

RESEARCH ARTICLE

# Does adaptation to vertebrate codon usage relate to flavivirus emergence potential?

Nicholas Di Paola<sup>1</sup>, Caio César de Melo Freire<sup>2</sup>, Paolo Marinho de Andrade Zanotto<sup>1\*</sup>

**1** Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute, University of Sao Paulo, Sao Paulo, Brazil, **2** Department of Genetics and Evolution, UFSCar—Federal University of Sao Carlos, Sao Carlos, Brazil

\* [pzanotto@usp.br](mailto:pzanotto@usp.br)



**OPEN ACCESS**

**Citation:** Di Paola N, Freire CCdM, Zanotto PMdA (2018) Does adaptation to vertebrate codon usage relate to flavivirus emergence potential? PLoS ONE 13(1): e0191652. <https://doi.org/10.1371/journal.pone.0191652>

**Editor:** Robyn S. Klein, Washington University, UNITED STATES

**Received:** July 25, 2017

**Accepted:** January 9, 2018

**Published:** January 31, 2018

**Copyright:** © 2018 Di Paola et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. Codon Usage Tables are made available for download at Github (<https://github.com/CaioFreire/CUB>).

**Funding:** P.M.A.Z received funding from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, <http://cnpq.br/>), grant number 441105/2016-5. N.D. received funding from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, <http://fapesp.br/>), grant number 2013/22136-1. The funders had no role in study

## Abstract

Codon adaptation index (CAI) is a measure of synonymous codon usage biases given a usage reference. Through mutation, selection, and drift, viruses can optimize their replication efficiency and produce more offspring, which could increase the chance of secondary transmission. To evaluate how higher CAI towards the host has been associated with higher viral titers, we explored temporal trends of several historic and extensively sequenced zoonotic flaviviruses and relationships within the genus itself. To showcase evolutionary and epidemiological relationships associated with silent, adaptive synonymous changes of viruses, we used codon usage tables from human housekeeping and antiviral immune genes, as well as tables from arthropod vectors and vertebrate species involved in the flavivirus maintenance cycle. We argue that temporal trends of CAI changes could lead to a better understanding of zoonotic emergences, evolutionary dynamics, and host adaptation. CAI appears to help illustrate historically relevant trends of well-characterized viruses, in different viral species and genetic diversity within a single species. CAI can be a useful tool together with *in vivo* and *in vitro* kinetics, phylodynamics, and additional functional genomics studies to better understand species trafficking and viral emergence in a new host.

## Introduction

When a virus first jumps into a human host from a non-human vector or reservoir host, conserved changes in viral population sequences can occur that could allow the virus to adapt to a new host [1,2]. Part of host adaptation relies on coping with a host's cellular environment, in particular, the translational machinery that will allow a virus to complete its replication cycle. Codon-bias analyses have proven to be effective in extrapolating information from a zoonotic virus infection in a new, vector or reservoir host [3–8]. During viral replication, the coding regions of viral genomes undergo both mutational and selective pressures that result in non-synonymous and synonymous changes that may or may not result in adaptive changes. Codon adaptation index (CAI) quantifies how well the codon preferences of the virus match that of the host [4,9]. For example, a host cell that uses more frequently AGA to code for Arginine will generally have a greater pool of tRNAs for AGA in its cellular environment, as opposed to a rarer, less used codon for Arginine such as CGU. Although cellular conditions are “tailored”

design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

for the host's translational needs, a virus could optimize its replicative and translational kinetics by selecting for codons more abundant in the host's cellular environment. Although it has been shown that the translational efficiency of some persistently infecting viruses is reduced to diminish protein expression to avoid immune surveillance [7], the opposite has been observed with acutely infecting viruses.

The availability of codon usage tables for a host's constitutively expressed or highly expressed genes could allow analysis of adaptive changes of acutely infecting zoonotic viruses [2], as they interact with new hosts. Convergence or divergence of viral codon adaptation towards or away from that of host can be observed over relatively short periods of time (*i.e.*, years), allowing side by side comparisons with their epidemic history. There is plausibility for increased CAI and the occurrence of human disease from acutely infecting zoonotic viruses. It can be argued that a virus, whose codon usage is calibrated to the host's usage, may have the capability to translate and replicate faster than a virus that has a lower codon adaptation to a specific host [10,11]. Because higher replication tends to increase viral titers, higher host viral load could allow for an increased chance of secondary transmission, virulence and higher chance of mortality from the infection [12].

A requirement for flaviviruses, and arboviruses in general, is the initial need to adapt to the codon choices of a vector species before the virus adapts to the codon usage choices of humans [13]. The alternation between mammals and arthropods impose constraints on arbovirus evolution. Specifically in the vector, albeit strong purifying selection, we can observe silent selection, that is changes in 4-fold degenerate codon sites, promoting differential codon usages in viral genomes [14–16].

Yellow Fever virus (YFV), Dengue virus (DENV) and West Nile virus (WNV) are globally prevalent flaviviruses that cause morbidity and mortality in humans. These three flaviviruses have extensive genetic and biological diversity within themselves, *i.e.* sylvatic and endemic strains of DENV and YFV, or the distinct genetic lineages of WNV [17,18], which vary in reservoir hosts and ability to infect humans in distinct biological behaviors [19–29]. To provide insights into the usage of temporally scaled CAI of YFV, DENV, and YFV in zoonotic emergences, we have performed a large analysis investigating different flavivirus vector-host relationships. We first created two codon usage tables for specific human genes that could play a role in viral codon adaptation. Next, we looked for broad codon adaptation patterns within the flavivirus genus subgroups. We used an in-depth approach to investigate the codon adaptation of YFV, Dengue 2 virus (DENV-2), and WNV to known hosts by using a large set of serially stamped sequences from widespread geographical origins. Our understanding of how viral species modulate translation and adapt to host cellular environments over time is limited. From these data, we show that flaviviruses have distinct patterns of codon adaptation linked vector-host subgroups and biology, and we provide evidence that species-specific codon adaptation is observable and possibly linked to the vector-host evolution across time in certain flaviviruses.

## Materials and methods

### Sequence databases

For each virus (Yellow Fever virus, Dengue 2 virus, Tobacco Mosaic virus, West Nile virus), all available complete genome sequences with collection date and country of isolation information were downloaded from Genbank. Open reading frames of sequences were exported aligned using MAFFT 7.222 [30] and then manually curated using the Geneious 9.0.5 software ([www.geneious.com](http://www.geneious.com)). See [S1 Dataset](#) for complete sequence information (Accession number, virus, collection date, country and lineage).

## Codon adaptation indices of viral coding genes

To investigate if our selected viruses have evidence for codon adaptation in humans, we calculated the codon adaptation indices (CAI) for each coding region of each respective virus. For Yellow Fever virus, Dengue 2 virus, and West Nile virus, the complete polyprotein was used to calculate CAI. For Tobacco Mosaic virus, coding regions were concatenated into frame and then used to generate CAI values.

To calculate normalized CAI, we first used the CAIcal program [31] to obtain a “raw” CAI (rCAI). Next, an “expected neutral CAI” (eCAI) value was calculated by generating 1000 random sequences using similar length, codon composition and GC-content. Normalized CAI values were then compared among different time points and viral lineages using a non-parametric rank test because central tendencies trend varied throughout time more than each time point variance. To obtain our normalized CAI threshold, rCAI/eCAI values were calculated. Values greater than 1 were taken as evidence for codon adaptation to the reference set of codon preferences [31]. Values lower than 1 was taken as evidence that mutational bias are driving codon selection.

## Codon usage tables

The codon usage tables for *Culex pipiens*, *Aedes aegypti*, *Macaca mulata*, *Columba livia* and *Rhipicephalus microplus* were downloaded directly from the publically available Codon Usage Database ([www.kazusa.or.jp/codon](http://www.kazusa.or.jp/codon)).

We used 3 in-house codon usage tables that aimed to represent human house-keeping genes, highly expressed human antiviral immune genes, and *Nicotiana tabacum* (tobacco) house-keeping genes. The human codon usage table available on the Kazusa Codon Usage Database was last updated in 2007. Since highly or constitutively expressed genes will manifest greater codon preference biases [1,2,9], we first identified which genes would be potential “targets” for viral codon adaptation. The 3804 identified human house-keeping genes by Eisenberg and Levanon in 2013 [32] were selected for the human house-keeping gene codon usage table. To identify and select highly expressed human antiviral genes, we looked for studies that quantified innate immunity gene expression during infection, specifically differences between mice and human immune responses [33,34], various protein functionalities of the innate immune system during viral infection [35–38], and immune responses during flavivirus infection [39,40]. Ultimately, we selected 25 genes that we believed to represent highly expressed antiviral immune genes. To address the concern that we may be incorrectly comparing a population’s (virus) versus an individual’s genetic components, we cross-referenced all human coding sequences used with haplotype data from the 1000 Genomes Project to look for single-nucleotide polymorphisms in coding regions of interest [41]. Variants in house-keeping and immune genes did not alter the overall codon frequencies of our generated codon usage tables. All sequences used to generate the human immune table and tobacco house-keeping gene table can be found in [S2 Dataset](#). Comparisons between human tables and between vector tables can be found in [S1 Text](#). Codon usage tables are made available at Github (<https://github.com/CaioFreire/CUB>).

