# Data-driven identification of potential Zika virus vectors

Michelle V. Evans[1,5*], Tad A. Dallas[1,2], Barbara A. Han[3], Courtney C. Murdock[1,4,5,6,7,8], John M. Drake[1,5,8]

**1** Odum School of Ecology, University of Georgia, Athens, GA, USA
**2** Department of Environmental Science and Policy, University of California-Davis, Davis, CA, USA
**3** Cary Institute of Ecosystem Studies, Millbrook, NY, USA
**4** Department of Infectious Disease, University of Georgia, Athens, GA, USA
**5** Center for the Ecology of Infectious Diseases, University of Georgia, Athens, GA, USA
**6** Center for Tropical Emerging Global Diseases, University of Georgia, Athens, GA, USA
**7** Center for Vaccines and Immunology, University of Georgia, Athens, GA, USA
**8** River Basin Center, University of Georgia, Athens, GA, USA

*Corresponding Author: mvevans@uga.edu

## Abstract

Zika is an emerging virus whose rapid spread is of great public health concern. Knowledge about transmission remains incomplete, especially concerning potential transmission in geographic areas in which it has not yet been introduced. To identify unknown vectors of Zika, we developed a data-driven model linking vector species and the Zika virus via vector-virus trait combinations that confer a propensity toward associations in an ecological network connecting flaviviruses and their mosquito vectors. Our model predicts that thirty-five species may be able to transmit the virus, seven of which are found in the continental United States, including *Culex quinquefasciatus* and *Cx. pipiens*. We suggest that empirical studies prioritize these species to confirm predictions of vector competence, enabling the correct identification of populations at risk for transmission within the United States.

# Introduction

In 2014, Zika virus was introduced into Brazil and Haiti, from where it rapidly spread throughout the Americas. By Jauary 2017, over 100,000 cases had been confirmed in 24 different states in Brazil (`http://ais.paho.org/phip/viz/ed_zika_cases.asp`), with large numbers of reports from many other counties in South and Central America (Faria et al. 2016). Originally isolated in Uganda in 1947, the virus remained poorly understood until it began to spread within the South Pacific, including an outbreak affecting 75% of the residents on the island of Yap in 2007 (49 confirmed cases) and over 32,000 cases in the rest of Oceania in 2013-2014, the largest outbreak prior to the Americas (2016-present) (Cao-Lormeau et al. 2016, Duffy et al. 2009). Guillian-Barré syndrome, a neurological pathology associated with Zika virus infection, was first recognized at this time (Cao-Lormeau et al. 2016). Similarly, an increase in newborn microcephaly was found to be correlated with the increase in Zika cases in Brazil in 2015 and 2016 (Schuler-Faccini et al. 2016). For this reason, in February 2016, the World Health Organization declared the American Zika virus epidemic to be a Public Health Emergency of International Concern.

Despite its public health importance, the ecology of Zika virus transmission has been poorly understood until recently. It has been presumed that *Aedes aegypti* and *Ae. albopictus* are the primary vectors due to epidemiologic association with Zika virus (Messina et al. 2016), viral isolation from and transmission experiments with field populations (especially in *Ae. aegypti* (Haddow et al. 2012, Boorman and Porterfield 1956, Haddow et al. 1964)), and association with related arboviruses (e.g. dengue fever virus, yellow fever virus). Predictions of the potential geographic range of Zika virus in the United States,and associated estimates for the size of the vulnerable population, are therefore primarily based on the distributions of *Ae. aegypti* and *Ae. albopictus*, which jointly extend across the Southwest, Gulf coast, and mid-Atlantic regions of the United States (Centers for Disease Control and Prevention 2016). We reasoned, however, that if other, presently unidentified Zika-competent mosquitoes exist in the Americas, then these projections may be too restricted and therefore optimistically biased. Additionally, recent experimental studies show that the ability of *Ae. aegypti* and *Ae. albopictus* to transmit the virus varies significantly across mosquito populations and geographic regions (Chouin-Carneiro et al. 2016), with some pop-

ulations exhibiting low dissemination rates even though the initial viral titer after inoculation may be high (Diagne et al. 2015). This suggests that in some locations other species may be involved in transmission. The outbreak on Yap, for example, was driven by a different species, *Ae. hensilli* (Ledermann et al. 2014). Closely related viruses of the *Flaviviridae* family are vectored by over nine mosquito species, on average (see Supplementary Data). Thus, because Zika virus may be associated with multiple mosquito species, we considered it necessary to develop a more comprehensive list of potential Zika vectors.

The gold standard for identifying competent disease vectors requires isolating virus from field-collected mosquitoes, followed by experimental inoculation and laboratory investigation of viral dissemination throughout the body and to the salivary glands (Barnett 1960, Hardy et al. 1983), and, when possible, successful transmission back to the vertebrate host (e.g. (Komar et al. 2003)). Unfortunately, these methods are costly, often underestimate the risk of transmission (Bustamante and Lord 2010), and the amount of time required for analyses can delay decision making during an outbreak (Day 2001). To address the problem of identifying potential vector candidates in an actionable time frame, we therefore pursued a data-driven approach to identifying candidate vectors aided by machine learning algorithms for identifying patterns in high dimensional data. If the propensity of mosquito species to associate with Zika virus is statistically associated with common mosquito traits, it is possible to rank mosquito species by the degree of risk represented by their traits – a comparative approach similar to the analysis of risk factors in epidemiology. For instance, a model could be constructed to estimate the statistical discrepancy between the traits of known vectors (i.e., *Ae. aegypti*, *Ae. albopictus*, and *Ae. hensilli*) and the traits of all possible vectors. Unfortunately, this simplistic approach would inevitably fail due to the small amount of available data (i.e., sample size of 3). Thus, we developed an indirect approach that leverages information contained in the associations among many virus-mosquito pairs to inform us about specific associations. Specifically, our method identifies covariates associated with the propensity for mosquito species to vector any flavivirus. From this, we constructed a model of the mosquito-flavivirus network and then extracted from this model the life history profile and species list of mosquitoes predicted to associate with Zika virus, which we recommend be experimentally

tested for Zika virus competence.

# Methods

## Data Collection and Feature Construction

Our dataset comprised a matrix of vector-virus pairs relating all known flaviviruses and their mosquito vectors. To construct this matrix, we first compiled a list of mosquito-borne flaviviruses to include in our study (Van Regenmortel et al. 2000, Kuno et al. 1998, Cook and Holmes 2005). Viruses that only infect mosquitoes and are not known to infect humans were not included. Using this list, we constructed a mosquito-virus pair matrix based on the Global Infectious Diseases and Epidemiology Network database (GIDEON 2016), the International Catalog of Arboviruses Including Certain Other Viruses of Vertebrates (ArboCat) (Karabatsos 1985), *The Encyclopedia of Medical and Veterinary Entomology* (Russell et al. 2013) and Mackenzie et al. (2012).

We defined a known vector-virus pair as one for which the full transmission cycle (i.e, infection of mosquito via an infected host (mammal or avian) or bloodmeal that is able to be transmitted via saliva) has been observed. Basing vector competence on isolation or intrathoracic injection bypasses several important barriers to transmission (Hardy et al. 1983), and may not be true evidence of a mosquito's ability to transmit an arbovirus. We found our definition to be more conservative than that which is commonly used in disease databases (e.g. Global Infectious Diseases and Epidemiology Network database), which often assumes isolation from wild-caught mosquitoes to be evidence of a mosquito's role as vector. Therefore, a supplementary analysis investigates the robustness of our findings with regards to uncertainty in vector status by comparing the analysis reported in the main text to a second analysis in which any kind of evidence for association, including merely isolating the virus in wild-caught mosquitoes, is taken as a basis for connection in the virus-vector network (see Appendix I for analysis and results).

