemiology*. 1975; 101:17–35. PMID: [235212](#)
14. Walton TE, Grayson MA. Venezuelan equine encephalomyelitis. In: Monath TP, editor. *The Arboviruses: Epidemiology and Ecology*. IV. Boca Raton, Florida: CRC Press; 1988. p. 203–31.
15. Brault AC, Powers AM, Chavez CL, Lopez RN, Cachon MF, Gutierrez LF, et al. Genetic and antigenic diversity among eastern equine encephalitis viruses from North, Central, and South America. *American Journal of Tropical Medicine and Hygiene*. 1999; 61:579–86. PMID: [10548292](#)
16. Oberste MS, Fraire M, Navarro R, Zepeda C, Zarate ML, Ludwig GV, et al. Association of Venezuelan equine encephalitis virus subtype IE with two equine epizootics in Mexico. *American Journal of Tropical Medicine and Hygiene*. 1998; 59:100–7. PMID: [9684636](#)
17. Rivas F, Diaz LA, Cardenas VM, Daza E, Bruzon L, Alcala A, et al. Epidemic Venezuelan equine encephalitis in La Guajira, Colombia, 1995. *Journal of Infectious Disease*. 1997; 175:828–32.
18. Weaver SC, Salas R, Rico-Hesse R, Ludwig GV, Oberste MS, Boshell J, et al. Re-emergence of epidemic Venezuelan equine encephalomyelitis in South America. VEE Study Group. *Lancet*. 1996; 348(9025):436–40. Epub 1996/08/17. S0140673696022751 [pii]. PMID: [8709783](#).
19. Brault AC, Powers AM, Holmes EC, Woelk CH, Weaver SC. Positively charged amino acid substitutions in the E2 envelope glycoprotein are associated with the emergence of Venezuelan equine encephalitis virus. *Journal of Virology*. 2002; 76:1718–30. <https://doi.org/10.1128/JVI.76.4.1718-1730.2002> PMID: [11799167](#)
20. Greene IP, Paessler S, Austgen L, Anischenko M, Brault AC, Bowen RA, et al. Envelope glycoprotein mutations mediate equine amplification and virulence of epizootic Venezuelan equine encephalitis virus. *Journal of Virology*. 2005; 79:9128–33. <https://doi.org/10.1128/JVI.79.14.9128-9133.2005> PMID: [15994807](#)
21. Anishchenko M, Bowen RA, Paessler S, Austgen L, Greene IP, Weaver SC. Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103(13):4994–9. <https://doi.org/10.1073/pnas.0509961103> PMID: [16549790](#).
22. Brault AC, Powers AM, Weaver SC. Vector infection determinants of Venezuelan equine encephalitis virus reside within the E2 envelope glycoprotein. *Journal of Virology*. 2002; 76:6387–92. <https://doi.org/10.1128/JVI.76.12.6387-6392.2002> PMID: [12021373](#)
23. Weaver SC, Ferro C, Barrera R, Boshell J, Navarro JC. Venezuelan equine encephalitis. *Ann Rev Entomology*. 2004; 49:141–74.
24. Aguilar PV, Adams AP, Suarez V, Beingolea L, Vargas J, Mancock S, et al. Genetic characterization of Venezuelan equine encephalitis virus from Bolivia, Ecuador and Peru: identification of a new subtype ID lineage. *PLoS Neglected Tropical Diseases*. 2009; 15:e514.
25. Kinney RM, Pfeffer M, Tsuchiya KR, Chang GJ, Roehrig JT. Nucleotide sequences of the 26S mRNA's of the viruses defining the Venezuelan equine encephalitis antigenic complex. *American Journal of Tropical Medicine and Hygiene*. 1998; 59:952–64. PMID: [9886206](#)
26. Quiroz E, Aguilar PV, Cisneros J, Tesh RB, Weaver SC. Venezuelan equine encephalitis in Panama: fatal endemic disease and genetic diversity of etiologic viral strains. *PLoS Neglected Tropical Diseases*. 2009; 3:e472. <https://doi.org/10.1371/journal.pntd.0000472> PMID: [19564908](#)
27. Weaver SC, Pfeffer M, Marriott K, Kang W, Kinney RM. Genetic evidence for the origins of Venezuelan equine encephalitis virus subtype IAB outbreaks. *American Journal of Tropical Medicine and Hygiene*. 1999; 60:441–8. PMID: [10466974](#)
28. Anischenko M, Bowen RA, Paessler S, Austgen L, Greene IP, Weaver SC. Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. *Proceedings of the National Academy of Sciences*. 2006; 103:4994–9.