## WNV phylogenetic inference and Malthusian fitness estimate

First, 790 West Nile Virus complete polyprotein sequences from Genbank were downloaded. Sequences were aligned and curated by coding sequences using Muscle and Geneious respectively. Using FastTree.v2 [42], a maximum-likelihood tree was inferred, and allowed us to partition sequences by previously described lineages [20,43]. Using 61 lineage 2 sequences, the rate of each substitution type under the general reversible model (GTR) substitution model,

the proportion of invariant sites ( $I$ ), and shape parameter of a gamma distribution with four rate categories ( $\Gamma_4$ ) was estimated from the data. Tip times corresponding to the year of virus sampling, a lognormal uncorrelated relaxed clock using continuous quantile parameterization [44], and a GMRF Bayesian Skyride coalescence model [45] was used and run using Beast 1.8.3 [46]. The MCMC analysis was run for 100 million chains to ensure convergence (achieved by a single run, with 10% of the run, discarded as representing “burn in”). A MCC was produced and annotated by the use of TreeAnnotator in the BEAST package. To infer the recent demographic history of WNV in Europe, we employed the Bayesian skyride method [45] and previously described methods to estimate the temporal dynamics of effective population size ( $N_e$ ) of WNV, which approximates the number of infections in time. To reveal the dynamics of viral population size growth, we calculated the Malthusian fitness ( $W_m$ ) for the polyprotein, which was approximated by the ratio of the population size in sequential time points ( $W_m = N_{e,t}/N_{e,t-1}$ ) [47].

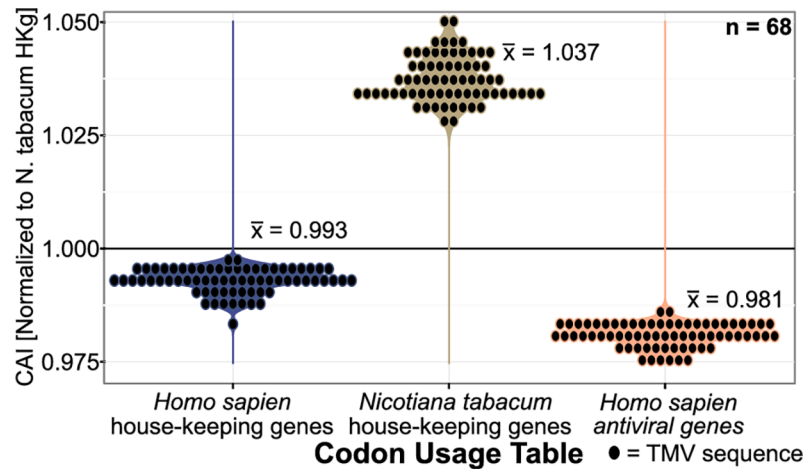
## Results

House-keeping genes are constitutively expressed in all human cells [32,48]; being therefore a likely venue for viral codon adaptation. Moreover, it is fair to assume that human immune genes shape many viral-host interactions and disease severity [33,38]. Usually during a viral acute infection, immune responses are triggered including an increased expression of antiviral genes [35,37,39,49,50]. We did not have any evidence supporting that the host’s tRNA pools alter during viral infection or if the codon preferences of a virus change when cellular immune processes initiate during a recognized viral infection. The immune antiviral genes table was created to provide an alternate human codon table that has different codon usages than that of human housekeeping genes. Between the two human tables, the average difference between the frequencies per thousand bases among triplets is 4.259 (*s.d.* =  $\pm 4.104$ ) with a max difference of 18.1 (All extended codon table comparisons can be found in Supplementary Data).

We first wanted to show that there are measurable and meaningful viral-host adaptations that could be observed using CAI. As a proof of concept for virus-host codon adaptation, we calculated the CAI of Tobacco Mosaic virus (TMV) using complete coding sequences against the codon preferences of tobacco (*Nicotiana tabacum*) housekeeping genes [51,52], TMV’s “natural” host (Fig 1). When TMV sequences were tested against the codon preferences of human genes, there was no evidence for codon adaptation. The average CAI of TMV to tobacco housekeeping genes was 1.037 (*s.d.* =  $\pm 0.005$ ), showing strong evidence for adaptation. On the other hand, TMV’s CAI against human housekeeping and antiviral genes showed no evidence for codon adaptation with an average CAI of 0.993 (*s.d.* =  $\pm 0.003$ ) and 0.981 (*s.d.* =  $\pm 0.003$ ), respectively.

After demonstrating that CAI could be used to measure a virus’ codon adaptation to specific hosts codon choices, we hypothesized that similar relationships could be observed in flaviviruses. They are arthropod-borne viruses that have the ability to infect and replicate in hosts of different phyla. Therefore, versatility in gene expression and protein synthesis is at premium, and certain changes in the viral RNA genome could affect the fitness of the virus in a specific host. Indeed, viral fitness changes relate to dinucleotide frequencies, codon preferences, and codon pair biases [5,14,15,53,54].

Ecology, different virus-host relationships and biogeographical migrations of flavivirus species may explain observed genetic differences among subgroups [13,15,16,55,56]. Phylogenetic studies have acknowledged three groups that reflect the evolutionary and ecological dynamics: tick-borne, mosquito-borne (*Aedes* and *Culex* species), and no known-vector flaviviruses [57–63]. We reconstructed the evolutionary history of the Flavivirus genus based on the complete



**Fig 1. The CAI of Tobacco Mosaic virus to house-keeping genes.** A violin dotplot of the CAI of complete Tobacco Mosaic virus (TMV) coding sequences to the codon usages of human house-keeping genes, antiviral/immune genes, and tobacco (*Nicotiana tabacum*) house-keeping genes. Average CAI values for each group are shown.

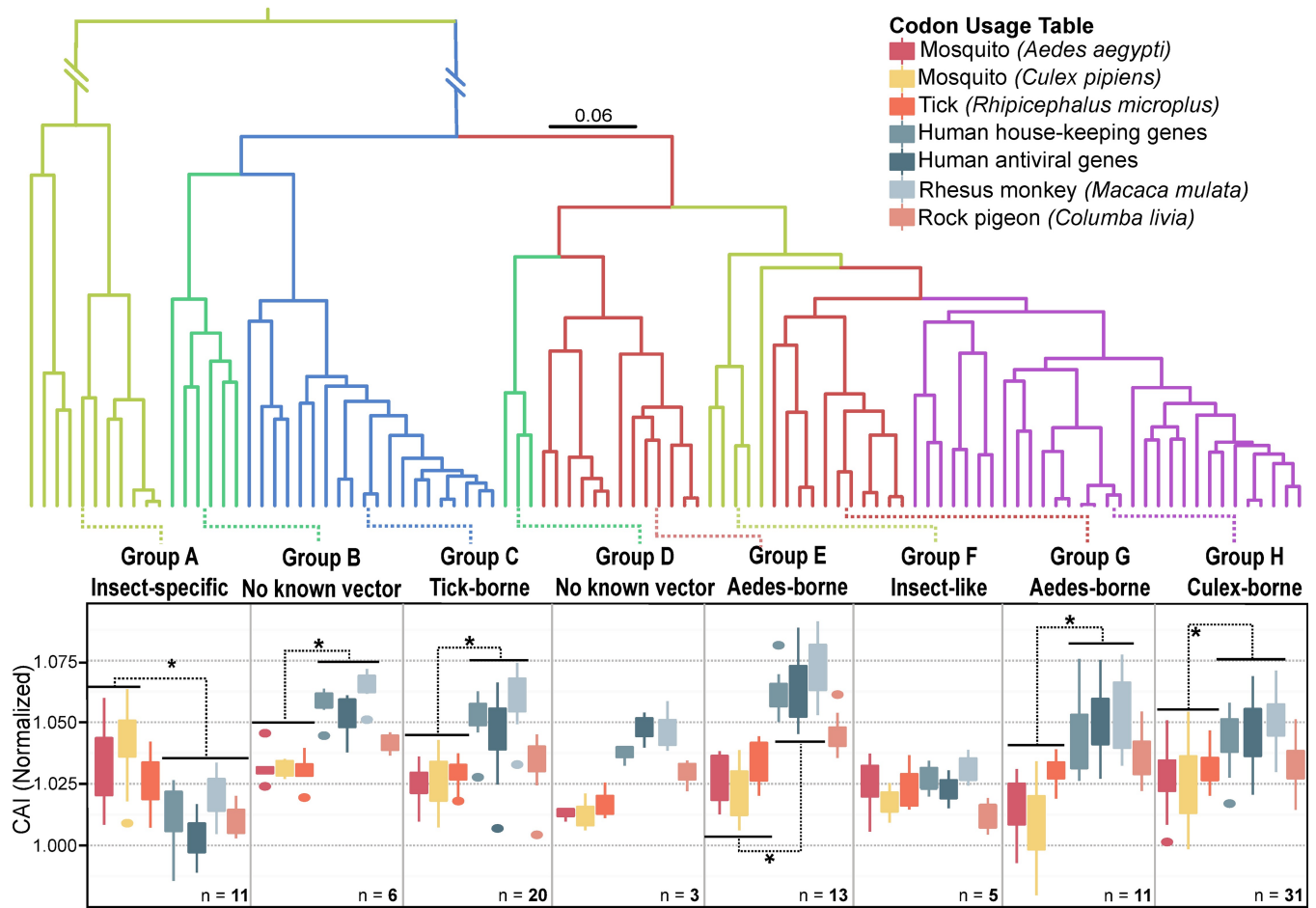
<https://doi.org/10.1371/journal.pone.0191652.g001>

