

<https://helda.helsinki.fi>

---

## An Unsupervised Algorithm for Host Identification in Flaviviruses

Truong Nguyen, Phuoc

Multidisciplinary Digital Publishing Institute

2021-05-14

---

Truong Nguyen, P.; Garcia-Vallvé, S.; Puigbò, P. An Unsupervised Algorithm for Host Identification in Flaviviruses. *Life* 2021, 11, 442.

---

<http://hdl.handle.net/10138/349010>

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

# An Unsupervised Algorithm for Host Identification in Flaviviruses

Phuoc Truong Nguyen <sup>1,2</sup>, Santiago Garcia-Vallvé <sup>3</sup> and Pere Puigbò <sup>1,4,5,\*</sup>

<sup>1</sup> Department of Biology, University of Turku, 20500 Turku, Finland; phuoc.truong@helsinki.fi

<sup>2</sup> Department of Virology, Faculty of Medicine, University of Helsinki, 00290 Helsinki, Finland

<sup>3</sup> Research Group in Cheminformatics & Nutrition, Department of Biochemistry and Biotechnology, Rovira i Virgili University, 43007 Tarragona, Catalonia, Spain; santi.garcia-vallve@urv.cat

<sup>4</sup> Department of Biochemistry and Biotechnology, Rovira i Virgili University, 43007 Tarragona, Catalonia, Spain

<sup>5</sup> Nutrition and Health Unit, Eurecat Technology Centre of Catalonia, 43204 Reus, Catalonia, Spain

\* Correspondence: pepuav@utu.fi

**Abstract:** Early characterization of emerging viruses is essential to control their spread, such as the Zika Virus outbreak in 2014. Among other non-viral factors, host information is essential for the surveillance and control of virus spread. Flaviviruses (genus *Flavivirus*), akin to other viruses, are modulated by high mutation rates and selective forces to adapt their codon usage to that of their hosts. However, a major challenge is the identification of potential hosts for novel viruses. Usually, potential hosts of emerging zoonotic viruses are identified after several confirmed cases. This is inefficient for deterring future outbreaks. In this paper, we introduce an algorithm to identify the host range of a virus from its raw genome sequences. The proposed strategy relies on comparing codon usage frequencies across viruses and hosts, by means of a normalized Codon Adaptation Index (CAI). We have tested our algorithm on 94 flaviviruses and 16 potential hosts. This novel method is able to distinguish between arthropod and vertebrate hosts for several flaviviruses with high values of accuracy (virus group 91.9% and host type 86.1%) and specificity (virus group 94.9% and host type 79.6%), in comparison to empirical observations. Overall, this algorithm may be useful as a complementary tool to current phylogenetic methods in monitoring current and future viral outbreaks by understanding host–virus relationships.



**Citation:** Truong Nguyen, P.; Garcia-Vallvé, S.; Puigbò, P. An Unsupervised Algorithm for Host Identification in Flaviviruses. *Life* **2021**, *11*, 442. <https://doi.org/10.3390/life11050442>

Academic Editors: Kuan Rong Chan, Yaw Shin Ooi, Yie Hou Lee and Eugenia Ong Ziyang

Received: 8 March 2021

Accepted: 12 May 2021

Published: 14 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** flavivirus; host identification; codon adaptation index; algorithm

## 1. Introduction

Recent viral pandemics have shown that rapid characterization of the virus is essential during the development of an outbreak [1–4]. Among other factors, host information is essential for surveillance and control of virus spread. However, emerging viruses are fully characterized only after several confirmed cases occur; this is an inefficient method of deterring current and future outbreaks [5]. Fast and reliable computational biology methods are needed to develop antiviral treatments, to improve medical diagnoses and to efficiently contain viral outbreaks [6]. Viral genomes are modulated by high mutation rates [7] and by selective forces to adapt their codon usage to that of their hosts, especially when the viruses can infect a wide host range, as is the case for flaviviruses [8]. Previous methods identify flavivirus host range based on an analysis of dinucleotides [9,10] based on the idea that a virus that infects multiple hosts has a weaker dinucleotide bias [11].

In this article, we introduce an unsupervised algorithm to identify putative virus host ranges based on only genome sequence information. The proposed methodology has been tested in 94 viruses of genus *Flavivirus* and 16 potential hosts. Several flaviviruses are major human pathogens, with potential host ranges from vertebrates to arthropods [12]. Flaviviruses are classified by vector type into mosquito-borne (MBFV), tick-borne (TBFV), insect-only (IOFV) and unknown vector (UVFV) [13] flaviviruses. In MBFVs, there exists a

paraphyletic subgroup of mosquito-specific viruses [14], also known as dual-host insect-only flaviviruses (dhIOFVs). However, certain annotation ambiguities exist; e.g., the Ecuador Paraíso Escondido Virus (EPEV) is defined as MBFV based on phylogeny, but may also be classified as dhIOFV [15]. Flaviviruses with the same host type tend to be monophyletic and are subject to the same selective pressures as the host; this situation is reflected in their codon usage and dinucleotide composition [16]. The most widespread and prevalent flaviviruses include Dengue virus (DENV), West Nile virus (WNV), Japanese encephalitis virus (JEV), and Zika virus (ZKV) [17].