29. Djikeng A, Halpin R, Kuzmickas R, DePasse J, Feldblyum J, Sengamalay N, et al. Viral genome sequencing by random priming methods. *BMC Genomics*. 2008; 9:5. <https://doi.org/10.1186/1471-2164-9-5> PMID: [18179705](#)

30. Djikeng A, Spiro D. Advancing full length genome sequencing for human viral pathogens. *Future Virology*. 2009; 4:47–53. <https://doi.org/10.2217/17460794.4.1.47> PMID: 19884976
31. Ghedin E, Laplante J, DePasse J, Wentworth DE, Santos RP, Lepow ML, et al. Deep sequencing reveals mixed infection with 2009 pandemic influenza A (H1N1) virus strains and the emergence of oseltamivir resistance. *J Infect Dis*. 2011; 203(2):168–74. Epub 2011/02/04. <https://doi.org/10.1093/infdis/jiq040> PMID: 21288815; PMC3071067.
32. Koren S, Treangen TJ, Hill CM, Pop M, Phillippy A. Automated ensemble assembly and validation of microbial genomes. *BMC Bioinformatics*. 2014; 15:126. <https://doi.org/10.1186/1471-2105-15-126> PMID: 24884846
33. Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*. 2010; 27:221–4. <https://doi.org/10.1093/molbev/msp259> PMID: 19854763
34. Forrester NL, Palacios G, Tesh RB, Savji N, Guzman H, Sherman M, et al. Genome scale phylogeny of the Alphavirus genus suggests a marine origin. *Journal of Virology*. 2012; 85:8709–17.
35. Forrester NL, Palacios G, Tesh RB, Savji N, Guzman H, Sherman M, et al. Genome-scale phylogeny of the alphavirus genus suggests a marine origin. *Journal of virology*. 2012; 86(5):2729–38. Epub 2011/12/23. <https://doi.org/10.1128/JVI.05591-11> PMID: 22190718.
36. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 1999; 41:95–8.
37. Swofford DL. PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sunderland, Massachusetts: Sinauer Associates; 1998.
38. Posada DC, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 1998; 14(9):817–8. PMID: 9918953
39. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*. 2001; 17:754–5. PMID: 11524383
40. Ronquist F, Huelsenbeck JP. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003; 19:1572–4. PMID: 12912839
41. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*. 2007; 7:214. <https://doi.org/10.1186/1471-2148-7-214> PMID: 17996036
42. Drummond AJ, Suchard MA, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*. 2012; 29:1969–73. <https://doi.org/10.1093/molbev/mss075> PMID: 22367748
43. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*. 2005; 22:1185–92. <https://doi.org/10.1093/molbev/msi103> PMID: 15703244
44. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biology*. 2006; 4:e88. <https://doi.org/10.1371/journal.pbio.0040088> PMID: 16683862
45. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*. 2012; 29:2157–67. <https://doi.org/10.1093/molbev/mss084> PMID: 22403239
46. Kosakovsky Pond SL, Frost SDW. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*. 2005; 21:2531–3. <https://doi.org/10.1093/bioinformatics/bti320> PMID: 15713735
47. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*. 2012; 8:e1002764. <https://doi.org/10.1371/journal.pgen.1002764> PMID: 22807683
48. Wertheim JO, Chu DK, Peiris JS, Kosakovsky Pond SL, Poon LL. A case for the ancient origin of coronaviruses. *Journal of Virology*. 2013; 87:7039–45. <https://doi.org/10.1128/JVI.03273-12> PMID: 23596293
49. Wertheim JO, Kosakovsky Pond SL. Purifying selection can obscure the ancient age of viral lineages. *Molecular Biology and Evolution*. 2011; 28:3355–65. <https://doi.org/10.1093/molbev/msr170> PMID: 21705379
50. Wertheim JO, Smith MD, Smith DM, Scheffler K, Kosakovsky Pond SL. Evolutionary origins of human herpes simplex viruses 1 and 2. *Molecular Biology and Evolution*. 2014; 31:2356–64. <https://doi.org/10.1093/molbev/msu185> PMID: 24916030
51. Kosakovsky Pond SL, Murrell B, Fourment M, Frost SD, Delpont W, Scheffler K. A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution*. 2011; 28:3033–43. <https://doi.org/10.1093/molbev/msr125> PMID: 21670087

52. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*. 2015; 32:1365–71.