coding sequences of species used in Moureau *et al.*'s 2015 paper [63], except that we used a nucleotide alignment. As biological and genetic differences are evident between each host/vector subgroup, we calculated the normalized CAI for each sequence in their respective clade against 7 codon usage tables: *i*) human housekeeping, *ii*) highly expressed human antiviral immune genes, *iii*) *Aedes aegypti*, *iv*) *Culex pipiens*, *v*) *Rhipicephalus microplus* (tick), *vi*) *Macaca mulata* (rhesus monkey) and *vii*) *Columba livia* (rock dove pigeon) (Fig 2).

Insect-specific groups A includes Cell fusing agent virus (CFAV), Culex flavivirus (CxFV), Aedes flavivirus (AEFV), Kamiti River virus (KMV) and seven others (S1 Fig): all viruses in this group are thought to be vertebrate-incompetent both *in vitro* and *in vivo* [62,64–70]. Both CAI values to human codon and pigeon tables compared to insect vectors were significantly different; with vector CAI values higher (Wilcoxon rank sum test, *p*-values ranged from  $1.985 \times 10^{-5}$  to 0.0457). Although all CAI results for vertebrate CAIs were significantly lower to the *Culex* and *Aedes* tables, we did not find any significant difference between tick and monkey codon adaptation within the group A clade.

The other insect-specific flavivirus group F clade includes 5 species (Chaoyang virus (CHAOV), Donggang virus (DONV), Lammi virus (LAMV), Barkedji virus (BJV) and Ilo-mantsi virus (ILOV) paralogous to mosquito-borne flaviviruses, supporting a more recent emergence and closer relation to *Aedes* and *Culex*-borne viruses. The general equality across all of the CAIs of group F is different than the other insect-specific group (A), where the CAI of vector species is visibly greater than vertebrate species. The group F species showed no clear relationship between vertebrates and vector species codon usages.

Mosquito-borne and tick-borne groups that have known associations with a vertebrate host revealed a different codon bias. In groups C (tick-borne), E (Yellow fever virus complex), G (Dengue virus complex) and H (West Nile virus complex), viral codon adaptation indexes were significantly higher in mammalian vertebrates than in vector species (Wilcoxon rank sum test, *p*-values ranged from  $5.387 \times 10^{-7}$  to 0.0192). Additionally, we found that CAI values were on average higher in monkeys than in human genes for vertebrate-associated flavivirus groups, and in some cases differences were significant. A higher CAI average in vertebrates compared to vectors was universal in all vertebrate-associated groups. Interestingly, the group B clade (no known-vector) shared a similar trend to mosquito-borne and tick-borne CAI



**Fig 2. Flavivirus-CAI dynamics to associated species.** (A) Bayesian-inferred phylogenetic tree of the complete ORF using nucleotide sequences. All nodes had a posterior probability > 0.9. Taxa were omitted for clarity. Associated viral vectors and vertebrate host groups are colored. The scale bar represents 0.06 mutations per site. (B) Box and whisker plot of CAI for each flavivirus species subgroup. For each group of sequences, the CAI was normalized by length, GC-percentage and amino-acid content. Black lines with asterisks signify CAI values that are significantly different.

<https://doi.org/10.1371/journal.pone.0191652.g002>

box plots, where mammalian vertebrate CAIs were significantly higher than vector species (Wilcoxon rank sum test, p-values ranged from 0.002 to 0.008).

The *Culex*-borne flavivirus clade (group H) emerged from *Aedes*-borne flaviviruses thousands of years ago [58,71]. *Culex*-borne species have a life cycle involving primarily avian species and a few species have even been isolated from rodents. For example, the rock dove pigeon (*Columba livia*) is an avian species with prior associations to *Culex*-borne viral infections *i.e.* West Nile virus [72]. Like the group G *Aedes*-borne flaviviruses species, CAI values to mammalian vertebrate tables were significantly higher than in mosquito species (Wilcoxon rank sum test, p-values ranged from  $5.83 \times 10^{-9}$  to  $1.032 \times 10^{-6}$ ). In the case of the rock pigeon, the *Culex*-borne flaviviruses had a lower CAI distribution to pigeons than to mammalian vertebrates (Wilcoxon rank sum test, p-values ranged from  $2.201 \times 10^{-8}$  to 0.000136). Furthermore, *Culex*-borne species had a significantly higher CAI to pigeons than to *Culex pipiens* (Wilcoxon rank sum test, p-value = 0.007) but only averaged higher than *Aedes aegypti* and tick CAI values. Across flavivirus subgroups that are known to harbor viral species with known capability of infecting mammalian vertebrates (B, C, D, E, G, H), the CAI to pigeons rests in the middle of vector and mammalian CAI values.

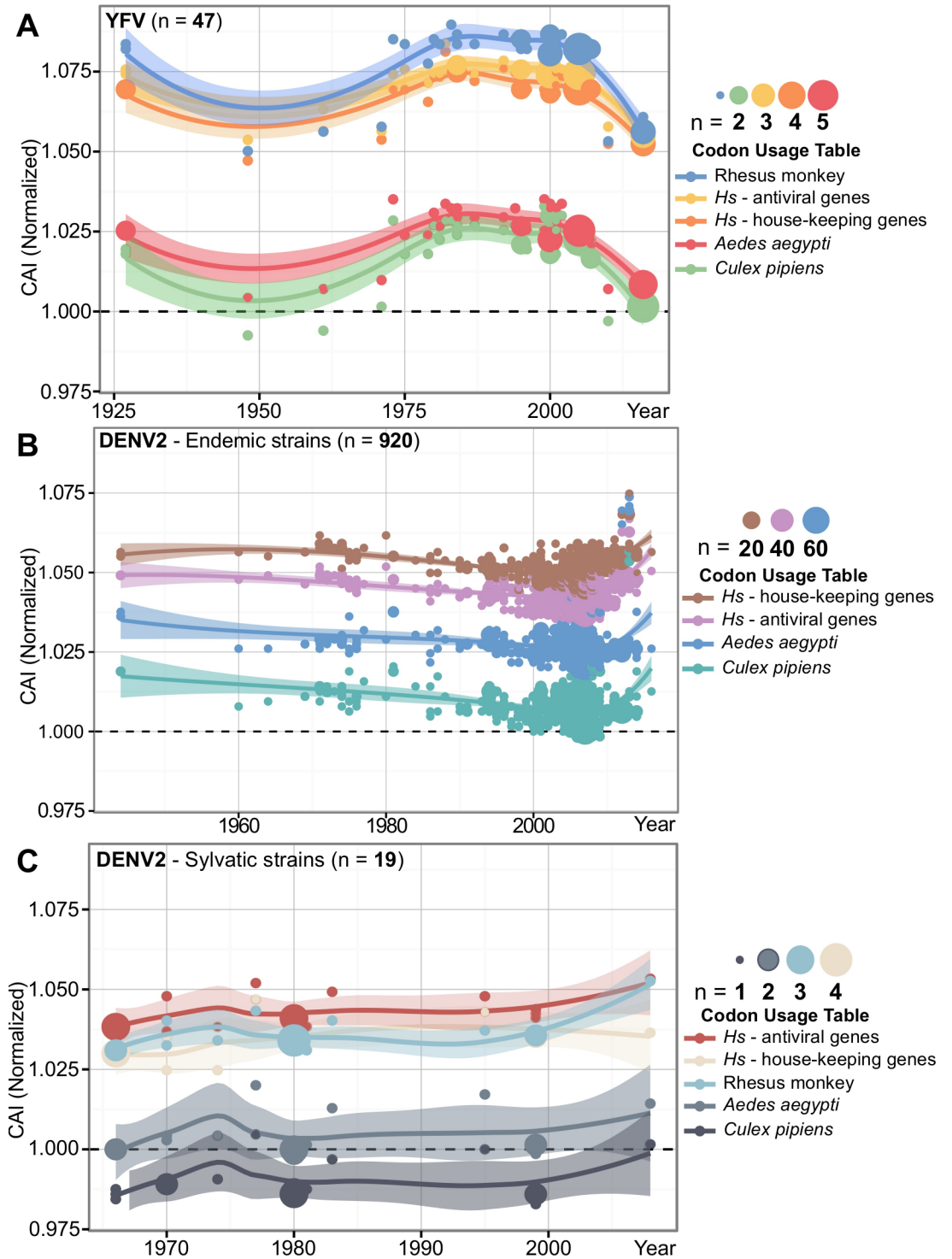
After observing CAI differences between the flavivirus genus subgroups, we explored if changes in CAI over time could be seen in specific viral species. Yellow fever virus (YFV, group E) and Dengue 2 virus (DENV-2, group G) are paralogous *Aedes*-borne flaviviruses who boast century-long interactions with human hosts [27,73,74]. Both had large numbers of serially stamped complete ORF sequences available, allowing us to investigate the temporal trends of YFV CAI, and compare differences in the CAI of DENV-2 endemic and sylvatic strains (Fig 3).