Several articles suggest that highly similar codon usage frequencies between viruses and hosts are indicative of a high virus–host adaptation level [18]. Thus, the codon adaptation index (CAI) [19] may be a robust indicator for determining putative hosts. Here, we use a normalized CAI (nCAI) and a correspondence analysis (CA) to compare codon usage frequencies across virus and host sequences (see Materials and Methods section). Therefore, the nCAI-CA algorithm provides a fast and reliable method of identifying the putative host range of a virus. This method requires only coding sequences (CDSs) without prior knowledge, and can be implemented with minimal computational equipment. In addition, we have developed an easy-to-use web server, available at <http://ppuigbo.me/programs/CAIcal/nCAI> (accessed on 13 May 2021), to calculate nCAI values.

## 2. Materials and Methods

The optimal host identification algorithm (Figure 1) consists of two phases. In the first phase, the algorithm computes the required codon usage tables through two subroutines: one for the host and the other for the virus. These tables, along with complete genomic CDSs, are then used as the input data for CAIcal [20]. This produces CAI data between the virus and host (CAI<sub>h</sub>) using virus CDSs, and host codon usage tables and CAI data for the virus itself (CAI<sub>s</sub>) using virus CDSs and virus codon usage tables. In the second phase, the CAI<sub>h</sub> values are normalized by dividing each by its respective CAI<sub>s</sub> as in Equation (1):

$$nCAI = \frac{CAI_h}{CAI_s} \quad (1)$$

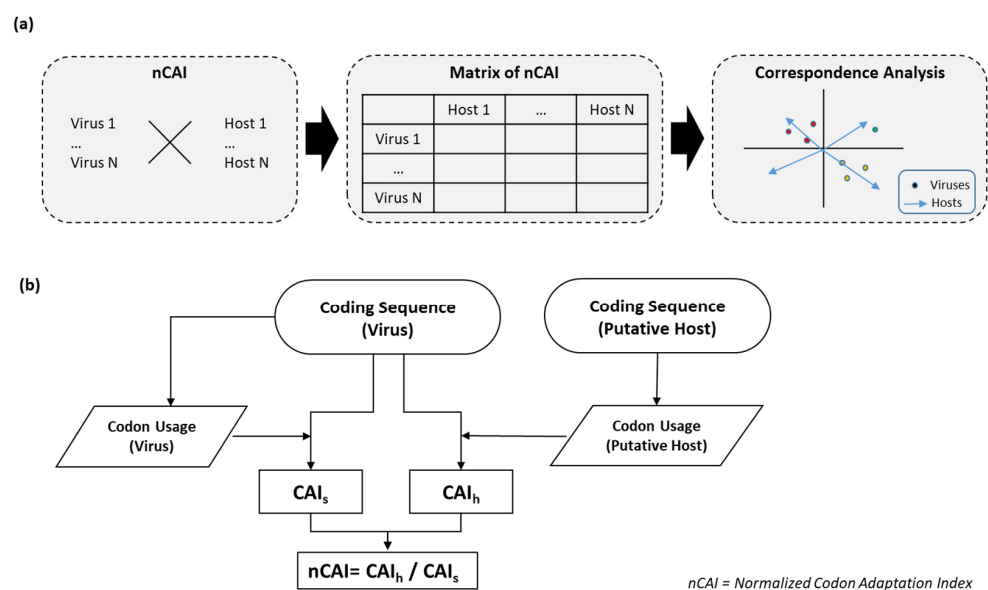
This yields the normalized CAI (nCAI) value, from which the optimal and likely hosts can be inferred depending on how similar the codon usage of a virus is to the codon usage of its host organisms. The nCAI values range between  $-\infty$  and  $+\infty$ , and the optimal value is 1.0, indicating identical codon usage between the virus and host and therefore perfect adaptation to the host. Values above and below 1.0 would indicate over- and underoptimization, respectively, and thus suboptimal adaptation to a host.

The nCAI calculations can be performed with the CAIcal tool in a dedicated web server, written in PHP, that works on any web browser (<http://ppuigbo.me/programs/CAIcal/nCAI>, accessed on 13 May 2021). The server requires two sets of inputs: complete DNA or RNA CDSs of the viruses of interest in FASTA format and the codon usage tables of the potential host animals in the format used by the Codon Usage Database [21]. CAIcal will then output the results in a tab-delimited table with the following values: name of the query sequence (Name); CAI of the virus to a host (CAI<sub>h</sub>); CAI of the virus to itself (CAI<sub>s</sub>); normalized CAI, calculated by dividing CAI<sub>h</sub> by CAI<sub>s</sub> (nCAI); length of the query sequence (Length); overall %GC; and GC content at the first, second or third nucleotide of each codon (%GC1–3).

Available flavivirus CDSs and their respective protein sequences were obtained from the RefSeq [22] and GenBank [23] databases. The viruses ( $n = 94$ ) were chosen according to phylogenetic studies [24–26] and current ICTV classifications [13] (Supplementary Table S1).