Fifteen mosquito traits (Appendix II, Table 1) and twelve virus traits (Appendix II, Table 2) were collected from the literature. For the mosquito species, the geographic range was defined as the number of countries in which the species has been collected, based on Walter Reed Biosys-

tematics Unit (2016). While there are uncertainties in species' ranges due to false absences, this represents the most comprehensive, standardized dataset available that includes both rare and common mosquito species. A species' continental extent was recorded as a binary value of its presence by continent. A species' host range was defined as the number of taxonomic classes the species is known to feed on, with the Mammalia class further split into non-human primates and other mammals, because of the important role primates play in zoonotic spillovers of vector-borne disease (e.g. dengue, chikungunya, yellow fever, and Zika viruses) (Weaver 2005, Diallo et al. 2005*a*, Weaver et al. 2016). The total number of unique flaviviruses observed per mosquito species was calculated from our mosquito-flavivirus matrix. All other traits were based on consensus in the literature (see Appendix III for sources by species). For three traits – urban preference, endophily (a proclivity to bite indoors), and salinity tolerance – if evidence of that trait for a mosquito was not found in the literature, it was assumed to be negative.

We collected data on the following virus traits: host range (Mahy 2009, Mackenzie et al. 2012, Chambers and Monath 2003, Cook and Zumla 2009), disease severity (Mackenzie et al. 2012), human illness (Chambers and Monath 2003, Cook and Zumla 2009), presence of a mutated envelope protein, which controls viral entry into cells (Grard et al. 2009), year of isolation (Karabatsos 1985), and host range (Karabatsos 1985). Disease severity was based on Mackenzie et al. (2012), ranging from no known symptoms (e.g. Kunjin virus) to severe symptoms and significant human mortality (e.g. yellow fever virus). For each virus, vector range was calculated as the number of mosquito species for which the full transmission cycle has been observed. Genome length was calculated as the mean of all complete genome sequences listed for each flavivirus in the Virus Pathogen Database and Analysis Resource (`http://www.viprbrc.org/`). For more recently discovered flaviviruses not yet cataloged in the above databases (i.e., New Mapoon Virus, Iquape virus), viral traits were gathered from primary literature (sources listed in Appendix III).

## Predictive model

Following Han et al. (2015), boosted regression trees (BRT) (Friedman 2001) were used to fit a logistic-like predictive model relating the status of all possible virus-vector pairs (0: not associated,

1: associated) to a predictor matrix comprising the traits of the mosquito and virus traits in each pair. Boosted regression trees circumvent many issues associated with traditional regression analysis (Elith et al. 2008), allowing for complex variable interactions, collinearity, non-linear relationships between covariates and response variables, and missing data. Additionally, this technique performs well in comparison with other logistic regression approaches (Friedman 2001). Trained boosted regression tree models are dependent on the split between training and testing data, such that each model might predict slightly different propensity values. To address this, we trained an ensemble of 25 internally cross-validated BRT models on independent partitions of training and testing data. The resulting model demonstrated low variance in relative variable importance and overall model accuracy, suggesting models all converged to a similar result.

Prior to the analysis of each model, we randomly split the data into training (70%) and test (30%) sets while preserving the proportion of positive labels (known associations) in each of the training and test sets. Models were trained using the `gbm` package in *R* (Ridgeway 2015), with the maximum number of trees set to 25,000, a learning rate of 0.001, and an interaction depth of 5. To correct for optimistic bias (Smith et al. 2014), we performed 10-fold cross validation and chose a bag fraction of 50% of the training data for each iteration of the model. We estimated the performance of each individual model with three metrics: Area Under the Receiver Operator Curve, specificity, and sensitivity. For specificity and sensitivity, which require a preset threshold, we thresholded predictions on the testing data based on the value which maximized the sum of the sensitivity and specificity, a threshold robust to the ratio of presence to background points in presence-only datasets (Liu et al. 2015). Variable importance was quantified by permutation (Breiman 2001) to assess the relative contribution of virus and vector traits to the propensity for a virus and vector to form a pair. Because we transformed many categorical variables into binary variables (e.g., continental range as binary presence or absence by continent), the sum of the relative importance for each binary feature was summed to obtain a single value for the entire variable.

Each of our twenty-five trained models was then used to predict novel mosquito vectors of Zika by applying the trained model to a data set consisting of the virus traits of Zika paired with the

traits of all mosquitoes for which flaviviruses have been isolated from wild caught individuals, and, depending on the species, may or may not have been tested in full transmission cycle experiments (a total of 180 mosquito species). This expanded dataset allowed us to predict over a large number of mosquito species, while reasonably limiting our dataset to those species suspected of transmitting flaviviruses. The output of this model was a propensity score ranging from 0 to 1. In our case, the final propensity score for each vector was the mean propensity score assigned by the twenty-five models. To label unobserved edges, we thresholded propensity scores at the value of lowest ranked known vector (Liu et al. 2013).

## Model Validation

In addition to conventional performance metrics, we conducted additional analyses to further validate both this method of prediction, and our model specifically. To account for uncertainty in the vector-virus links in our initial matrix, we repeated our analysis for a vector-virus matrix with a less conservative definition of a postive link (field isolation and above), referred to as our supplementary model. Vector competence is a dynamic trait, and there exists significant intraspecific variation in the ability of a vector to transmit a virus for certain species of mosquitoes (Diallo et al. 2005*b*, Gubler et al. 1979). Our supplementary model is based on a less conservative definition of vector competence and includes species implicated as vectors, but not yet verified through laboratory competence studies, and therefore accounts for additional uncertainty such as intraspecific variation.

While this approach is well-tested in epidemiological applications (Parascandola 2004), it has only recently been applied to predict ecological associations, and, as such, has limitations unique to this application. To further evaluate this prediction method, we performed a modified "leave-one-out" analysis, whereby we trained a model to a dataset from which a well-studied virus had been ommitted, and then predicted vectors for this virus and compared them against a list of known vectors. We repeated this analysis for West Nile, dengue, and yellow fever viruses, following the same method of training as for our original model. While this analysis differs from our original method, it provides a more stringent evaluation of this method of prediction because the model is

trained on an incomplete dataset and predicts on unfamiliar data, a more difficult task than that posed to our original model.

# Results

In total, we identified 132 vector-virus pairs, consisting of 77 mosquito species and 37 flaviviruses. The majority of these species were *Aedes* (32) or *Culex* (24) species. Our supplementary dataset consisted of an additional 103 mosquito species suspected to transmit flaviviruses, but for which evidence of a full transmission cycle does not exist. This resulted in 180 potential mosquito-Zika pairs on which to predict with our trained model. As expected, closely related viruses, such as the four strains of dengue, shared many of the same vectors and were clustered in our network diagram (Fig. 1). The distribution of vectors to viruses was uneven, with a few viruses vectored by many mosquito species, and rarer viruses vectored by only one or two species. The virus with the most known competent vectors was West Nile virus (31 mosquito vectors), followed by yellow fever virus (24 mosquito vectors). In general, encephalitic viruses such as West Nile virus were found to be more commonly vectored by *Culex* mosquitoes and hemorrhagic viruses were found to be more commonly vectored by *Aedes* mosquitoes (see Gould and Solomon (2008) for further distinctions within *Flaviviridae*) (Fig. 1).