53. Martin DP, Lemey P, Lott M, Moulton V, Posada DC, Lefevre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*. 2010; 26:2462–3. <https://doi.org/10.1093/bioinformatics/btq467> PMID: 20798170
54. Weaver SC, Winegar R, Manger ID, Forrester NL. Alphaviruses: Population genetics and determinants of emergence. *Antiviral Research*. 2012; 94:242–57. <https://doi.org/10.1016/j.antiviral.2012.04.002> PMID: 22522323
55. Weaver SC, Barrett AD. Transmission cycles, host range, evolution and emergence of arboviral disease. *Nature Reviews Microbiology*. 2004; 2:789–801. <https://doi.org/10.1038/nrmicro1006> PMID: 15378043
56. Walton TE, Alvarez O Jr, Buckwalter RM, Johnson KM. Experimental infections of horses with an attenuated Venezuelan equine encephalomyelitis vaccine (strain TC-83). *Infection and Immunity*. 1972; 5:570–6.
57. Brault AC, Powers AM, Ortiz D, Estrada-Franco JG, Navarro-Lopez R, Weaver SC. Venezuelan equine encephalitis emergence: Enhanced vector infection from a single amino acid substitution in the envelope glycoprotein. *Proc Natl Acad Sci USA*. 2004; 101(31):11344–9. <https://doi.org/10.1073/pnas.0402905101> PMID: 15277679.
58. Greene IP, Paessler S, Anischenko M, Smith DR, Brault AC, Frolov I, et al. Venezuelan equine encephalitis virus in the guinea pig model: evidence for epizootic determinants outside the E2 envelope glycoprotein gene. *American Journal of Tropical Medicine and Hygiene*. 2005; 72:330–8. PMID: 15772331
59. Brault AC, Powers AM, Ortiz D, Estrada-Franco JG, Navarro-Lopez R, Weaver SC. Venezuelan equine encephalitis emergence: Enhanced vector infection from a single amino acid substitution in the envelope glycoprotein. *Proceedings of the National Academy of Sciences*. 2004; 101:11344–9.
60. Tsetsarkin KA, Chen R, Leal G, Forrester NL, Higgs S, Huang J, et al. Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. *Proceedings of the National Academy of Sciences*. 2011; 108:7872–7.