In this study, a vector is defined as an organism capable of transmitting a virus to another type of organism. This definition does not take into account whether the virus is virulent within a vector, i.e., there is no differentiation between a vector and a vector-host. A host, on the other hand, is an organism in which the virus primarily replicates, and it does not directly transmit the virus to another organism of the same type. The host

organisms ( $N = 16$ ) for this study were chosen based on information primarily provided by the Virus–Host Database [27], which includes representative arthropod (mosquitoes and tick) and vertebrate (mammals, birds, reptiles and amphibians) host species. Additionally, a more comprehensive list of hosts and vectors for each flavivirus is included in Supplementary Table S1. This table includes only confirmed cases of viruses sequenced from an organism, or cases in which viruses have successfully infected the cells of a host in a laboratory experiment. It is important to note that not all host animals listed in the database are primary hosts, as they might have acquired the viruses through happenstance. We computed a codon usage reference table for 16 putative hosts representing all possible flavivirus host types among vertebrates (mammals, birds, reptiles, and amphibians) and arthropods (mosquitoes and ticks). We analyzed genomes that contained over 10,000 CDSs to reflect actual codon usage frequencies, as well as those of *Gallus gallus* (6017) and *Sus scrofa* (2953).



**Figure 1.** Scheme of the algorithm used to calculate the normalized codon adaptation index (nCAI). (a) Pipeline to identify putative hosts based on nCAI values. The complete coding sequences of hosts and viruses are used to compute nCAI values, which are put into a table. These values are then subjected to correspondence analysis to identify optimal hosts and, thus, the likelihood of a virus infecting an organism. (b) Algorithm to calculate nCAI. The CAI values for possible hosts and viruses of interest are computed from the complete coding sequences (CDSs) and the codon usage tables, which are calculated from the same sequences. The CAI values of the host (CAI<sub>h</sub>) are calculated from virus CDSs and host codon usages, and the CAI values of the viruses (CAI<sub>s</sub>) are computed using virus CDSs and the codon usage values of the viruses themselves. The resulting CAI values are then normalized by dividing each CAI<sub>h</sub> by its respective CAI<sub>s</sub>.

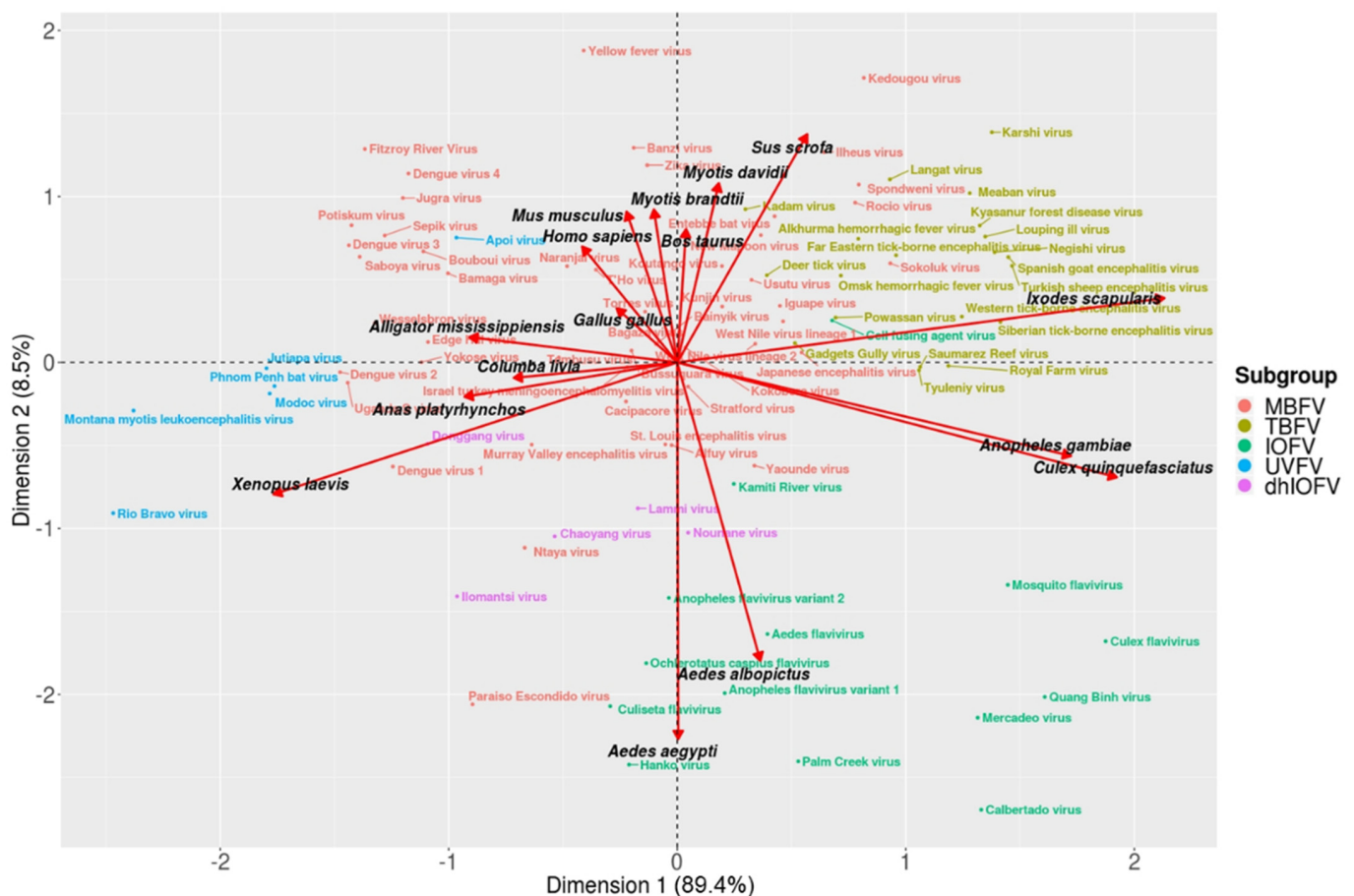
The %GC and relative synonymous codon usage (RSCU) values were calculated from the CDSs of the flaviviruses. The RSCU describes the preference bias for a codon to be used to encode an amino acid [28]. This can be calculated by dividing the observed number of a codon by the expected frequency of the same codon, assuming that individual codons for amino acids were used at equal frequency [29]. CA was performed for two different types of nCAI datasets. The first analysis included all known flaviviruses, and the second included separate datasets, containing only the values for DENV, JEV, WNV and ZKV. The correspondence analyses were performed with the “ca” package (version 0.70) and then plotted with the “ggplot2” package (version 2.2.1) in R (version 3.4.4). The nCAI values of all MBFVs were plotted in a heat map with the “pheatmap” package (version 1.0.10) in R. The virus phylogenetic tree was computed with the following steps: first, the amino