To start, we calculated their CAI to human gene codon usage datasets, the Rhesus monkey (*Macaca mulata*) *Culex pipiens* and the *Aedes aegypti* codon usage tables. For the YFV dataset (Fig 3A), the 5 CAI trendlines were highly correlated (Spearman's rank correlation test,  $p$ -value  $\leq 2.2 \times 10^{-16}$ ,  $\rho$  ranged from 0.977 to 0.999). YFV demonstrated codon adaptation for both vectors and humans across all sampled strains. YFV CAI to Rhesus monkey codon usage was significantly higher than human and vector tables (Wilcoxon rank sum test,  $p$ -values  $\leq 6.375 \times 10^{-7}$ ). We also observed significantly higher CAI for *Aedes aegypti* than in *Culex pipiens* (Wilcoxon rank sum test,  $p$ -value = 0.006). Across all three datasets, YFV had a higher CAI to the mammalian codon usage tables, followed by DENV-2 endemic and sylvatic strains respectively.

A higher CAI to its *Aedes*-associated vector species compared to the common *Culex*-borne vector was also true for endemic and sylvatic strains of DENV-2 (Wilcoxon rank sum test,  $p$ -values  $< 1.711 \times 10^{-5}$ ) (Fig 3B and 3C). DENV-2 endemic strains, showed highly correlated trend lines across all CAI tests (Spearman's rank correlation test,  $p$ -values  $< 2.2 \times 10^{-16}$ ,  $\rho$  values ranged from 0.875 to 0.996). Since we were more interested in the interplay of DENV-2 endemic strains with humans, particularly when transmission alters between mosquitoes and humans in an urban setting, we removed the CAI data of DENV-2 endemic strains to the Rhesus monkey table. DENV-2 endemic strains showed a significantly higher CAI to housekeeping genes over time (Wilcoxon rank sum test,  $p$ -value  $< 2.2 \times 10^{-16}$ ). This is divergent to YFV, where CAI over time was greater to the immune genes codon usage table (Wilcoxon rank sum test,  $p$ -value =  $5.775 \times 10^{-5}$ ).

In DENV-2 sylvatic strains, we saw a different pattern (Fig 3C). DENV-2 sylvatic strains had a higher CAI towards immune genes than they did in house-keeping genes (Wilcoxon rank sum test,  $p$ -value =  $1.919 \times 10^{-5}$ ). Furthermore, sylvatic strains did not show any evidence for codon adaptation towards *Culex pipiens*, as the CAI for all sequences was below 1. Interestingly, codon adaptation to the Rhesus monkey shows an increase in CAI from 1995 to the 2008 sequences, positively correlating with the same increase in CAI that takes place later in DENV-2 endemic strains to human genes from 2003 to 2016 (Spearman's rank correlation test,  $p$ -value  $\leq 2.2 \times 10^{-16}$ ,  $\rho$  ranged from 0.943 to 0.996). In fact, this same positive correlation was shared in DENV-2 sylvatic strains codon adaptation to *Aedes aegypti*. (Spearman's rank correlation test,  $p$ -value  $\leq 2.2 \times 10^{-16}$ ,  $\rho$  ranged from 0.946 to 0.995).

After exploring the CAI trends of YFV and DENV-2, both *Aedes*-borne flaviviruses, we then asked if a similar result could be found in a *Culex*-borne flavivirus such as West Nile virus (WNV). WNV is transmitted primarily by the *Culex pipiens* mosquito and can cause severe neurological disease in vertebrates: including humans, horses and birds [75]. Furthermore, the virus has been isolated in every continent except Antarctica [20]. In addition to its vast geographical spread, the WNV species shows high genetic diversity and exists in at least 7 different lineages [17]. WNV lineage 2 (L2) was historically thought as a milder, less pathogenic lineage than its lineage 1 counterpart and was believed to be restricted to sub-Saharan Africa [76,77]. In 2004, the known boundaries of L2 changed when a Hungarian goshawk (*Accipiter gentilis*) had neurological symptoms that were later identified as WNV disease from a L2 strain [78]. L2 was detected in birds, mosquitoes and humans sporadically over the next few years in Eastern



**Fig 3.** CAI changes across time for (A) Yellow fever virus, (B) endemic strains of Dengue 2 virus, and (C) sylvatic strains of Dengue 2 virus. For each codon usage table, the CAI was normalized by length, GC% and amino acid content for each dataset. Area of plot points reflects the density of sequences at a specific coordinate. A trend line was generated using LOESS, a non-parametric regression method, with 0.95 confidence interval shading. For B), CAI data to monkeys was removed for clarity, but was positioned in between the human table trend lines.

<https://doi.org/10.1371/journal.pone.0191652.g003>



European nations [79]. In 2010, a large WNV L2 human outbreak in Greece was reported, with 81 confirmed cases of West Nile neuroinvasive disease [80]. From 2010 to 2014, WNV human outbreaks were reported in Russia, Czech Republic, Italy, Austria, Serbia, Hungary, Belgium and Romania [81].

With the availability of African and recently sequenced European WNV L2 complete genomes, we investigated the CAI changes of L2 during its recent emergence out of Africa and into Europe. We first show that there were at least two different introductions of L2 into Europe since 2003 (Fig 4A). Interestingly, the epidemic sublineage ( $n = 43$ ) had a higher CAI than the 3 European sequences outside the clade for vertebrate and mosquito tables (Wilcoxon rank sum test,  $p$ -values ranged from 0.001 to 0.02). We then calculated the Malthusian fitness ( $W_m$ ) for L2 from 2004 to 2014, a time period that coincided with L2 strains isolated and sequenced in Europe (Fig 4B).  $W_m$  allows us to observe fitness in a very short period of time, which can show variation in population sizes of a viral species [47].

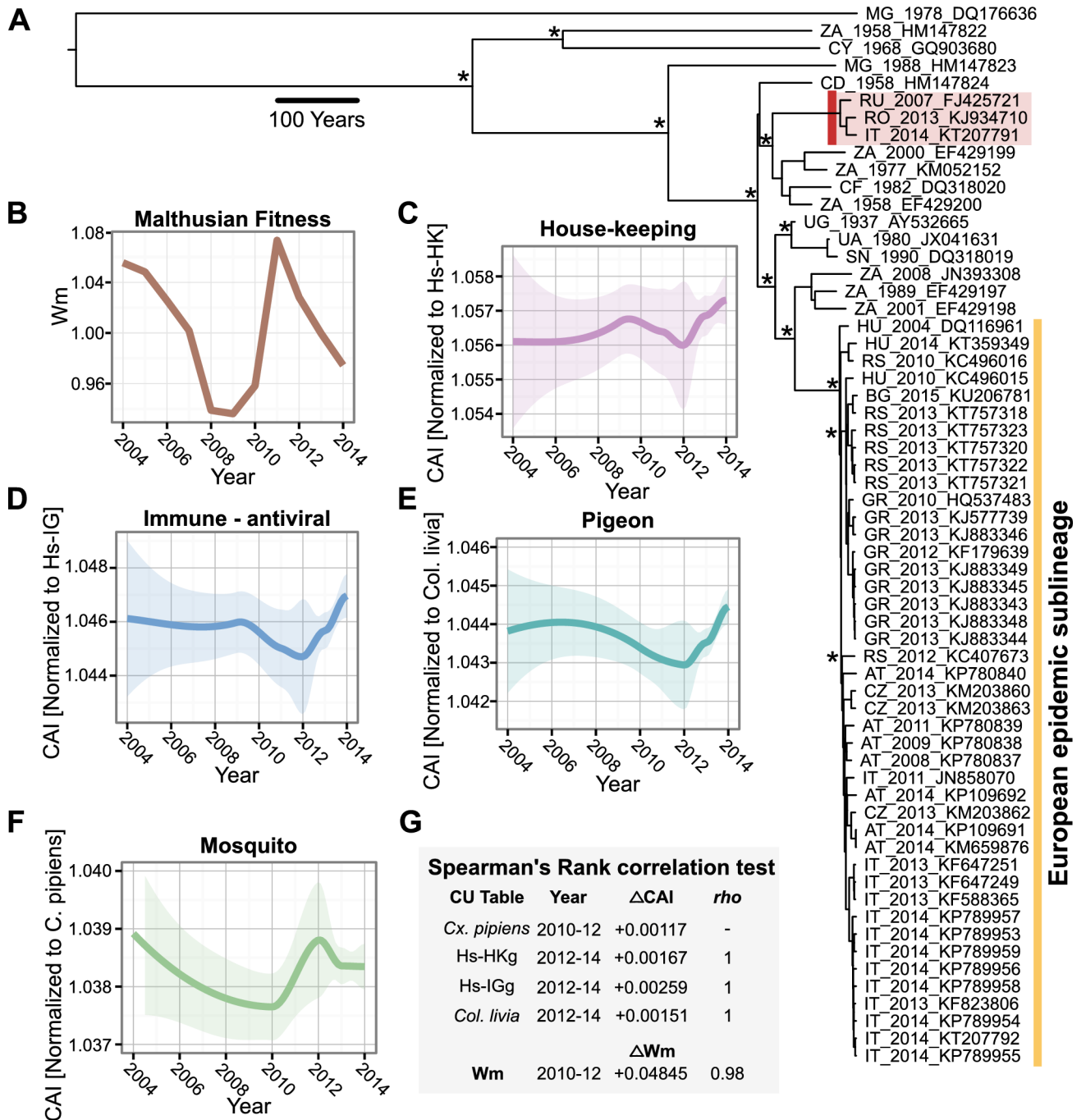
We then calculated the CAI of L2 sequences against human, mosquito (*Culex pipiens*) and the pigeon (*Columba livia*) codon usage tables. With  $W_m$ , a measure of relative fitness over time, we were able to compare two independent measurements of viral fitness that span before and during the European WNV L2 epidemic (Fig 4B–4F). We first looked for any trends of  $W_m$  or CAI from 2004 to 2014 that were similar. We discovered that the trend lines for mosquitoes and  $W_m$  were correlated (Wilcoxon rank sum test,  $p = 0.005$ ,  $\rho = 0.8$ ). Notably, the joint increase of  $W_m$  and mosquito CAI during the 2010 to 2012 period was highly correlated (Wilcoxon rank correlation test,  $p = 0.017$ ,  $\rho = 1$ ). Interestingly, an increase of CAI in pigeon and human tables followed in 2012 to 2014 that was highly correlated to the earlier increase in mosquitoes and  $W_m$  (Spearman's rank correlation test,  $p$ -value  $\leq 0.017$ ,  $\rho \geq 0.98$ ). The increase in vector CAI that precedes the increase in human CAI was similar to what we observed in sylvatic DENV-2 CAI to *Aedes aegypti* and rhesus monkeys, followed by a later increase in the CAI of endemic DENV-2 strains in humans.