Our ensemble of BRT models trained on common vector and virus traits predicted mosquito vector-virus pairs in the test dataset with high accuracy ($AUC = 0.92 \pm 0.02$; $sensitivity = 0.858 \pm 0.04$; $specificity = 0.872 \pm 0.04$). Due to non-monotonicity and existence of interactions among predictor variables within our model, one cannot make general statements about the directionality of effect. Thus, we focus on the relative importance of different variables to model performance. The most important variable for accurately predicting the presence of vector-virus pair was the subgenus of the mosquito species, followed by continental range (e.g. continents on which species are present). The number of viruses vectored by a mosquito species and number of mosquito vectors of a virus were the third and fifth most important variables, respectively. Unsurprisingly, this suggests that, when controlling for other variables, mosquitoes and viruses with more known vector-virus pairs (i.e., more viruses vectored and more hosts infected, respectively), are more

likely to be part of a predicted pair by the model. Mosquito ecological traits such as larval habitat and salinity tolerance were generally less important than a species' phylogeny or geographic range (Figure 2).

When applied to the 180 potential mosquito-Zika pairs, the model predicted thirty-five vectors to be ranked above the threshold (set at the value of the lowest-ranked known vector), for a total of nine known vectors and twenty-six novel, predicted mosquito vectors of Zika (Table 1). Of these vectors, there were twenty-four *Aedes* species, nine *Culex* species, one *Psorophora* species, and one *Runchomyia* species. The GBM model's top two ranked vectors for Zika are the most highly-suspected vectors of Zika virus, *Ae. aegypti* and *Ae. albopictus*.

## Model Validation

Our supplementary and primary models generally concur and their ranking of potential Zika virus vectors are highly correlated ($\rho = 0.508$ and $\rho = 0.693$ on raw and thresholded predictions, respectively). As one might expect, the supplementary model assigned fewer scores of low propensity (App. 1, Fig. 2), suggesting that incorporating this additional uncertainty in the training dataset eroded the models ability to distinguish negative links. The supplementary model's performance on the testing data ($AUC = 0.84 \pm 0.02$), however, indicates that the additional uncertainty did not impede model performance.

When trained on "leave-one-out" datasets, all three models were able to predict the testing data with high accuracy ($AUC = 0.91$, $AUC = 0.91$, $AUC = 0.92$ for West Nile, dengue, and yellow fever viruses, respectively). Performance varied when models were validated against predictions of "known outcomes". A model trained without West Nile virus predicted highly linked vectors reasonably well ($AUC = 0.69$), however it assigned low scores to rarer "known" vectors, such as *Culiseta inornata*, which was only associated with West Nile virus. Similarly, the model trained on the dengue-omitted dataset predicted training data and vectors of dengue itself with high accuracy ($AUC = 0.92$). While the model trained without yellow fever performed well on the testing data, it performed poorly when predicting vectors of yellow fever virus ($AUC = 0.47$). Unlike West Nile and dengue viruses, the majority of the known vectors of yellow fever are only

associated with yellow fever (i.e. a single vector-virus link), and so were excluded completely from the training data when all yellow fever links were omitted. Additionally, several of the vectors species are of the *Haemagogus* genus, which was completely absent from the training data. Given the importance of phylogeny of the vector species in predicting vector-virus links, it follows that a dataset with a novel subgenus would be difficult for the model to predict on, resulting in low model performance. The low performance of this model illustrates that incorporating common traits and additional vector-virus links improves model prediction. When traits were not available in the training dataset, model performance was much lower, suggesting that there exists a statistical association between a vectors' traits and its ability to transmit a virus.

# Discussion

Zika virus is unprecedented among emerging arboviruses in its combination of severe public health hazard, rapid spread, and poor scientific understanding. Particularly crucial to public health preparedness is knowledge about the geographic extent of potentially at risk populations and local environmental conditions for transmission, which are determined by the presence of competent vectors. Until now, identifying additional competent vector species has been a low priority because Zika virus has historically been geographically restricted to a narrow region of equatorial Africa and Asia (Petersen et al. 2016), and the mild symptoms of infection made its range expansion since the 1950's relatively unremarkable. However, with its relatively recent and rapid expansion into the Americas and its association with severe neurological disorders, the prediction of potential disease vectors in non-endemic areas has become a matter of critical public health importance. We identify these potential vector species by developing a data-driven model that identifies candidate vector species of Zika virus by leveraging data on traits of mosquito vectors and their flaviviruses. We suggest that empirical work should prioritize these species in their evaluation of vector competence of mosquitoes for Zika virus.

Our model predicts that fewer than one third of the potential mosquito vectors of Zika virus have been identified, with over twenty-five additional mosquito species worldwide that may have the capacity to contribute to transmission. The continuing focus in the published literature on

two species known to transmit Zika virus (*Ae. aegypti* and *Ae. albopictus*) ignores the potential role of other vectors, potentially misrepresenting the spatial extent of risk. In particular, four species predicted by our model to be competent vectors – *Ae. vexans*, *Culex quinquefasciatus*, *Cx. pipiens*, and *Cx. tarsalis* – are found throughout the continental United States. Further, the three *Culex* species are primary vectors of West Nile virus (Farajollahi et al. 2011). *Cx. quinquefasciatus* and *Cx. pipiens* were ranked 3rd and 17th by our model, respectively, and together these species were the highest-ranking species endemic to the United States after the known vectors (*Ae. aegypti* and *Ae. albopictus*). *Cx. quinquefasciatus* has previously been implicated as an important vector of encephalitic flaviviruses, specifically West Nile virus and St. Louis encephalitis (Turrell et al. 2005, Hayes et al. 2005), and a hybridization of the species with *Cx. pipiens* readily bites humans (Fonseca et al. 2004). The empirical data available on the vector competence of *Cx. pipiens* and *Cx. quinquefasciatus* is currently mixed, with some studies finding evidence for virus transmission and others not (Guo et al. 2016, Aliota et al. 2016, Fernandes et al. 2016, Huang et al. 2016). These results suggest, in combination with evidence for significant genotype $\times$ genotype effects on the vector competence of *Ae. aegypti* and *Ae. albopictus* to transmit Zika (Chouin-Carneiro et al. 2016), that the vector competence of *Cx. pipiens* and *Cx. quinquefasciatus* for Zika virus could be highly dependent upon the genetic background of the mosquito-virus pairing, as well as local environmental conditions. Thus, considering their anthropophilic natures and wide geographic ranges, *Cx. quinquefasciatus* and *Cx. pipiens* could potentially play a larger role in the transmission of Zika in the continental United States. Further experimental research into the competence of populations of *Cx. pipiens* to transmit Zika virus across a wider geographic range is therefore highly recommended, and should be prioritized.

The vectors predicted by our model have a combined geographic range much larger than that of the currently suspected vectors of Zika (Fig. 3), suggesting that, were these species to be confirmed as vectors, a larger population may be at risk of Zika infection than depicted by maps focusing solely on *Ae. aegypti* and *Ae. albopictus*. The range of *Cx. pipiens* includes the Pacific Northwest and the upper mid-West, areas that are not within the known range of *Ae. aegypti* or *Ae. albopictus* (Darsie and Ward 2005). Furthermore, *Ae. vexans*, another predicted vector of

Zika virus, is found throughout the continental US and the range of *Cx. tarsalis* extends along the entire West coast (Darsie and Ward 2005). On a finer scale, these species use a more diverse set of habitats, with *Ae. aegypti* and *Cx. quinquefasciatus* mainly breeding in artificial containers, and *Ae. vexans* and *Ae. albopictus* being relatively indiscriminate in their breeding sites, including breeding in natural sites such as tree holes and swamps. Therefore, in addition to the wider geographic region supporting potential vectors, these findings suggest that both rural and urban areas could serve as habitat for potential vectors of Zika. We recommend experimental tests of these species for competency to transmit Zika virus, because a confirmation of these vectors would necessitate expanding public health efforts into these areas not currently considered at risk.