acid sequences of each genome were aligned using MUSCLE [30]. Tree construction was performed with FastTree [31]. Host trees were built based on NCBI taxonomy [32]. Each of the viruses and host organisms were sorted to match their respective phylogenies. For the DENV, JEV, WNV and ZKV genomes, their results were clustered based on k-means (5) in the heat map (Supplementary Figure S9). The clustering of each subgroup was performed and visualized by computing centroids based on the multivariate normal distribution of each subgroup with a confidence level of 0.95. This was achieved with the “ggplot2” package (version 2.2.1) in R (version 3.4.4). The virus subgroups included MBFV, TBFV, IOFV, UVFV and dhIOFV, and the host type subgroups were vertebrates, mosquitoes, and ticks.

### 3. Results

First, to assess whether a codon usage methodology could distinguish subgroups within a viral species, we performed nCAI-CA analysis for all the available CDSs of DENV, WNV, JEV, and ZKV, which numbered 4865, 297, 1619 and 494, respectively. Each viral subgroup formed a distinct cluster based on relative synonymous codon usage (Supplementary Figures S3 and S4) and GC content (%GC) (Supplementary Figures S5–S7). In addition, we determined the interspecies and intraspecies variability of the RSCU and %GC in the DENV, JEV, WNV and ZKV genomes. The results show that the RSCU values could differentiate viral subgroups within species and that their distances mostly reflected the evolutionary histories of the viruses (Supplementary Figure S5). The %GC was not a discriminating factor at the intraspecific level (Supplementary Figure S6). At the interspecies level, the clustering patterns based on the RSCU were only slightly more similar to the evolutionary histories of the viruses than the %GC (Supplementary Figure S7).

Next, we used the nCAI-CA algorithm (Figure 1) to identify the optimal hosts of 94 flaviviruses, based on only complete CDSs and codon usage tables from 16 potential hosts (vertebrates: mammals, birds, reptiles and amphibians; arthropods: mosquitoes and ticks) (Supplementary Tables S1–S3). The nCAI-CA algorithm was able to accurately determine host types for MBFVs and UVFVs (vertebrates), IOFVs (*Aedes* mosquitoes), and TBFVs (*Ixodes scapularis*) (Figure 2). The paraphyletic group of dhIOFVs clustered between *Aedes* mosquitoes and vertebrates (Supplementary Figure S1). The CA plot shows a partial overlap between the MBFV and TBFV groups (Supplementary Figure S1); however, on average, TBFVs had higher nCAI values (0.813) than MBFVs (0.765) for *I. scapularis*, suggesting a higher degree of optimization for tick hosts (Table 1 and Supplementary Table S2). The nCAI-CA analysis also revealed unexpected findings for individual viruses, e.g., WNVs clustered within MBFVs but near TBFVs, which aligns with the results of previous infectivity tests [33] and some observational studies (Supplementary Table S1). All the viruses could be classified into two general host groups: vertebrates (MBFVs, TBFVs, UVFVs and dhIOFVs) and mosquitoes (IOFVs) (Supplementary Figure S2). Our analysis shows that no group clusters near *Culex quinquefasciatus*, suggesting that this is not an optimal host for most flaviviruses. However, *Culex* mosquitoes are relatively good vector-hosts for certain flaviviruses (e.g., JEV and WNV) and in many cases the preferred mosquito vector-host is debatable [34–38]. Our results indicate a higher adaptation of MBFV towards *Aedes*; however, a higher genomic adaptation does not imply that *Aedes* is currently the most common host-vector for all MBFV, as additional factors should be considered. Moreover, our algorithm mostly rules out *Anopheles gambiae* as the main host-vector, in agreement with the literature [39], although there are few notable exceptions [40,41]. Overall, the nCAI-CA algorithm is able to predict virus groups and host types with high values of accuracy and specificity in comparison to empirical observations (Table 2 and Supplementary Table S4).