## Discussion

Observing changes that may be involved in translational efficiency offers a promising avenue for a better comprehension to how some flaviviruses can infect and replicate in human hosts. Codon usage bias studies are inherently common, but they rarely consider the changes in codon adaptation to hosts that occur over time, as we have done here. Although we observed significant differences in the normalized CAI of different viruses, we still lack the knowledge to infer what biological consequences these small changes could cause. Although the critique of sequencing bias, including the scarcity of aged samples, and lack of non-human sequences can limit CAI analyses, we have shown that the results can be informative.

One advantage of calculating the CAI of flaviviruses is the single open-reading frame that codes for around 3400 amino acids. Using such complete polyprotein reading frames, we were able to observe differences in CAI between previously inferred vector and vertebrate phylogenetic groups [5,63,82,83]. In particular, insect-specific groups favored the codon choices of vector species, while mammal-associated flavivirus groups were better adapted to vertebrate codon usage. Insect-specific flaviviruses are thought to sustain their populations in their respective insect vectors in the absence of mammal reservoirs [84], so lower translational efficiency in vertebrates could be expected.

As expected, some of our results could not be connected to some of the known biological aspects of flaviviruses. This could suggest that there may be uncharacterized features of the translational mechanisms at play. For example, the results we obtained using the tick (*Rhipicephalus microplus*) table did not support the biological and epidemiology of tick-borne viruses



**Fig 4. CAI and phylodynamics of West Nile virus lineage 2 sequences.** A) Bayesian maximum clade credibility tree representing a time scaled phylogeny of a WNV lineage 2 polyprotein sequences. Bayesian posterior probabilities > 0.9 are marked with an asterisk at major nodes. Averages for the European epidemic lineage (yellow bar) and a 2<sup>nd</sup> European lineage (highlighted in red) are shown. B) Malthusian fitness ( $W_m$ ) was calculated from 2004–2014, and compared to LOESS trend lines generated from CAI values to the codon usages of C) human house-keeping genes, D) human immune/antiviral genes, E) pigeon (*Columba livia*) genes, and F) mosquito (*Culex pipiens*) genes. G) The Spearman's rank correlation test was used to test if there were any correlations between the 2010 to 2012 CAI increase in mosquitoes and the CAI increase in vertebrate species from 2012–2014, as well as  $W_m$ . The  $\Delta CAI$ ,  $\Delta W_m$  and  $\rho$  are shown for clarity. # = for all correlations,  $p$ -values were < 0.05.

<https://doi.org/10.1371/journal.pone.0191652.g004>

(Fig 2). Indeed, tick-borne viruses cause dead-end infections in humans [85], so we did not expect CAI to be as high, especially compared to some mosquito-borne viruses that can cause secondary infections in humans [2]. Some tick-borne viruses use rodents as their reservoir hosts, and have associations with other mammalian vertebrates such as dogs, horses, and other large mammals [86], which could have led to tick-borne viruses having similar codon usages to humans and monkeys. Nevertheless, we did not investigate the CAI of flaviviruses to rodents so far.

The number of insect-specific and arboviral sequences as a whole are rapidly increasing [87,88], which will allow for more in-depth analyses in the future. The capabilities of CAI analysis are limited to the availability of serial-stamped sequencing data, specifically older isolates. Furthermore, the public availability of complete genome sequences has increased as cost for sequencing has decreased and improved over time. Occasionally, we also encountered a limitation as to what inferences we could make on the biology of flaviviruses. For example, in Fig 3C, there is only one complete DENV-2 sylvatic genome isolated in 2008 that suggested a CAI increase in rhesus monkeys and *Aedes aegypti* since the turn of the millennia that was later followed by an increase in human CAI. This was also the case for West Nile virus, where L2 showed an increase in Malthusian fitness and *Culex pipiens* prior to an increase in human CAI (Fig 4B–4G). Although there was not enough DENV-2 sylvatic sequences to calculate Malthusian fitness, comparing Malthusian fitness and CAI could help elucidate evolutionary dynamics of an emerging virus [89].

Crucially, just one sequence could be enough to make statistical inferences regarding changes in codon adaptation to a specific host. This is due to the fact that every one of the 3400 codons in the genome is counted, and tested against 500 random sequences with the same length and amino acid content to determine if codon changes are adaptive or random. Therefore, arguably every flavivirus sequence included in our analysis provides substantial amount of data.

During an ongoing epidemic where sufficient sampling and sequencing have been generated, we could use these sequences to observe changes in fitness of a virus that could coincide with fluctuations of CAI to different hosts. This was the case for WNV L2, for which we had access to 46 complete polyprotein sequences sampled in Europe since 2008. Our phylogenetic inference supported results from previous works that suggest multiple entries of WNV into Europe [22,90–92]. Interestingly, we found that a small group of European sequences (highlighted in red in Fig 4) had significantly lower CAI values in mammals than the epidemic sublineage composed of the rest of the European L2 sequences. This could be explained by sampling bias or possibly that these strains lacked the translational efficiency to spread with the same force as the other European strains (Fig 4A). We lack *in vivo* experimental data, but we can hypothesize how our results can shed light on the complexity of viral translational adaptation over time in different vectors and hosts.

Coupling CAI analysis with *in vivo* and *in vitro* viral titer quantification, and functional genetics could improve the understanding of the phenotypic impact of synonymous changes, such as in the cases of the comparison between the African and Brazilian strains of the Zika virus [10]. In a similar manner, a possible follow-up study would be to test *in vivo/in vitro* kinetics (see [93] for an example) for endemic and sylvatic strains of DENV2 from the same sampling location. There are many gaps in our understanding of how synonymous and non-synonymous changes lead to functional changes in viral replication, virulence and host adaptation [94]. Nevertheless, temporal analyses of CAI can be informative in better identifying time-stamped sequences with non-synonymous changes that could provide an advantage in different hosts. In sum, we hope more studies will identify housekeeping genes for other peridomestic species capable of transmitting zoonotic viruses, such as flavivirus vectors and *Culex*-associated migratory birds.