While transmission requires a competent vector, vector competence does not necessarily equal transmission risk or inform vectorial capacity. There are many biological factors that, in conjunction with positive vector competence, determine a vector's role in disease transmission. For example, although *Ae. aegypti* mosquitoes are efficient vectors of West Nile virus, they prefer to feed on humans, which are dead-head hosts for the disease, and therefore have low potential to serve as a vector (Turell et al. 2005). *Psorophora ferox*, although predicted by our model as a potential vector of Zika virus, would likely play a limited role in transmission because it rarely feeds on humans (Molaei et al. 2008). Additionally, vector competence is dynamic, and may be mediated by environmental factors that influence viral development and mosquito immunity (Muturi and Alto 2011). Therefore, our list of potential vectors of Zika represents a comprehensive starting point, which should be furthered narrowed by empirical work and consideration of biological details that impact transmission dynamics. Given the severe neurological side-effects of Zika virus infection, beginning with the most conservative method of vector prediction ensures that risk is not underestimated, and allows public health agencies to interpret the possibility of Zika transmission given local conditions.

Our model serves as a starting point to streamlining empirical efforts to identify areas and populations at risk for Zika transmission. While our model enables data-driven predictions about the geographic area at potential risk of Zika transmission, subsequent empirical work investigating Zika vector competence and transmission efficiency is required for model validation, and to inform

future analyses of transmission dynamics. For example, in spite of its low transmission efficiency in certain geographic regions (Chouin-Carneiro et al. 2016), *Ae. aegypti* is anthropophilic (Powell et al. 2013), and may therefore pose a greater risk of human-to-human Zika virus transmission than mosquitoes that bite a wider variety of animals. On the other hand, mosquito species that prefer certain hosts in rural environments are known to alter their feeding behaviors to bite alternative hosts (e.g., humans and rodents) in urban settings, due to changes in host community composition (Chaves et al. 2010). Environmental factors such as precipitation and temperature directly influence mosquito populations, and determine the density of vectors in a given area (Thomson et al. 2006), an important factor in transmission risk. Additionally, socio-economic factors such as housing type and lifestyle can decrease a populations' contact with mosquito vectors, and lower the risk of transmission to humans (Moreno-Madriñán and Turell 2017). Effective risk modeling and forecasting the range expansion of Zika virus in the United States will depend on validating the vector status of these species, as well as resolving behavioral and biological details that impact transmission dynamics.

Although we developed this model with Zika virus in mind, our findings have implications for other emerging flaviviruses and contribute to recently developed methodology applying machine learning methods to the prediction of unknown agents of infectious diseases. This technique has been used to predict rodent reservoirs of disease (Han et al. 2015) and bat carriers of filoviruses (Han et al. 2016) by training models with host-specific data. Our model, however, incorporates additional data by constructing a vector-virus network that is used to inform predictions of vector-virus associations. The combination of common virus traits with vector-specific traits enabled us to predict potential mosquito vectors of specific flaviviruses, and to train the model on additional information distributed throughout the the flavivirus-mosquito network.

Uncertainty in our model arises through uncertainty inherent in our datasets. Vector status is not static (e.g. mutation in the chikungunya virus to increase transmission by *Ae. albopictus* (Weaver and Forrester 2015)) and can vary across vector populations (Bennett et al. 2002). When incorporating uncertainty in vector status through our supplementary model, our predictions generally agreed with that of our original model. However, the increased uncertainty did reduce the

models' ability to distinguish negative links, resulting in higher uncertainty in propensities scores (as measured by standard deviation) and a larger number of predicted vectors. Additionally, the model performs poorly when predicting on vector-virus links with trait levels not included in the training data set, as was the case when omitting yellow fever virus. Another source of uncertainty is regarding vector and virus traits. In addition to intraspecific variation in biological traits, many vectors are understudied, and common traits such as biting activity are unknown to the level of species. Additional study into the behavior and biology of less common vector species would increase the accuracy of prediction techniques such as this, and allow for a better of understanding of species' potential role as vectors.

Interestingly, our constructed flavivirus-mosquito network generally concurs with the proposed dichotomy of *Aedes* species vectoring hemorrhagic or febrile arboviruses and *Culex* species vectoring neurological or encephalitic viruses (Grard et al. 2009) (Fig. 1). However, there are several exceptions to this trend, notably West Nile virus, which is vectored by several *Aedes* species. Additionally, our model predicts several *Culex* species to be possible vectors of Zika virus. While this may initially seem contrary to the common phylogenetic pairing of vectors and viruses noted above, Zika's symptoms, like West Nile virus, are both febrile and neurological. Thus, its symptoms do not follow the conventional hemorrhagic/encephalitic division. The ability of Zika virus to be vectored by a diversity of mosquito vectors could have important public health consequences, as it may expand both the geographic range and seasonal transmission risk of Zika virus, and warrants further empirical investigation.

Considering our predictions of potential vector species and their combined ranges, species on the candidate vector list need to be validated to inform the response to Zika virus. Vector control efforts that target *Aedes* species exclusively may ultimately be unsuccessful in controlling transmission of Zika because they do not control other, unknown vectors. For example, the release of genetically modified *Ae. aegypti* to control vector density through sterile insect technique is species-specific and would not control alternative vectors (Alphey et al. 2009). Additionally, species' habitat preferences differ, and control efforts based singularly on reducing *Aedes* larval habitat will not be as successful at controlling *Cx. quinquefasciatus* populations (Rey et al. 2006).

Predicted vectors of Zika virus must be empirically tested and, if confirmed, vector control efforts would need to respond by widening their focus to control the abundance of all predicted vectors of Zika virus. Similarly, if control efforts are to include all areas at potential risk of disease transmission, public health efforts would need to expand to address regions such as the northern Midwest that fall within the range of the additional vector species predicted by our model. An understanding of the capacity of mosquito species to vector Zika virus is necessary to prepare for potential establishment of Zika virus in the United States, and we recommend that experimental work start with this list of candidate vector species.

# Acknowledgements

The authors acknowledge Pasha Feinberg for assistance with data collection.

# References


Aliota, M. T., S. A. Peinado, J. E. Osorio, and L. C. Bartholomay. 2016. *Culex pipiens* and *Aedes triseriatus* mosquito susceptibility to Zika virus [letter]. Emerging Infectious Diseases .

Alphey, L., M. Benedict, R. Bellini, G. G. Clark, D. A. Dame, M. W. Service, and S. L. Dobson. 2009. Sterile-Insect Methods for Control of Mosquito-Borne Diseases: An Analysis. Vector-Borne and Zoonotic Diseases **10**:295–311.

Barnett, H., 1960. The Incrimination of Arthropods as Vectors of Disease. Pages 341–345 *in* Proceedings of the 11th International Congress of Entomolgoy, volume 1962. Strouhal H, Beier M, editors.

Bennett, K. E., K. E. Olson, M. d. L. Muñoz, I. Fernandez-Salas, J. A. Farfan-Ale, S. Higgs, W. C. Black, and B. J. Beaty. 2002. Variation in Vector Competence for Dengue 2 Virus among 24 Collections of Aedes Aegypti from Mexico and the United States. Am J Trop Med Hyg **67**:85–92.

Boorman, J. P. T., and J. S. Porterfield. 1956. A simple technique for infection of mosquitoes with viruses transmission of Zika virus. Transactions of The Royal Society of Tropical Medicine and Hygiene **50**:238–242.

Breiman, L. 2001. Random Forests. Machine learning **45**:5–32.

Bustamante, D. M., and C. C. Lord. 2010. Sources of Error in the Estimation of Mosquito Infection Rates Used to Assess Risk of Arbovirus Transmission. American Journal of Tropical Medicine and Hygiene **82**:1172–1184.