**Figure 2.** Correspondence analysis of the normalized codon adaptation index (nCAI) values of flaviviruses (genus *Flavivirus*;  $n = 94$ ). The plot shows that nCAI can differentiate multiple subgroups of flaviviruses based on their degree of codon usage optimization relative to their host organisms. Mosquito-borne flaviviruses are generally optimized for vertebrate hosts, while tick-borne flaviviruses are optimized for ticks, and insect-only flaviviruses are optimized for mosquitoes. Dual-host insect-only flaviviruses show optimization for both mosquitoes and vertebrates, and unknown vector flaviviruses are also optimized for vertebrates. Dimension 1 explains 89.4% of the variation, and Dimension 2 explains 8.5% of the variation. MBFV: mosquito-borne flaviviruses, TBFV: tick-borne flaviviruses, IOFV: insect-only flaviviruses and UVFV: unknown vector flaviviruses.

The heat map of nCAI values for all flaviviruses shows common adaptation patterns (Supplementary Figure S8). Likely optimal hosts (within an nCAI range of 0.9–1.1) include mammals (*Myotis brandtii*, *M. davidii*, *Mus musculus*, *Bos taurus*, and *Homo sapiens*) and *Aedes* mosquitoes (*Aedes aegypti* and *A. albopictus*). Unlikely hosts due to low adaptation (nCAI < 0.9) include *C. quinquefasciatus*, *A. gambiae*, *I. scapularis* and *S. scrofa*. These results are in accordance with previous studies and observations, e.g., most MBFVs have a reproductive cycle that includes *Aedes* (host-vector) or *Culex* (vector) mosquitoes and a primary mammalian host (Supplementary Table S1). Based on these analyses, flaviviruses are potentially less adapted to reproduction in *Culex* mosquitoes due to the differences in %GC between *Culex* and *Aedes* mosquitoes. Moreover, our analysis suggests that TBFV is a group of flaviviruses optimized to reproduce in vertebrates and use ticks as vectors (and we speculate that they may occasionally reproduce in ticks). In general, flavivirus codon usage is overoptimized (nCAI > 1.1) for birds (*Columba livia*, *G. gallus*, *Anas platyrhynchos*), amphibians (*Xenopus laevis*) and reptiles (*Alligator mississippiensis*).

**Table 1.** Values of %GC3 and nCAI (Mean ± Standard deviation) by flavivirus and host groups.

		Tick	Aedes	Anopheles	Culex	Mammals	Other Vertebrates
<b>Flavivirus Groups</b>	<b>%GC3</b>	<b>72.0%</b>	<b>58.1%</b>	<b>69.6%</b>	<b>69.3%</b>	<b>60.9%</b>	<b>52.9%</b>
		-	±1.8%	-	-	±2.5%	±4.5%
<b>dhIOFV</b> (n = 5)	<b>50.0%</b> ±2.3%	<b>0.738</b> ±0.048	<b>0.949</b> ±0.053	<b>0.729</b> ±0.051	<b>0.764</b> ±0.049	<b>0.889</b> ±0.062	<b>1.027</b> ±0.049
<b>IOFV</b> (n = 14)	<b>54.3%</b> ±3.8%	<b>0.748</b> ±0.046	<b>0.938</b> ±0.044	<b>0.734</b> ±0.042	<b>0.780</b> ±0.047	<b>0.866</b> 0.051	<b>0.982</b> ±0.037
<b>MBFV</b> (n = 49)	<b>52.1%</b> ±3.7%	<b>0.765</b> ±0.033	<b>0.967</b> ±0.033	<b>0.752</b> ±0.029	<b>0.792</b> ±0.032	<b>0.925</b> ±0.051	<b>1.053</b> ±0.035
<b>TBFV</b> (n = 20)	<b>58.9%</b> ±1.8%	<b>0.812</b> ±0.021	<b>0.979</b> ±0.024	<b>0.795</b> ±0.020	<b>0.842</b> ±0.020	<b>0.938</b> ±0.043	<b>1.045</b> ±0.028
<b>UVFV</b> (n = 6)	<b>44.4%</b> ±3.3%	<b>0.705</b> ±0.012	<b>0.948</b> ±0.028	<b>0.706</b> ±0.010	<b>0.746</b> ±0.010	<b>0.907</b> ±0.053	<b>1.057</b> ±0.038
<b>Mosquito</b> (n = 14)	<b>54.3%</b> ±3.8%	<b>0.748</b> ±0.046	<b>0.938</b> ±0.044	<b>0.734</b> ±0.042	<b>0.780</b> ±0.047	<b>0.866</b> ±0.051	<b>0.982</b> ±0.037
<b>Tick</b> (n = 20)	<b>58.9%</b> ±1.8%	<b>0.812</b> ±0.021	<b>0.979</b> ±0.024	<b>0.795</b> ±0.020	<b>0.842</b> ±0.020	<b>0.938</b> ±0.043	<b>1.045</b> ±0.028
<b>Vertebrate</b> (n = 60)	<b>51.2%</b> ±4.3%	<b>0.757</b> ±0.038	<b>0.963</b> ±0.035	<b>0.746</b> ±0.033	<b>0.785</b> ±0.035	<b>0.920</b> ±0.053	<b>1.052</b> ±0.037