61. Tsetsarkin KA, Weaver SC. Sequential adaptive mutations enhance efficient vector switching by Chikungunya virus and its epidemic emergence. *PLoS Pathogens*. 2011; 7:e1002412. <https://doi.org/10.1371/journal.ppat.1002412> PMID: 22174678
62. Adams AP, Navarro-Lopez R, Ramirez-Aguilar FJ, Lopez-Gonzalez I, Leal G, Flores-Mayorga JM, et al. Venezuelan equine encephalitis virus activity in the Gulf Coast region of Mexico, 2003–2010. *PLoS Neglected Tropical Diseases*. 2012; 6:e1875. <https://doi.org/10.1371/journal.pntd.0001875> PMID: 23133685
63. Ferro C, Boshell J, Moncayo AC, Gonzalez M, Ahumada ML, Kang W, et al. Natural enzootic vectors of Venezuelan equine encephalitis virus, Magdalena Valley, Columbia. *Emerging Infectious Diseases*. 2003; 9:49–54. <https://doi.org/10.3201/eid0901.020136> PMID: 12533281
64. Scherer WF, Weaver SC, Taylor CA, Cupp EW, Dickerman RW, Rubino HH. Vector competence of *Culex (Melanoconion) taeniopus* for allopatric and epizootic Venezuelan equine encephalomyelitis viruses. *American Journal of Tropical Medicine and Hygiene*. 1987; 36:194–7. PMID: 3812882
65. Scherer WF, Weaver SC, Taylor CA, Cupp EW. Vector incompetency: its implication in the disappearance of epizootic Venezuelan equine encephalomyelitis virus from Middle America. *Journal of Medical Entomology*. 1986; 23:23–9. PMID: 3950927
66. Sirivanakarn S, Belkin JN. The identity of *Culex (Melanoconion) taeniopus* and related species with notes on synonymy and description of a new species (Diptera: Culicidae). *Mosquito Systematics*. 1980; 12:7–24.
67. Mendez W, Liria J, Navarro JC, Garcia CZ, Freier JE, Salas R, et al. Spatial dispersion of adult mosquitoes (Diptera: Culicidae) in a sylvatic focus of Venezuelan equine encephalitis virus. *Journal of Medical Entomology*. 2001; 38:813–21. PMID: 11761379
68. Navia-Gine WG, Loaiza JR, Miller MJ. Mosquito-host interactions during and after an outbreak of equine viral encephalitis in Eastern Panama. *PLoS One*. 2013; 8:e81788. <https://doi.org/10.1371/journal.pone.0081788> PMID: 24339965
69. Pecor JE, Jones T, Turell MJ, Fernandez R, Carbajal F, O'Guinn ML, et al. Annotated checklist of hte mosquito species encountered during arboviral studies in Iquitos, Peru (Diptera: Culicidae). *Journal of the American Mosquito Control Association*. 2000; 16:210–8. PMID: 11081648
70. Sallum MA, Forattini OP. Revision of the Spissipes section of *Culex (Melanoconion)* (Diptera: Culicidae). *Journal of the American Mosquito Control Association*. 1996; 12.

71. Turell MJ, O'Guinn ML, Navarro R, Romero G, Estrada-Franco JG. Vector competence of Mexican and Honduran Mosquitoes (Diptera:Culicidae) for enzootic (IE) and Epizootic (IC) strains of Venezuelan equine encephalomyelitis virus. *Journal of Medical Entomology*. 2003; 40:306–10. PMID: [12943109](#)
72. Barrera R, Ferro C, Navarro JC, Freier JE, Liria J, Salas R, et al. Contrasting sylvatic foci of Venezuelan equine encephalitis virus in northern South America. *American Journal of Tropical Medicine and Hygiene*. 2002; 67:324–34. PMID: [12408676](#)
73. Navarro JC, Weaver SC. Molecular phylogeny of the Vomerifer and Pedroi groups in the Spissipes section of the subgenus *Culex* (Melanoconion). *Journal of Medical Entomology*. 2004; 41:575–81. PMID: [15311446](#)
74. Sirivanakarn S, Galindo P. *Culex* (Melanoconion) *adamesi*, a new species from Panama. *Mosquito Systematics*. 1980; 12:25–34.
75. Aitken T, Galindo P. On the identity of *Culex* (Melanoconion) *portesi senevet & abonnenc* 1941. *Proceedings of the Entomological Society of Washington*. 1966; 68:198–208.
76. Scherer WF, Weaver SC, Taylor CA, Cupp EW, Dickerman RW, Rubino HH. Vector competence of *Culex* (Melanoconion) *taeniopus* for allopatric and epizootic Venezuelan equine encephalomyelitis viruses. *The American journal of tropical medicine and hygiene*. 1987; 36(1):194–7. Epub 1987/01/01. PMID: [3812882](#).