Cao-Lormeau, V.-M., A. Blake, S. Mons, S. L. Lastère, C. R. Roche, J. Vanhomwegen, T. Dub, L. Baudouin, A. Teissier, P. Larre, A.-L. Vial, C. Decam, V. Choumet, S. K Halstead, H. J. Willison, L. Musset, J.-C. Manuguerra, P. Despres, E. Fournier, H.-P. Mallet, D. Musso, A. Fontanet, J. Neil, and F. Ghawché. 2016. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. The Lancet **387**:1–9.

Centers for Disease Control and Prevention, 2016. Estimated range of *Aedes albopictus* and *Aedes aegypti* in the United States, 2016. `http://www.cdc.gov/zika/vector/range.html`.

Chambers, T. J., and T. P. Monath, editors. 2003. The Flaviviruses: Detection, Diagnosis and Vaccine Development. Academic Press.

Chaves, L. F., L. C. Harrington, C. L. Keogh, A. M. Nguyen, and U. D. Kitron. 2010. Blood feeding patterns of mosquitoes: random or structured? Frontiers in Zoology **7**:3–11.

Chouin-Carneiro, T., A. Vega-Rua, M. Vazeille, A. Yebakima, R. Girod, D. Goindin, M. Dupont-Rouzeyrol, R. Lourenço-de Oliveira, and A.-B. Failloux. 2016. Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus. PLoS Neglected Tropical Diseases **10**:e0004543–11.

Cook, G. C., and A. Zumla. 2009. Manson's Tropical Diseases. Elsevier Health Sciences.

Cook, S., and E. C. Holmes. 2005. A multigene analysis of the phylogenetic relationships among the flaviviruses (Family: Flaviviridae) and the evolution of vector transmission. Archives of Virology **151**:309–325.

Darsie, R. F., and R. A. Ward. 2005. Identification and Geographical Distribution of the Mosquitos of North America, North of Mexico. University Press of Florida.

Day, J. F. 2001. Predicting St. Louis Encephalitis Virus Epidemics: Lesson from Recent, and Not So Recent, Outbreaks. Annual Review of Entomology **46**:111–138.

Diagne, C. T., D. Diallo, O. Faye, Y. Ba, O. Faye, A. Gaye, I. Dia, S. C. Weaver, A. A. Sall, and M. Diallo. 2015. Potential of selected Senegalese Aedes spp. mosquitoes (Diptera: Culicidae) to transmit Zika virus. BMC Infectious Diseases **15**.

Diallo, M., A. A. Sall, A. C. Moncayo, Y. Ba, Z. Fernandez, D. Ortiz, L. L. Coffey, C. Mathiot, R. B. Tesh, and S. C. Weaver. 2005*a*. Potential role of sylvatic and domestic African mosquito species in dengue emergence. American Journal of Tropical Medicine and Hygiene **73**:445–449.

Diallo, M., A. A. Sall, A. C. Moncayo, Y. Ba, Z. Fernandez, D. Ortiz, L. L. Coffey, C. Mathiot, R. B. Tesh, and S. C. Weaver. 2005*b*. Potential Role of Sylvatic and Domestic African Mosquito Species in Dengue Emergence. American Journal of Tropical Medicine and Hygiene **73**:445–449.

Duffy, M. R., T.-H. Chen, W. T. Hancock, A. M. Powers, J. L. Kool, R. S. Lanciotti, M. Pretrick, M. Marfel, S. Holzbauer, C. Dubray, L. Guillaumot, A. Griggs, M. Bel, A. J. Lambert, J. Laven, O. Kosoy, A. Panella, B. J. Biggerstaff, M. Fischer, and E. B. Hayes. 2009. Zika virus outbreak on Yap Island, federated states of Micronesia. New England Journal of Medicine **360**:2536–2543.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology **77**:802–813.

Farajollahi, A., D. M. Fonseca, L. D. Kramer, and A. M. Kilpatrick. 2011. "Bird biting" mosquitoes and human disease: A review of the role of Culex pipiens complex mosquitoes in epidemiology. Infection, Genetics and Evolution **11**:1577–1585.

Faria, N. R., R. d. S. d. S. Azevedo, M. U. G. Kraemer, R. Souza, M. S. Cunha, S. C. Hill, J. Theze, M. B. Bonsall, T. A. Bowden, I. Rissanen, I. M. Rocco, J. S. Nogueira, A. Y. Maeda, F. G. d. S. Vasami, F. L. d. L. Macedo, A. Suzuki, S. G. Rodrigues, A. C. R. Cruz, B. T. Nunes, D. B. d. A. Medeiros, D. S. G. Rodrigues, A. L. Nunes Queiroz, E. V. P. d. Silva, D. F. Henriques, E. S. Travassos da Rosa, C. S. de Oliveira, L. C. Martins, H. B. Vasconcelos, L. M. N. Casseb, D. d. B. Simith, J. P. Messina, L. Abade, J. Lourenco, L. C. J. Alcantara, M. M. d. Lima, M. Giovanetti, S. I. Hay, R. S. de Oliveira, P. d. S. Lemos, L. F. d. Oliveira, C. P. S. de Lima, S. P. da Silva, J. M. d. Vasconcelos, L. Franco, J. F. Cardoso, J. L. d. S. G. Vianez-Junior, D. Mir, G. Bello, E. Delatorre, K. Khan, M. Creatore, G. E. Coelho, W. K. de Oliveira, R. Tesh, O. G. Pybus, M. R. T. Nunes, and P. F. C. Vasconcelos. 2016. Zika virus in the Americas: Early epidemiological and genetic findings. Science **352**:345–349.

Fernandes, R. S., S. S. Campos, A. Ferreira-de Brito, R. M. d. Miranda, K. A. B. d. Silva, M. G. d. Castro, L. M. S. Raphael, P. Brasil, A.-B. Failloux, M. C. Bonaldo, and R. Loureno-de Oliveira. 2016. Culex quinquefasciatus from Rio de Janeiro Is Not Competent to Transmit the Local Zika Virus. PLOS Negl Trop Dis **10**:e0004993.

Fonseca, D. M., N. Keyghobadi, C. A. Malcolm, C. Mehmet, F. Schaffner, M. Mogi, R. C. Fleischer, and R. C. Wilkerson. 2004. Emerging Vectors in the Culex pipiens Complex. Science **303**:1535–1538.

Friedman, J. H. 2001. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics **29**:1189–1232.

GIDEON, 2016. Global Infectious Diseases and Epidemiology Network.

Gould, E. A., and T. Solomon. 2008. Pathogenic flaviviruses. The Lancet **371**:500–509.

Grard, G., G. Moureau, R. N. Charrel, E. C. Holmes, E. A. Gould, and X. de Lamballerie. 2009. Genomics and evolution of Aedes-borne flaviviruses. Journal of General Virology **91**:87–94.

Gubler, D. J., S. Nalim, R. Tan, H. Saipan, and J. Sulianti Saroso. 1979. Variation in Susceptibility to Oral Infection with Dengue Viruses among Geographic Strains of Aedes Aegypti. The American Journal of Tropical Medicine and Hygiene **28**:1045–1052.

Guo, X.-x., C.-x. Li, Y.-q. Deng, D. Xing, Q.-m. Liu, Q. Wu, A.-j. Sun, Y.-d. Dong, W.-c. Cao, C.-f. Qin, and T.-y. Zhao. 2016. Culex pipiens quinquefasciatus: a potential vector to transmit Zika virus. Emerging Microbes & Infections **5**:e102.

Haddow, A. D., A. J. Schuh, C. Y. Yasuda, M. R. Kasper, V. Heang, R. Huy, H. Guzman, R. B. Tesh, and S. C. Weaver. 2012. Genetic Characterization of Zika Virus Strains: Geographic Expansion of the Asian Lineage. PLoS Neglected Tropical Diseases **6**:e1477–7.