Host types: Tick (*Ixodes scapularis*); Aedes (*Aedes albopictus* and *Aedes aegypti*); Anopheles (*Anopheles gambiae*); Culex (*Culex quinquefasciatus*); Mammals (*Homo sapiens*, *Bos taurus*, *Sus scrofa*, *Mus musculus*, *Myotis daividdi* and *Myotis brandtii*); and Other Vertebrates (*Alligator mississippiensis*, *Xenopus laevis*, *Anas platyrhynchos*, *Gallus gallus* and *Columba livia*). Complete list of flaviviruses is available in Supplementary Table S1. nCAI = CAIh/CAIs; nCAI: normalized Codon Adaptation Index; CAIh: Codon Adaptation Index calculated using host codon usage as a reference; CAIs: Codon Adaptation Index calculated with virus codon usage as a reference. %GC3: Percentage of guanine and cytosine at the third codon position. MBFV: mosquito-borne flaviviruses, TBFV: tick-borne flaviviruses, IOFV: insect-only flaviviruses and UVFV: unknown vector flaviviruses.

**Table 2.** Values of specificity and accuracy between nCAI predictions and empirical observations from Supplementary Table S1.

	Virus Group	dhIOFV	IOFV	MBFV	TBFV	UVFV	Host Type	Mosquito	Tick	Vertebrate
Accuracy <sup>1</sup>	91.9%	94.7%	96.8%	81.9%	95.7%	90.4%	86.1%	78.8%	86.7%	92.9%
Specificity <sup>2</sup>	94.9%	94.4%	100.0%	95.6%	94.6%	90.9%	79.6%	75.3%	84.0%	79.5%

$$^1 (TP + TN)/(TP + FN + FP + TN); ^2 TN/(TN + FP).$$

#### 4. Discussion

Despite the high values of specificity and accuracy produced by the nCAI-CA (Table 2 and Supplementary Table S4), there are certain limitations in its application. The algorithm is based on the assumption that there is a selection pressure to optimize the relative use of synonymous codons in the virus. However, it is well known that some viruses use the opposite strategy, and some viruses deoptimize codon usage to hide from host defense mechanisms [42]. The highest level of optimization is at nCAI = 1.0, when the relative use of synonymous codons in the virus and host is identical. Virus–host adaptations were also evaluated with a correspondence analysis (CA) plot (Figure 2). Flaviviruses able to infect a wide range of hosts (generalists) tend to be in the center of the plot, whereas host-specific flaviviruses move away from the center, towards their optimal hosts (Supplementary Figures S1 and S2). Viruses with overoptimized codon usage (nCAI >> 1) might be explained by multiple factors, e.g., adaptation to multiple hosts, effects of extreme %GC bias or adaptation to highly expressed genes [43]. Moreover, some gene-specific codon usage biases may better explain adaptations in certain viruses.

Nevertheless, further empirical investigations are necessary to determine reliable confidence intervals for nCAI. Host determination may be uncertain if viruses display approximately equal optimizations for different host types; for example, although dhIOFV codon usage is optimized for both vertebrate and mosquito hosts, they are insect-specific [14]. Although common host preference patterns are observed, the optimal hosts vary depending on the virus or subgroup and may not reflect documented cases (Supplementary Table S1). The observed host ranges also do not distinguish between vectors and hosts, and classical phylogenomic methods cannot determine potential hosts without confirmed cases. Our nCAI-based method overcomes this limitation by directly measuring the adaptation of viruses to the translational machinery of their hosts.

## 5. Conclusions

In conclusion, this novel algorithm provides a fast and proactive method to assess the potential host ranges and the risk of zoonotic host shift for new and emerging viruses. In flaviviruses, this method distinguishes between arthropod and vertebrate hosts with high accuracy. However, it might produce ambivalent results for viruses undergoing host shifts. Overall, this nCAI-based algorithm may be used as a complement to current phylogenetic methods to monitor current and future outbreaks.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/life11050442/s1>, Supplementary Figures S1–S9 and Supplementary Tables S1–S4.