## Supporting information

**S1 Dataset. Data used to generate figures.** The dataset is divided by the data used for each figure. In each tab, accession numbers and normalized CAI values are included.  
(XLSX)

**S2 Dataset. Codon usage table genes.** Genes used in the creating of in-house codon usage tables, including tobacco house-keeping genes and human immune/antiviral genes.  
(XLSX)

**S1 Fig. Bayesian-inferred phylogenetic tree of the complete ORF using nucleotide sequences.** Taxon are included and all nodes had a posterior probability  $> 0.9$ . Associated viral vectors and vertebrate host groups are colored. The scale bar represents 0.06 mutations per site.  
(EPS)

**S1 Text. Codon usage table comparisons using *codcmp*.** Using the *codcmp* program from the EMBOSS package, raw outputs from relevant comparisons between human, tobacco, and vector species codon usage tables are shown.  
(TXT)

## Acknowledgments

We would like to thank Manfred Weidmann, Amadou A. Sall, Moussa M. Diagne, Martin Faye, Erica Mendes, Kristina Ulrich, Jason Ladner, Michael Wiley and especially Dr. James D. Brien for their helpful comments and critiques in the development and improvement of this work.

## Author Contributions

**Conceptualization:** Nicholas Di Paola, Caio César de Melo Freire.

**Data curation:** Nicholas Di Paola, Caio César de Melo Freire.

**Formal analysis:** Nicholas Di Paola.

**Funding acquisition:** Nicholas Di Paola.

**Investigation:** Nicholas Di Paola.

**Methodology:** Nicholas Di Paola, Caio César de Melo Freire.

**Project administration:** Paolo Marinho de Andrade Zanotto.

**Resources:** Nicholas Di Paola.

**Software:** Nicholas Di Paola, Caio César de Melo Freire.

**Supervision:** Paolo Marinho de Andrade Zanotto.

**Validation:** Nicholas Di Paola, Caio César de Melo Freire.

**Visualization:** Nicholas Di Paola.

**Writing – original draft:** Nicholas Di Paola.

**Writing – review & editing:** Nicholas Di Paola, Caio César de Melo Freire, Paolo Marinho de Andrade Zanotto.

## References

1. Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. The Evolution and Genetics of Virus Host Shifts. *PLoS Pathog.* 2014; 10. <https://doi.org/10.1371/journal.ppat.1004395> PMID: 25375777

2. Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol*. Nature Publishing Group; 2009; 5: 311. <https://doi.org/10.1038/msb.2009.71> PMID: 19888206
3. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*. Nature Publishing Group; 2011; 12: 32–42. <https://doi.org/10.1038/nrg2899> PMID: 21102527
4. Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res*. 1986; 14: 5125–43. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=311530&tool=pmcentrez&rendertype=abstract> PMID: 3526280
5. Jenkins GM, Pagel M, Gould E a, Zanotto PM de A, Holmes EC. Evolution of Base Composition and Codon Usage Bias in the Genus *Flavivirus*. *J Mol Evol*. 2001; 52: 383–390. <https://doi.org/10.1007/s002390010168> PMID: 11343134
6. Butt AM, Nasrullah I, Qamar R, Tong Y. Evolution of codon usage in Zika virus genomes is host and vector specific. *Emerg Microbes Infect*. Nature Publishing Group; 2016; 5: e107. <https://doi.org/10.1038/emi.2016.106> PMID: 27729643
7. Shin YC, Bischof GF, Lauer WA, Desrosiers RC. Importance of codon usage for the temporal regulation of viral gene expression. *Proc Natl Acad Sci*. 2015; 112: 201515387. <https://doi.org/10.1073/pnas.1515387112> PMID: 26504241
8. Urrutia a O, Hurst LD. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics*. 2001; 159: 1191–9. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461876&tool=pmcentrez&rendertype=abstract> PMID: 11729162
9. Sharp PM, Li W-H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. 1987; 15: 1281–1295. <https://doi.org/10.1093/nar/15.3.1281> PMID: 3547335
10. Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JLM, Guimarães KP, et al. The Brazilian Zika virus strain causes birth defects in experimental models. *Nature*. 2016; 534: 267–271. <https://doi.org/10.1038/nature18296> PMID: 27279226
11. Andersen KG, Shapiro BJ, Matranga CB, Sealfon R, Lin AE, Moses LM, et al. Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus. *Cell*. 2015; 162: 738–750. <https://doi.org/10.1016/j.cell.2015.07.020> PMID: 26276630
12. Faye O, Andronico A, Faye O, Salje H, Bo??lle PY, Magassouba N, et al. Use of Viremia to Evaluate the Baseline Case Fatality Ratio of Ebola Virus Disease and Inform Treatment Studies: A Retrospective Cohort Study. *PLoS Med*. 2015; 12: 1–14. <https://doi.org/10.1371/journal.pmed.1001908> PMID: 26625118
13. Vasilakis N, Deardorff ER, Kenney JL, Rossi SL, Hanley KA, Weaver SC. Mosquitoes Put the Brake on Arbovirus Evolution: Experimental Evolution Reveals Slower Mutation Accumulation in Mosquito Than Vertebrate Cells. Andino R, editor. *PLoS Pathog*. 2009; 5: e1000467. <https://doi.org/10.1371/journal.ppat.1000467> PMID: 19503824
14. Behura SK, Severson DW. Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol Rev*. 2013; 88: 49–61. <https://doi.org/10.1111/j.1469-185X.2012.00242.x> PMID: 22889422
15. Behura SK, Severson DW. Biclust pattern of codon context usages between flavivirus and vector mosquito *Aedes aegypti*: relevance to infection and transcriptional response of mosquito genes. *Mol Genet Genomics*. 2014; 289: 885–894. <https://doi.org/10.1007/s00438-014-0857-x> PMID: 24838953
16. Shen SH, Stauff CB, Gorbatshevych O, Song Y, Ward CB, Yurovsky A, et al. Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. *Proc Natl Acad Sci*. 2015; 112: 4749–4754. <https://doi.org/10.1073/pnas.1502864112> PMID: 25825721
17. Pachler K, Lebl K, Berer D, Rudolf I, Hubalek Z, Nowotny N, et al. Putative New West Nile Lineage in *Uranotaenia unguiculata* Mosquitoes, Austria, 2013. *Emerg Infect Dis*. 2014; 20: 2119–2122. <https://doi.org/10.3201/eid2012.140921> PMID: 25418009
18. Fall G, Diallo M, Loucoubar C, Faye O, Sall AA. Vector competence of *Culex neavei* and *Culex quinquefasciatus* (Diptera: Culicidae) from Senegal for lineages 1, 2, Koutango and a putative new lineage of West Nile virus. *Am J Trop Med Hyg*. 2014; 90: 747–754. <https://doi.org/10.4269/ajtmh.13-0405> PMID: 24567319
19. Peterson D, Frank D, Pace N, Gordon J. Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. *Cell Host Microbe*. 2008; 3: 417–427. <https://doi.org/10.1016/j.chom.2008.05.001> Metagenomic PMID: 18541218

20. May FJ, Davis CT, Tesh RB, Barrett ADT. Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. *J Virol.* 2011; 85: 2964–74. <https://doi.org/10.1128/JVI.01963-10> PMID: 21159871
21. Murray KO, Mertens E, Despres P. West Nile virus and its emergence in the United States of America. *Vet Res.* 2010; 41: 67. <https://doi.org/10.1051/vetres/20100039> PMID: 21188801
22. McMullen AR, Albayrak H, May FJ, Davis CT, Beasley DWC, Barrett ADT. Molecular evolution of lineage 2 West Nile virus. *J Gen Virol.* 2013; 94: 318–25. <https://doi.org/10.1099/vir.0.046888-0> PMID: 23136360
23. Villabona-Arenas CJ, Zanotto PM de A. Worldwide spread of Dengue virus type 1. *PLoS One.* 2013; 8: e62649. <https://doi.org/10.1371/journal.pone.0062649> PMID: 23675416
24. Gubler DJ. The changing epidemiology of yellow fever and dengue, 1900 to 2003: Full circle? *Comp Immunol Microbiol Infect Dis.* 2004; 27: 319–330. <https://doi.org/10.1016/j.cimid.2004.03.013> PMID: 15225982
25. Gubler DJ. Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev. Am Soc Microbiol;* 1998; 11: 480–496.
26. Gubler DJ, Gubler DJ, Reiter P, Reiter P, Ebi KL, Ebi KL, et al. Climate variability and change in the United States: potential impacts on vector- and rodent-borne diseases. *Environ Health Perspect.* 2001;109 Suppl: 223–33. <https://doi.org/10.2307/3435012>
27. Bryant JE, Holmes EC, Barrett ADT. Out of Africa: A molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathog.* 2007; 3: 0668–0673. <https://doi.org/10.1371/journal.ppat.0030075> PMID: 17511518
28. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda K a, et al. Global spread of dengue virus types: mapping the 70 year history. *Trends Microbiol. Elsevier Ltd;* 2014; 22: 138–46. <https://doi.org/10.1016/j.tim.2013.12.011> PMID: 24468533
29. Beck A, Guzman H, Li L, Ellis B, Tesh RB, Barrett ADT. Phylogeographic Reconstruction of African Yellow Fever Virus Isolates Indicates Recent Simultaneous Dispersal into East and West Africa. *PLoS Negl Trop Dis.* 2013; 7. <https://doi.org/10.1371/journal.pntd.0001910> PMID: 23516640
30. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability Article Fast Track. *Mol Biol Evol.* 2013; 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
31. Puigbò P, Bravo IG, Garcia-Vallvé S. E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics.* 2008; 9: 65. <https://doi.org/10.1186/1471-2105-9-65> PMID: 18230160
32. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet. Elsevier Ltd;* 2013; 29: 569–574. <https://doi.org/10.1016/j.tig.2013.05.010> PMID: 23810203
33. Shay T, Jojic V, Zuk O, Rothamel K, Puyraimond-Zemmour D, Feng T, et al. Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc Natl Acad Sci.* 2013; 110: 2946–2951. <https://doi.org/10.1073/pnas.1222738110> PMID: 23382184
34. Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* 2009; 19: 1316–1323. <https://doi.org/10.1101/gr.080531.108> PMID: 19498102
35. Taniguchi T, Takaoka A. The interferon- $\alpha/\beta$  system in antiviral responses: A multimodal machinery of gene regulation by the IRF family of transcription factors. *Curr Opin Immunol.* 2002; 14: 111–116. [https://doi.org/10.1016/S0952-7915\(01\)00305-3](https://doi.org/10.1016/S0952-7915(01)00305-3) PMID: 11790540
36. Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol.* 2004; 5: 730–737. <https://doi.org/10.1038/ni1087> PMID: 15208624
37. Yoneyama M, Kikuchi M, Matsumoto K, Imaizumi T, Miyagishi M, Taira K, et al. Shared and Unique Functions of the DExD/H-Box Helicases RIG-I, MDA5, and LGP2 in Antiviral Innate Immunity. *J Immunol.* 2005; 175: 2851–2858. <https://doi.org/10.4049/jimmunol.175.5.2851> PMID: 16116171
38. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. *Cell.* 2006; 124: 783–801. <https://doi.org/10.1016/j.cell.2006.02.015> PMID: 16497588
39. Thio CL. Host Genetic Factors and Antiviral Immune Responses to Hepatitis C Virus. *Clin Liver Dis.* 2008; 12: 713–726. <https://doi.org/10.1016/j.cld.2008.03.002> PMID: 18625436
40. Liu ZY, Li XF, Jiang T, Deng YQ, Ye Q, Zhao H, et al. Viral RNA switch mediates the dynamic control of flavivirus replicase recruitment by genome cyclization. *Elife.* 2016; 5: 1–27. <https://doi.org/10.7554/eLife.17636> PMID: 27692070

41. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature*. 2015; 526: 68–74. <https://doi.org/10.1038/nature15393> PMID: [26432245](https://pubmed.ncbi.nlm.nih.gov/26432245/)
42. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010; 5: e9490. <https://doi.org/10.1371/journal.pone.0009490> PMID: [20224823](https://pubmed.ncbi.nlm.nih.gov/20224823/)
43. Bakonyi T, Hubálek Z, Rudolf I, Nowotny N. Novel Flavivirus or New Lineage of West Nile Virus, Central Europe. *Emerg Infect Dis*. 2005; 11: 225–231. <https://doi.org/10.3201/eid1102.041028> PMID: [15752439](https://pubmed.ncbi.nlm.nih.gov/15752439/)
44. Li WLS, Drummond AJ. Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol*. 2012; 29: 751–761. <https://doi.org/10.1093/molbev/msr232> PMID: [21940644](https://pubmed.ncbi.nlm.nih.gov/21940644/)
45. Minin VN, Bloomquist EW, Suchard MA. Smooth Skyride through a Rough Skyline: Bayesian Coalescent-Based Inference of Population Dynamics. *Mol Biol Evol*. 2000; 25: 1459–1471. <https://doi.org/10.1093/molbev/msn090> PMID: [18408232](https://pubmed.ncbi.nlm.nih.gov/18408232/)
46. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007; 7: 214. <https://doi.org/10.1186/1471-2148-7-214> PMID: [17996036](https://pubmed.ncbi.nlm.nih.gov/17996036/)
47. Day T, Otto SP. Fitness. *Encyclopedia of Life Sciences*. Chichester, UK: John Wiley & Sons, Ltd; 2001. pp. 1–6. <https://doi.org/10.1038/npg.els.0001745>
48. Eisenberg E, Levanon EY. Human housekeeping genes are compact. *TRENDS Genet. Elsevier*; 2003; 19: 362–365. [https://doi.org/10.1016/S0168-9525\(03\)00140-9](https://doi.org/10.1016/S0168-9525(03)00140-9) PMID: [12850439](https://pubmed.ncbi.nlm.nih.gov/12850439/)
49. Saffert RT, Kalejta RF. Inactivating a cellular intrinsic immune defense mediated by Daxx is the mechanism through which the human cytomegalovirus pp71 protein stimulates viral immediate-early gene expression. *J Virol*. 2006; 80: 3863–71. <https://doi.org/10.1128/JVI.80.8.3863-3871.2006> PMID: [16571803](https://pubmed.ncbi.nlm.nih.gov/16571803/)
50. Zanotto PM, Kallas EG, de Souza RF, Holmes EC. Genealogical evidence for positive selection in the nef gene of HIV-1. *Genetics*. 1999; 153: 1077–89. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10545443> PMID: [10545443](https://pubmed.ncbi.nlm.nih.gov/10545443/)
51. Schmidt GW, Delaney SK. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Mol Genet Genomics*. 2010; 283: 233–241. <https://doi.org/10.1007/s00438-010-0511-1> PMID: [20098998](https://pubmed.ncbi.nlm.nih.gov/20098998/)
52. Nicot N. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot*. 2005; 56: 2907–2914. <https://doi.org/10.1093/jxb/eri285> PMID: [16188960](https://pubmed.ncbi.nlm.nih.gov/16188960/)
53. Behura SK, Severson DW. Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS One. Public Library of Science*; 2012; 7: e43111. <https://doi.org/10.1371/journal.pone.0043111> PMID: [22912801](https://pubmed.ncbi.nlm.nih.gov/22912801/)
54. Simmen MW. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. *Genomics. Elsevier*; 2008; 92: 33–40. <https://doi.org/10.1016/j.ygeno.2008.03.009> PMID: [18485662](https://pubmed.ncbi.nlm.nih.gov/18485662/)
55. Coffey LL, Forrester N, Tssetsarkin K, Vasilakis N, Weaver SC. Factors shaping the adaptive landscape for arboviruses: implications for the emergence of disease. *Future Microbiol*. 2013; 155–176. <https://doi.org/10.2217/fmb.12.139> PMID: [23374123](https://pubmed.ncbi.nlm.nih.gov/23374123/)
56. Coffey LL, Vasilakis N, Brault AC, Powers AM, Triplet F, Weaver SC. Arbovirus evolution in vivo is constrained by host alternation. *Proc Natl Acad Sci U S A*. 2008; 105: 6970–6975. <https://doi.org/10.1073/pnas.0712130105> PMID: [18458341](https://pubmed.ncbi.nlm.nih.gov/18458341/)
57. Grard G, Moureau G, Charrel RN, Lemasson JJ, Gonzalez JP, Gallian P, et al. Genetic characterization of tick-borne flaviviruses: New insights into evolution, pathogenetic determinants and taxonomy. *Virology*. 2007; 361: 80–92. <https://doi.org/10.1016/j.virol.2006.09.015> PMID: [17169393](https://pubmed.ncbi.nlm.nih.gov/17169393/)
58. Gaunt MW, Sall AA, Lamballerie X De, Gould E a, Falconar AKI, Dzhivaniyan TI, et al. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J Gen Virol*. 2001; 82: 1867–1876. <https://doi.org/10.1099/0022-1317-82-8-1867> PMID: [11457992](https://pubmed.ncbi.nlm.nih.gov/11457992/)
59. Grard G, Moureau G, Charrel RN, Holmes EC, Gould EA, de Lamballerie X. Genomics and evolution of Aedes-borne flaviviruses. *J Gen Virol*. 2010; 91: 87–94. <https://doi.org/10.1099/vir.0.014506-0> PMID: [19741066](https://pubmed.ncbi.nlm.nih.gov/19741066/)
60. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB. Phylogeny of the genus Flavivirus. *J Virol*. 1998; 72: 73–83. Available: <http://jvi.asm.org/content/72/1/73.short> PMID: [9420202](https://pubmed.ncbi.nlm.nih.gov/9420202/)
61. Cook S, Moureau G, Kitchen A, Gould EA, de Lamballerie X, Holmes EC, et al. Molecular evolution of the insect-specific flaviviruses. *J Gen Virol*. 2012; 93: 223–234. <https://doi.org/10.1099/vir.0.036525-0> PMID: [22012464](https://pubmed.ncbi.nlm.nih.gov/22012464/)