Haddow, A. J., M. C. Williams, J. P. Woodall, D. I. H. Simpson, and L. K. H. Goma. 1964. Twelve isolations of Zika virus from Aedes (Stegomyia) africanus (Theobald) taken in and above a Uganda forest. Bulletin of the World Health Organization **31**:57–69.

Han, B. A., J. P. Schmidt, L. W. Alexander, S. E. Bowden, D. T. S. Hayman, and J. M. Drake. 2016. Undiscovered Bat Hosts of Filoviruses. PLoS Neglected Tropical Diseases **10**:e0004815–10.

Han, B. A., J. P. Schmidt, S. E. Bowden, and J. M. Drake. 2015. Rodent reservoirs of future zoonotic diseases. Proceedings of the National Academy of Sciences pages 1–6.

Hardy, J. L., E. J. Houk, L. D. Kramer, and W. C. Reeves. 1983. Intrinsic factors affecting vector competence of mosquitoes for arboviruses. Annual Review of Entomology **28**:229–262.

Hayes, E. B., N. Komar, R. S. Nasci, S. P. Montgomery, D. R. OLeary, and G. L. Campbell. 2005. Epidemiology and Transmission Dynamics of West Nile Virus Disease. Emerging Infectious Diseases **11**:1167–1173.

Huang, Y.-J. S., V. B. Ayers, A. C. Lyons, I. Unlu, B. W. Alto, L. W. Cohnstaedt, S. Higgs, and D. L. Vanlandingham. 2016. Culex Species Mosquitoes and Zika Virus. Vector-Borne and Zoonotic Diseases .

Karabatsos, N. 1985. International Catalog of Arboviruses Including Certain Other Viruses of Vertebrates. The American Journal of Tropical Medicine and Hygiene **27**:372–440.

Komar, N., S. Langevin, S. Hinten, N. Nemeth, E. Edwards, D. Hettler, B. Davis, R. Bowen, and M. Bunning. 2003. Experimental infection of north American birds with the New York 1999 strain of West Nile virus. Emerging Infectious Diseases **9**:311–322.

Kuno, G., G.-J. J. Chang, K. R. Tsuchiya, N. Karabatsos, and C. B. Cropp. 1998. Phylogeny of the Genus Flavivirus. Journal of Virology **72**:73–83.

Ledermann, J. P., L. Guillaumot, L. Yug, S. C. Saweyog, M. Tided, P. Machieng, M. Pretrick, M. Marfel, A. Griggs, M. Bel, M. R. Duffy, W. T. Hancock, T. Ho-Chen, and A. M. Powers. 2014. Aedes hensilli as a Potential Vector of Chikungunya and Zika Viruses. PLoS Neglected Tropical Diseases **8**:e3188–9.

Liu, C., G. Newell, and M. White. 2015. On the selection of thresholds for predicting species occurrence with presence-only data. Ecology and Evolution **6**:337–348.

Liu, C., M. White, and G. Newell. 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. Journal of Biogeography **40**:778–789.

Mackenzie, J., A. D. T. Barrett, and V. Deubel. 2012. Japanese Encephalitis and West Nile Viruses. Springer Science & Business Media.

Mahy, B. W. J. 2009. The Dictionary of Virology. Academic Press.

Messina, J. P., M. U. G. Kraemer, O. J. Brady, D. M. Pigott, and F. Shearer. 2016. Mapping global environmental suitability for Zika virus. eLife pages 1–22.

Molaei, G., T. G. Andreadis, P. M. Armstrong, and M. Diuk-Wasser. 2008. Host-Feeding Patterns of Potential Mosquito Vectors in Connecticut, USA: Molecular Analysis of Bloodmeals from 23 Species of Aedes, Anopheles, Culex, Coquillettidia, Psorophora, and Uranotaenia. Journal of Medical Entomology **45**:1143–1151.

Moreno-Madriñán, M. J., and M. Turell. 2017. Factors of Concern Regarding Zika and Other Aedes Aegypti-Transmitted Viruses in the United States. J. Med. Entomol. page tjw212.

Muturi, E. J., and B. W. Alto. 2011. Larval Environmental Temperature and Insecticide Exposure Alter Aedes aegyptiCompetence for Arboviruses. Vector-Borne and Zoonotic Diseases **11**:1157–1163.

Parascandola, M. 2004. Skepticism, Statistical Methods, and the Cigarette: A Historical Analysis of a Methodological Debate. Perspect. Biol. Med. **47**:244–261.

Petersen, E., M. E. Wilson, S. Touch, B. McCloskey, P. Mwaba, M. Bates, O. Dar, F. Mattes, M. Kidd, G. Ippolito, E. I. Azhar, and A. Zumla. 2016. Unexpected and Rapid Spread of Zika Virus in The Americas - Implications for Public Health Preparedness for Mass Gatherings at the 2016 Brazil Olympic Games. International Journal of Infectious Diseases pages 1–5.

Powell, J. R., W. J. Tabachnick, J. R. Powell, and W. J. Tabachnick. 2013. History of domestication and spread of Aedes aegypti - A Review. Memrias do Instituto Oswaldo Cruz **108**:11–17.

Rey, J. R., N. Nishimura, B. Wagner, M. A. H. Braks, S. M. O'Connell, and L. P. Lounibos. 2006. Habitat segregation of mosquito arbovirus vectors in south Florida. Journal of Medical Entomology **43**:1134–1141.

Ridgeway, G., 2015. gbm: Generalized Boosted Regression Models.

Russell, R. C., D. Otranto, and R. L. Wall. 2013. The Encyclopedia of Medical and Veterinary Entomology. CABI.

Schuler-Faccini, L., E. M. Ribeiro, I. M. L. Feitosa, D. D. G. Horovitz, D. P. Cavalcanti, A. Pessoa, M. J. R. Doriqui, J. I. Neri, J. M. d. P. Neto, H. Y. C. Wanderley, M. Cernach, A. S. El-Husny, M. V. S. Pone, C. L. C. Serao, M. T. V. Sanseverino, and Brazilian Medical Genetics Society–Zika Embryopathy Task Force. 2016. Possible Association Between Zika Virus Infection and Microcephaly — Brazil, 2015. MMWR. Morbidity and Mortality Weekly Report **65**:1–4.

Smith, G. C. S., S. R. Seaman, A. M. Wood, P. Royston, and I. R. White. 2014. Correcting for Optimistic Prediction in Small Data Sets. American Journal of Epidemiology **180**:318–324.

Thomson, M. C., F. J. Doblas-Reyes, S. J. Mason, R. Hagedorn, S. J. Connor, T. Phindela, A. P. Morse, and T. N. Palmer. 2006. Malaria Early Warnings Based on Seasonal Climate Forecasts from Multi-Model Ensembles. Nature **439**:576–9.

Turell, M. J., D. J. Dohm, M. R. Sardelis, M. L. O'guinn, T. G. Andreadis, and J. A. Blow. 2005. An Update on the Potential of North American Mosquitoes (Diptera: Culicidae) to Transmit West Nile Virus. Journal of Medical Entomology **42**:57–62.

Turrell, M. J., D. J. Dohm, M. R. Sardelis, M. L. OGuinn, T. G. Andreadis, and J. A. Blow. 2005. An Update on the Potential of North American Mosquitoes (Diptera: Culicidae) to Transmit West Nile Virus. Journal of Medical Entomology **42**:57–62.

Van Regenmortel, M. H. V., C. M. Fauquet, and D. H. L. Bishop. 2000. Virus taxonomy : classification and nomenclature of viruses : seventh report of the International Committee on Taxonomy of Viruses. San Diego : Academic Press, c2000.

Walter Reed Biosystematics Unit, W. D., Smithsonian Institution, 2016. Walter Reed Biosystematics Unit Systematic catalog of Culicidae. `http://www.mosquitocatalog.org/`. Accessed: 2016-05-02.