**Author Contributions:** Conceptualization, P.P.; methodology, P.P.; software, P.P.; validation, P.P. and S.G.-V.; formal analysis, P.T.N.; investigation, P.T.N. and P.P.; resources, P.T.N. and P.P.; data curation, P.T.N. and P.P.; writing—original draft preparation, P.T.N.; writing—review and editing, P.P. and S.G.-V.; visualization, P.T.N. and P.P.; supervision, P.P.; project administration, P.P.; funding acquisition, P.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Turku Collegium for Science and Medicine (Turku, Finland).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data used in this study is included within the manuscript or a Supplementary Material.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Kindhauser, M.K.; Allen, T.; Frank, V.; Santhana, R.S.; Dye, C. Zika: The origin and spread of a mosquito-borne virus. *Bull. World Health Organ.* **2016**, *94*, 675C–686C. [[CrossRef](#)] [[PubMed](#)]
- Omilabu, S.A.; Salu, O.B.; Oke, B.O.; James, A.B. The West African ebola virus disease epidemic 2014–2015: A commissioned review. *Niger. Postgrad. Med. J.* **2016**, *23*, 49–56. [[CrossRef](#)]
- Wang, C.; Horby, P.W.; Hayden, F.G.; Gao, G.F. A novel coronavirus outbreak of global health concern. *Lancet* **2020**, *395*, 470–473. [[CrossRef](#)]
- Girard, M.P.; Tam, J.S.; Assossou, O.M.; Kiény, M.P. The 2009 A (H1N1) influenza virus pandemic: A review. *Vaccine* **2010**, *28*, 4895–4902. [[CrossRef](#)] [[PubMed](#)]
- Gates, B. Responding to Covid-19-A Once-in-a-Century Pandemic? *N. Engl. J. Med.* **2020**, *382*, 1677–1679. [[CrossRef](#)]
- Smith, D. Applications of bioinformatics and computational biology to influenza surveillance and vaccine strain selection. *Vaccine* **2003**, *21*, 1758–1761. [[CrossRef](#)]
- Sanjuán, R.; Nebot, M.R.; Chirico, N.; Mansky, L.M.; Belshaw, R. Viral mutation rates. *J. Virol.* **2010**, *84*, 9733–9748. [[CrossRef](#)]
- Kuno, G. Host range specificity of flaviviruses: Correlation with in vitro replication. *J. Med. Entomol.* **2007**, *44*, 93–101. [[CrossRef](#)] [[PubMed](#)]
- Di Giallonardo, F.; Schlub, T.E.; Shi, M.; Holmes, E.C. Dinucleotide Composition in Animal RNA Viruses Is Shaped More by Virus Family than by Host Species. *J. Virol.* **2017**, *91*. [[CrossRef](#)] [[PubMed](#)]
- Parry, R.; Asgari, S. Discovery of Novel Crustacean and Cephalopod Flaviviruses: Insights into the Evolution and Circulation of Flaviviruses between Marine Invertebrate and Vertebrate Hosts. *J. Virol.* **2019**, *93*. [[CrossRef](#)]
- Jenkins, G.M.; Holmes, E.C. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.* **2003**, *92*, 1–7. [[CrossRef](#)]
- Chambers, T.J.; Hahn, C.S.; Galler, R.; Rice, C.M. Flavivirus genome organization, expression, and replication. *Annu. Rev. Microbiol.* **1990**, *44*, 649–688. [[CrossRef](#)] [[PubMed](#)]
- Simmonds, P.; Becher, P.; Bukh, J.; Gould, E.A.; Meyers, G.; Monath, T.; Muerhoff, S.; Pletnev, A.; Rico-Hesse, R.; Smith, D.B.; et al. Ictv Report Consortium ICTV virus taxonomy profile: Flaviviridae. *J. Gen. Virol.* **2017**, *98*, 2–3. [[CrossRef](#)] [[PubMed](#)]
- Huhtamo, E.; Cook, S.; Moureau, G.; Uzcátegui, N.Y.; Sironen, T.; Kuivanen, S.; Putkuri, N.; Kurkela, S.; Harbach, R.E.; Firth, A.E.; et al. Novel flaviviruses from mosquitoes: Mosquito-specific evolutionary lineages within the phylogenetic group of mosquito-borne flaviviruses. *Virology* **2014**, *464–465*, 320–329. [[CrossRef](#)]
- Alkan, C.; Zapata, S.; Bichaud, L.; Moureau, G.; Lemey, P.; Firth, A.E.; Gritsun, T.S.; Gould, E.A.; de Lamballerie, X.; Depaquit, J.; et al. Ecuador Paraiso Escondido Virus, a New Flavivirus Isolated from New World Sand Flies in Ecuador, Is the First Representative of a Novel Clade in the Genus Flavivirus. *J. Virol.* **2015**, *89*, 11773–11785. [[CrossRef](#)]
- Lobo, F.P.; Mota, B.E.F.; Pena, S.D.J.; Azevedo, V.; Macedo, A.M.; Tauch, A.; Machado, C.R.; Franco, G.R. Virus-host coevolution: Common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS ONE* **2009**, *4*, e6282. [[CrossRef](#)] [[PubMed](#)]