62. Blitvich B, Firth A. Insect-Specific Flaviviruses: A Systematic Review of Their Discovery, Host Range, Mode of Transmission, Superinfection Exclusion Potential and Genomic Organization [Internet]. *Viruses*. 2015. <https://doi.org/10.3390/v7041927> PMID: 25866904
63. Moureau G, Cook S, Lemey P, Nougairede A, Forrester NL, Khasnatinov M, et al. New Insights into Flavivirus Evolution, Taxonomy and Biogeographic History, Extended by Analysis of Canonical and Alternative Coding Sequences. *PLoS One*. 2015; 10: e0117849. <https://doi.org/10.1371/journal.pone.0117849> PMID: 25719412
64. Hoshino K, Takahashi-Nakaguchi A, Isawa H, Sasaki T, Higa Y, Kasai S, et al. Entomological surveillance for flaviviruses at migratory bird stopover sites in Hokkaido, Japan, and a new insect flavivirus detected in *Aedes galloisi* (Diptera: Culicidae). *J Med Entomol*. The Oxford University Press; 2012; 49: 175–182. PMID: 22308786
65. Hoshino K, Isawa H, Tsuda Y, Sawabe K, Kobayashi M. Isolation and characterization of a new insect flavivirus from *Aedes albopictus* and *Aedes flavopictus* mosquitoes in Japan. *Virology*. Elsevier; 2009; 391: 119–129. <https://doi.org/10.1016/j.virol.2009.06.025> PMID: 19580982
66. Bolling BG, Eisen L, Moore CG, Blair CD. Insect-specific flaviviruses from *Culex* mosquitoes in Colorado, with evidence of vertical transmission. *Am J Trop Med Hyg*. ASTMH; 2011; 85: 169–177. <https://doi.org/10.4269/ajtmh.2011.10-0474> PMID: 21734144
67. Vázquez A, Sánchez-Seco M-P, Palacios G, Molero F, Reyes N, Ruiz S, et al. Novel flaviviruses detected in different species of mosquitoes in Spain. *Vector-Borne Zoonotic Dis*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA; 2012; 12: 223–229. <https://doi.org/10.1089/vbz.2011.0687> PMID: 22022811
68. Crabtree MB, Nga PT, Miller BR. Isolation and characterization of a new mosquito flavivirus, Quang Binh virus, from Vietnam. *Arch Virol*. 2009; 154. <https://doi.org/10.1007/s00705-009-0373-1> PMID: 19347244
69. Crabtree MB, Sang RC, Stollar V, Dunster LM, Miller BR. Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus. *Arch Virol*. Springer; 2003; 148: 1095–1118. <https://doi.org/10.1007/s00705-003-0019-7> PMID: 12756617
70. Kim DY, Guzman H, Bueno R, Dennett JA, Auguste AJ, Carrington C V. Characterization of *Culex* Flavivirus (Flaviviridae) strains isolated from mosquitoes in the United States and Trinidad. *Virology*. 2009; 386. <https://doi.org/10.1016/j.virol.2008.12.034> PMID: 19193389
71. Pettersson JHO, Fiz-Palacios O. Dating the origin of the genus *Flavivirus* in the light of Beringian biogeography. *J Gen Virol*. 2014; 95: 1969–1982. <https://doi.org/10.1099/vir.0.065227-0> PMID: 24914065
72. Allison AB, Mead DG, Gibbs SEJ, Hoffman DM, Stalknecht DE. West Nile virus viremia in wild rock pigeons. *Emerg Infect Dis*. Centers for Disease Control and Prevention; 2004; 10: 2252. <https://doi.org/10.3201/eid1012.040511> PMID: 15663876
73. Zanotto PM, Gould E a, Gao GF, Harvey PH, Holmes EC. Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc Natl Acad Sci U S A*. 1996; 93: 548–553. <https://doi.org/10.1073/pnas.93.2.548> PMID: 8570593
74. Sall AA, Faye O, Diallo M, Firth C, Kitchen A, Holmes EC. Yellow fever virus exhibits slower evolutionary dynamics than dengue virus. *J Virol*. 2010; 84: 765–72. <https://doi.org/10.1128/JVI.01738-09> PMID: 19889759
75. Marra PP, Griffing SM, McLean RG. West Nile Virus and Wildlife. *Emerg Infect Dis*. 2003; 9: 7–9.
76. Petersen LR, Roehrig JT. West Nile virus: a reemerging global pathogen. *Rev Biomed*. 2001; 12.
77. Beasley DWC, Li L, Suderman MT, Barrett ADT. Mouse Neuroinvasive Phenotype of West Nile Virus Strains Varies Depending upon Virus Genotype. *Virology*. Elsevier; 2002; 296: 17–23. <https://doi.org/10.1006/viro.2002.1372> PMID: 12036314
78. Bakonyi TT, Ivanics E, Erdleyi K, Ursu K, Ferenczi EE, Weissenbock H, et al. Lineage 1 and 2 strains of encephalitic West Nile virus, Central Europe. *Emerg Infect Dis*. 2006; 12: 618–623. <https://doi.org/10.3201/eid1204.051379> PMID: 16704810
79. Bakonyi T, Ferenczi E, Erdélyi K, Kutasi O, Csörgő T, Seidel B, et al. Explosive spread of a neuroinvasive lineage 2 West Nile virus in Central Europe, 2008/2009. *Vet Microbiol*. 2013; 165: 61–70. <https://doi.org/10.1016/j.vetmic.2013.03.005> PMID: 23570864
80. Papa A, Danis K, Baka A, Bakas A, Douglas G, Lytras T, et al. Ongoing outbreak of West Nile virus infections in humans in Greece, July–August 2010. *Age (Omaha)*. 2010; 20: 20–29.
81. ECDC. West Nile fever maps. In: European Centre for Disease Prevention and Control [Internet]. 2017 [cited 27 Jun 2017]. Available: [http://ecdc.europa.eu/en/healthtopics/west\\_nile\\_fever/West-Nile-fever-maps/Pages/historical-data.aspx](http://ecdc.europa.eu/en/healthtopics/west_nile_fever/West-Nile-fever-maps/Pages/historical-data.aspx)
82. Billoir F, de CR, Tolou H, de MP, Gould EA, de LX. Phylogeny of the genus flavivirus using complete coding sequences of arthropod-borne viruses and viruses with no known vector. *J Gen Virol*. 2000; 81 Pt 9. <https://doi.org/10.1099/0022-1317-81-9-2339> PMID: 10950993



83. Gould EA, Zlotoff PM de A, Holmes EC. The genetic evolution of flaviviruses. 1997.
84. Kuno G. Host range specificity of flaviviruses: correlation with in vitro replication. *J Med Entomol*. 2007; 44. <https://doi.org/10.1093/jmedent/41.5.93>
85. Labuda M, Nuttall P a. Tick-borne viruses. *Parasitology*. 2004;129 Suppl: S221–S245. <https://doi.org/10.1017/S0031182004005220>
86. Dobler G. Zoonotic tick-borne flaviviruses. *Vet Microbiol*. 2010; 140: 221–228. <https://doi.org/10.1016/j.vetmic.2009.08.024> PMID: 19765917
87. Shi M, Lin X-D, Vasilakis N, Tian J-H, Li C-X, Chen L-J, et al. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the *Flaviviridae* and related viruses. *J Virol*. 2015; 90: JVI.02036–15. <https://doi.org/10.1128/JVI.02036-15> PMID: 26491167
88. Cook S, Moureau G, Kitchen A, Gould EA, de Lamballerie X, Holmes EC, et al. Molecular evolution of the insect-specific flaviviruses. *J Gen Virol*. 2012; 93: 223–234. <https://doi.org/10.1099/vir.0.036525-0> PMID: 22012464
89. Freire CC de M, Iamarino A, Neto DF de L, Sall AA, Zlotoff PM de A. Spread of the pandemic Zika virus lineage is associated with NS1 codon usage adaptation in humans [Internet]. 2015 Nov. <https://doi.org/10.1101/032839>
90. Bagnarelli P, Marinelli K, Trotta D, Monchetti A, Tavio M, Gobbo R Del, et al. Human case of autochthonous West Nile virus lineage 2 infection in Italy, September 2011. *Eurosurveillance*. 2011; 16: 1–4.
91. Papa A, Bakonyi TT, Xanthopoulou K, Vázquez A, Tenorio A, Nowotny N, et al. Genetic Characterization of West Nile Virus Lineage 2, Greece, 2010. *Emerg Infect Dis*. 2011; 17: 920–922. <https://doi.org/10.3201/eid1705.101759> PMID: 21529413
92. Sambri V, Capobianchi M, Charrel R, Fyodorova M, Gaibani P, Gould E, et al. West Nile virus in Europe: emergence, epidemiology, diagnosis, treatment, and prevention. *Clin Microbiol Infect*. 2013; 19: 699–704. <https://doi.org/10.1111/1469-0691.12211> PMID: 23594175
93. Fall G, Di Paola N, Faye M, Dia M, Freire CC de M, Loucoubar C, et al. Biological and phylogenetic characteristics of West African lineages of West Nile virus. *PLoS Negl Trop Dis*. Public Library of Science; 2017; 11: e0006078. Available: <https://doi.org/10.1371/journal.pntd.0006078> PMID: 29117195
94. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, et al. Pathways to zoonotic spillover. *Nat Rev Microbiol*. Nature Publishing Group; 2017; 1–29. <https://doi.org/10.1038/nrmicro.2017.45> PMID: 28555073