Weaver, S. C., 2005. Host range, amplification and arboviral disease emergence. Pages 33–44 *in*

P. D. C. J. Peters and P. C. H. Calisher, editors. Infectious Diseases from Nature: Mechanisms of Viral Emergence and Persistence. Springer Vienna.

Weaver, S. C., F. Costa, M. A. Garcia-Blanco, A. I. Ko, G. S. Ribeiro, G. Saade, P.-Y. Shi, and N. Vasilakis. 2016. Zika virus: History, emergence, biology, and prospects for control. Antiviral Research **130**:69–80.

Weaver, S. C., and N. L. Forrester. 2015. Chikungunya: Evolutionary History and Recent Epidemic Spread. Antiviral Res. **120**:32–39.

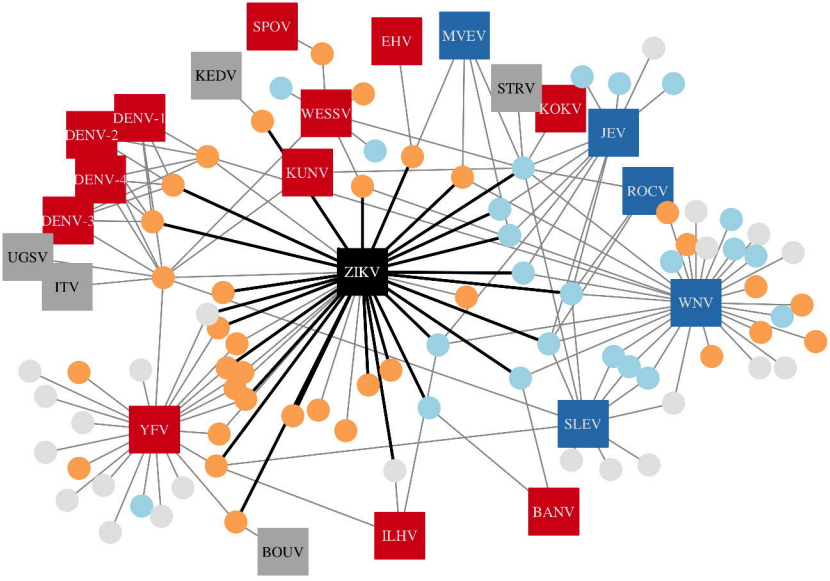
Table 1: **Predicted vectors of Zika virus, as reported by our model.** Mosquito species endemic to the continental United States are bolded. A species is defined as a known vector of Zika virus if a full transmission cycle (see main text) has been observed.

| **Species** | **GBM Prediction ± *SD*** | **Known Vector?** |
|---|---|---|
| ***Aedes aegypti*** | $0.81 \pm 0.12$ | Yes |
| ***Ae. albopictus*** | $0.54 \pm 0.14$ | Yes |
| ***Culex quinquefasciatus*** | $0.38 \pm 0.14$ | No |
| *Ae. polynesiensis* | $0.36 \pm 0.13$ | No |
| *Ae. scutellaris* | $0.33 \pm 0.13$ | No |
| *Ae. africanus* | $0.32 \pm 0.11$ | No |
| *Ae. furcifer* | $0.31 \pm 0.16$ | Yes |
| *Ae. vittatus* | $0.30 \pm 0.20$ | Yes |
| *Ae. taylori* | $0.30 \pm 0.16$ | Yes |
| *Ae. luteocephalus* | $0.25 \pm 0.12$ | Yes |
| *Ae. tarsalis* | $0.18 \pm 0.11$ | Yes |
| *Ae. metallicus* | $0.16 \pm 0.08$ | No |
| *Ae. minutus* | $0.16 \pm 0.09$ | No |
| *Ae. opok* | $0.14 \pm 0.06$ | No |
| *Ae. bromeliae* | $0.11 \pm 0.06$ | No |
| *Ae. scapularis* | $0.10 \pm 0.04$ | No |
| ***Cx. pipiens*** | $0.10 \pm 0.04$ | No |
| *Ae. hensilli* | $0.10 \pm 0.06$ | Yes |
| *Ae. vigilax* | $0.10 \pm 0.05$ | No |
| *Cx. annulirostrix* | $0.08 \pm 0.03$ | No |
| ***Psorophora ferox*** | $0.08 \pm 0.05$ | No |
| *Cx. rubinotus* | $0.08 \pm 0.07$ | No |
| ***Cx. tarsalis*** | $0.08 \pm 0.03$ | No |
| *Ae. occidentalis* | $0.08 \pm 0.05$ | No |
| *Ae. flavicolis* | $0.07 \pm 0.04$ | No |
| *Ae. serratus* | $0.07 \pm 0.04$ | No |
| *Cx. p. molestus* | $0.07 \pm 0.04$ | No |
| ***Ae. vexans*** | $0.06 \pm 0.04$ | No |
| *Cx. neavei* | $0.06 \pm 0.02$ | No |
| *Runchomyia frontosa* | $0.06 \pm 0.04$ | No |
| *Ae. neoafricanus* | $0.06 \pm 0.03$ | No |
| *Ae. chemulpoensis* | $0.06 \pm 0.03$ | No |
| *Cx. vishnui* | $0.05 \pm 0.01$ | No |
| *Cx. tritaeniorhynchus* | $0.05 \pm 0.01$ | No |
| *Ae. fowleri* | $0.04 \pm 0.03$ | Yes |

Figure 1: **A network diagram of mosquito vectors (circles) and their flavivirus pairs (rectangles).** The *Culex* mosquitoes (light blue) and primarily encephalitic viruses (blue) are more clustered than the *Aedes* (orange) and hemmorhagic viruses (red). Notably, West Nile Virus is vectored by both *Aedes* and *Culex* species. Predicted vectors of Zika are shown by bolded links in black. The inset shows predicted vectors of Zika and species names, ordered by the model's propensity scores. Included flaviviruses are Banzi virus (BANV), Bouboui virus (BOUV), dengue virus strains 1, 2, 3 & 4 (DENV-1,2,3,4), Edge Hill virus (EHV), Ilheus virus (ILHV), Israel turkey meningoencephalomyelitis virus (ITV), Japanese encephalitis virus (JEV), Kedougou virus (KEDV), Kokobera virus (KOKV), Kunjin virus (KUNV), Murray Valley encephalitis virus (MVEV), Rocio virus (ROCV), St. Louis encephalitis virus (SLEV), Spondwendi virus (SPOV), Stratford virus (STRV), Uganda S virus (UGSV), Wesselsbron virus (WESSV), West Nile Virus (WNV), yellow fever virus (YFV), and Zika virus (ZIKV).

Figure 2: **Variable importance by permutation, averaged over 25 models.** Because some categorical variables were treated as binary by our model (i.e. continental range), the relative importance of each binary variable was summed to result in the overall importance of the categorical variable. Mosquito and virus traits are shown in blue and maroon, respectively. Error bars represent the standard error from 25 models.