17. Guarner, J.; Hale, G.L. Four human diseases with significant public health impact caused by mosquito-borne flaviviruses: West Nile, Zika, dengue and yellow fever. *Semin. Diagn. Pathol.* **2019**, *36*, 170–176. [[CrossRef](#)]
18. 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.* **2009**, *5*, 311. [[CrossRef](#)] [[PubMed](#)]
19. Sharp, P.M.; 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. [[CrossRef](#)]
20. Puigbò, P.; Bravo, I.G.; Garcia-Vallve, S. CAIcal: A combined set of tools to assess codon usage adaptation. *Biol. Direct* **2008**, *3*, 38. [[CrossRef](#)]
21. Nakamura, Y.; Gojobori, T.; Ikemura, T. Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Res.* **2000**, *28*, 292. [[CrossRef](#)] [[PubMed](#)]
22. O’Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [[CrossRef](#)]
23. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2013**, *41*, D36–D42. [[CrossRef](#)] [[PubMed](#)]
24. Gaunt, M.W.; Sall, A.A.; de Lamballerie, X.; Falconar, A.K.; Dzhivanian, T.I.; Gould, E.A. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J. Gen. Virol.* **2001**, *82*, 1867–1876. [[CrossRef](#)] [[PubMed](#)]
25. Kuno, G.; Chang, G.J.; Tsuchiya, K.R.; Karabatsos, N.; Cropp, C.B. Phylogeny of the genus *Flavivirus*. *J. Virol.* **1998**, *72*, 73–83. [[CrossRef](#)] [[PubMed](#)]
26. Grard, G.; Moureau, G.; Charrel, R.N.; Holmes, E.C.; Gould, E.A.; de Lamballerie, X. Genomics and evolution of Aedes-borne flaviviruses. *J. Gen. Virol.* **2010**, *91*, 87–94. [[CrossRef](#)] [[PubMed](#)]
27. Mihara, T.; Nishimura, Y.; Shimizu, Y.; Nishiyama, H.; Yoshikawa, G.; Uehara, H.; Hingamp, P.; Goto, S.; Ogata, H. Linking Virus Genomes with Host Taxonomy. *Viruses* **2016**, *8*, 66. [[CrossRef](#)]
28. Sharp, P.M.; Li, W.H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **1986**, *24*, 28–38. [[CrossRef](#)]
29. Sharp, P.M.; Tuohy, T.M.; Mosurski, K.R. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* **1986**, *14*, 5125–5143. [[CrossRef](#)]
30. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
31. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **2009**, *26*, 1641–1650. [[CrossRef](#)]
32. Federhen, S. The NCBI Taxonomy database. *Nucleic Acids Res.* **2012**, *40*, D136–D143. [[CrossRef](#)]
33. Lawrie, C.H.; Uzcátegui, N.Y.; Armesto, M.; Bell-Sakyi, L.; Gould, E.A. Susceptibility of mosquito and tick cell lines to infection with various flaviviruses. *Med. Vet. Entomol.* **2004**, *18*, 268–274. [[CrossRef](#)]
34. Sotomayor-Bonilla, J.; Tolsá-García, M.J.; García-Peña, G.E.; Santiago-Alarcon, D.; Mendoza, H.; Alvarez-Mendizabal, P.; Rico-Chávez, O.; Sarmiento-Silva, R.E.; Suzán, G. Insights into the Host Specificity of Mosquito-Borne Flaviviruses Infecting Wild Mammals. *Ecohealth* **2019**, *16*, 726–733. [[CrossRef](#)] [[PubMed](#)]
35. Wagner, S.; Mathis, A.; Schönenberger, A.C.; Becker, S.; Schmidt-Chanasit, J.; Silaghi, C.; Veronesi, E. Vector competence of field populations of the mosquito species *Aedes japonicus japonicus* and *Culex pipiens* from Switzerland for two West Nile virus strains. *Med. Vet. Entomol.* **2018**, *32*, 121–124. [[CrossRef](#)] [[PubMed](#)]
36. Liu, Z.; Zhou, T.; Lai, Z.; Zhang, Z.; Jia, Z.; Zhou, G.; Williams, T.; Xu, J.; Gu, J.; Zhou, X.; et al. Competence of *Aedes aegypti*, *Ae. albopictus*, and *Culex quinquefasciatus* Mosquitoes as Zika Virus Vectors, China. *Emerg. Infect. Dis.* **2017**, *23*, 1085–1091. [[CrossRef](#)] [[PubMed](#)]
37. Vaidyanathan, R.; Scott, T.W. Geographic variation in vector competence for West Nile virus in the *Culex pipiens* (Diptera: Culicidae) complex in California. *Vector Borne Zoonotic Dis.* **2007**, *7*, 193–198. [[CrossRef](#)]
38. Ndiaye, E.H.; Fall, G.; Gaye, A.; Bob, N.S.; Talla, C.; Diagne, C.T.; Diallo, D.; Yamar, B.A.; Dia, I.; Kohl, A.; et al. Vector competence of *Aedes vexans* (Meigen), *Culex poicilipes* (Theobald) and *Cx. quinquefasciatus* Say from Senegal for West and East African lineages of Rift Valley fever virus. *Parasit. Vectors* **2016**, *9*, 94. [[CrossRef](#)]
39. Nanfack Minkeu, F.; Vernick, K.D. A systematic review of the natural virome of anopheles mosquitoes. *Viruses* **2018**, *10*, 222. [[CrossRef](#)]
40. Lequime, S.; Lambrechts, L. Discovery of flavivirus-derived endogenous viral elements in Anopheles mosquito genomes supports the existence of Anopheles-associated insect-specific flaviviruses. *Virus Evol.* **2017**, *3*, vew035. [[CrossRef](#)]
41. Colmant, A.M.G.; Hobson-Peters, J.; Bielefeldt-Ohmann, H.; van den Hurk, A.F.; Hall-Mendelin, S.; Chow, W.K.; Johansen, C.A.; Fros, J.; Simmonds, P.; Watterson, D.; et al. A New Clade of Insect-Specific Flaviviruses from Australian Anopheles Mosquitoes Displays Species-Specific Host Restriction. *mSphere* **2017**, *2*. [[CrossRef](#)] [[PubMed](#)]

- 
42. Zhou, J.; Gao, Z.; Zhang, J.; Chen, H.; Pejsak, Z.; Ma, L.; Ding, Y.; Liu, Y. Comparative [corrected] codon usage between the three main viruses in pestivirus genus and their natural susceptible livestock. *Virus Genes* **2012**, *44*, 475–481. [[CrossRef](#)] [[PubMed](#)]
  43. Rocha, E.P.C. Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. *Genome Res.* **2004**, *14*, 2279–2286. [[CrossRef](#)] [[PubMed](#)]