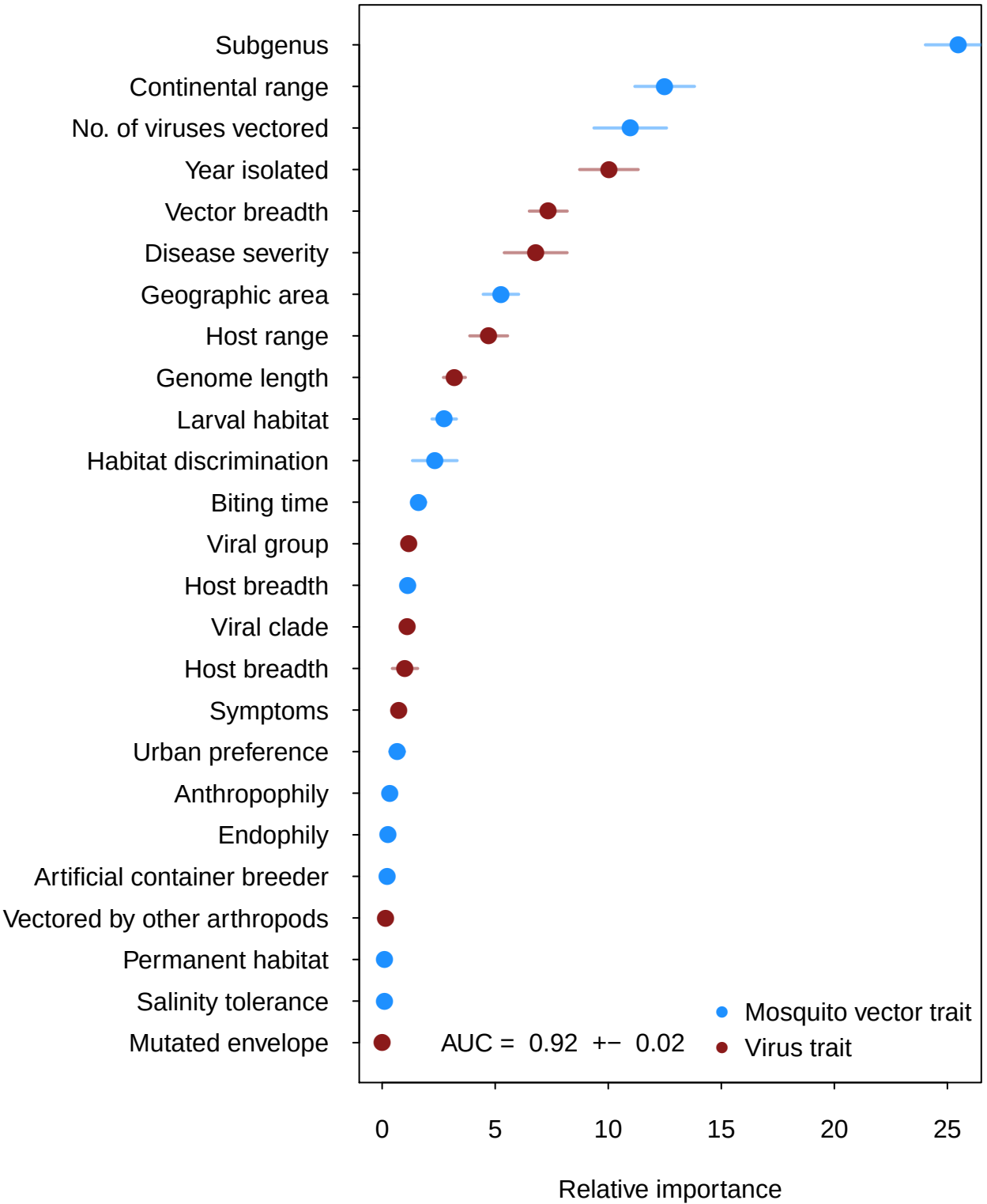
Figure 3: **Distribution maps of predicted vectors of Zika virus in the continental US.** Maps of *Aedes* species are based on Centers for Disease Control and Prevention (2016). All other species' distributions are georectified maps from Darsie and Ward (2005).



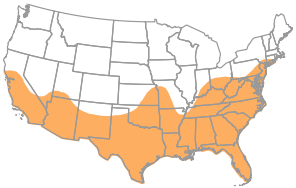

Aedes species
Culex species
Other Mosq.
Hemmorhagic
Encephalitic
Symptoms Unknown
Known Link
Predicted Link

**Predicted Zika Links**

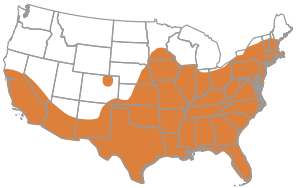
- *Aedes aegypti*
- *Aedes albopictus*
- *Culex quinquefasciatus*
- *Aedes polynesiensis*
- *Aedes scutellaris*
- *Aedes africanus*
- *Aedes furcifer*
- *Aedes vittatus*
- *Aedes taylori*
- *Aedes luteocephalus*
- *Aedes tarsalis*
- *Aedes metallicus*
- *Aedes minutus*
- *Aedes opok*
- *Aedes bromeliae*
- *Aedes scapularis*
- *Culex pipiens*
- *Aedes hensilli*
- *Aedes vigilax*
- *Culex annulirostris*
- *Psorophora ferox*
- *Culex rubinotus*
- *Culex tarsalis*
- *Aedes occidentalis*
- *Aedes flavicollis*
- *Aedes serratus*
- *Culex p. molestus*
- *Aedes vexans*
- *Culex neavei*
- *Runchomyia frontosa*
- *Aedes neoafricanus*
- *Aedes chemulpoensis*
- *Culex vishnui*
- *Culex tritaeniorhynchus*
- *Aedes fowleri*


Subgenus
Continental range
No. of viruses vectored
Year isolated
Vector breadth
Disease severity
Geographic area
Host range
Genome length
Larval habitat
Habitat discrimination
Biting time
Viral group
Host breadth
Viral clade
Host breadth
Symptoms
Urban preference
Anthropophily
Endophily
Artificial container breeder
Vectored by other arthropods
Permanent habitat
Salinity tolerance
Mutated envelope
AUC = 0.92 +− 0.02
Mosquito vector trait
Virus trait
0
5
10
15
20
25
Relative importance

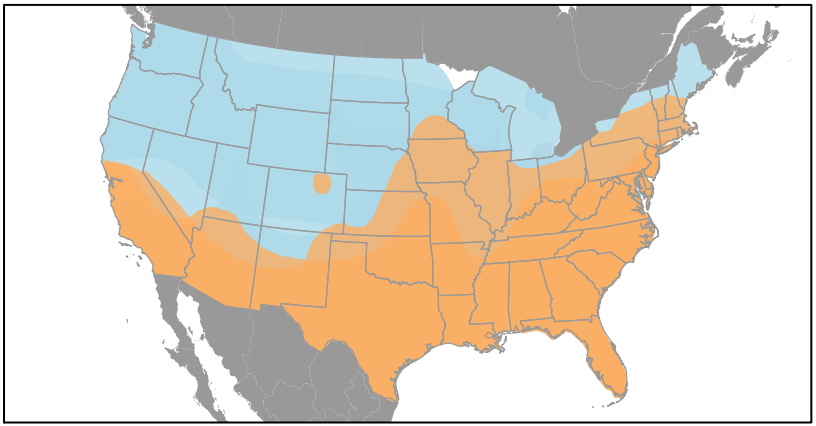
*Ae. aegypti*



*Ae. albopictus*

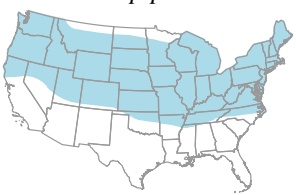


*Ps. ferox*

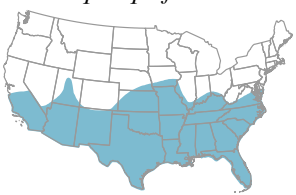
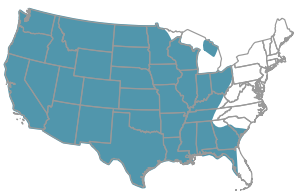


Distribution maps of predicted vectors of ZIKV in the continental US. *Aedes* species are shown in orange, *Culex* in blue and other genera in gray. Inset map represents overlay of all predicted vectors. The range of *Ae. vexans* encompasses the entire continental US and is not shown for clarity.

*Cx. pipiens*



*Cx. quinquefasciatus*



*Cx. tarsalis*