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], John M. Drake[1,5]

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

*Corresponding Author: mvevans@uga.edu

Abstract

Zika is an emerging, mosquito-borne virus recently introduced to the Americas, whose rapid spread is unprecedented and of great public health concern. Knowledge about transmission – which depends on the presence of competent vectors – remains incomplete, especially concerning potential transmission in geographic areas in which it has not yet been introduced. To identify presently unknown vectors of Zika, we developed a data-driven model linking candidate vector species and the Zika virus via vector-virus trait combinations that confer a propensity toward associations in the larger ecological network connecting flaviviruses and their mosquito vectors. Our model predicts that thirty-five species may be able to transmit the virus, twenty-six of which are not currently known vectors of Zika virus. Seven of these species are found in the continental United States, including *Culex quinquefasciatus* and *Cx. pipiens*, both of which are common mosquito pests and vectors of West Nile Virus. Because the range of these predicted species is wider than that of known vectors *Aedes aeygpti* and *Ae. albopictus*, we reason that a larger geographic area is at risk for autochthonous transmission of Zika virus than reported by maps constructed from the ranges of only the two *Aedes* species. Consequently, the reach of existing vector control activities and public health campaigns may need to be expanded.

Introduction

In 2014, Zika virus was introduced into Brazil and Haiti, from where it rapidly spread throughout the Americas. By June 2016, over 300,000 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 of 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-Barre's 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 field populations (especially from *Ae. aegypti* (Haddow et al. 2012)), and association with related arboviruses (e.g. dengue fever virus, chikungunya virus). Predictions of the potential geographic range of Zika virus in the Americas, 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 populations 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 (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 a suitable 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. Finally, we constructed new maps of the potential Zika virus distribution in North America using this larger list of potentially competent species.

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, transmission from infected host to vector to susceptible host) 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 assume 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 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 Supplement I for analysis and results).

Fifteen mosquito traits (Supplement II, Table 1) and twelve virus traits (Supplement 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 Biosystematics Unit (2016). A species' continental extent was recorded as a binary value of its presence by continent. A species' host breadth 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, 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 Supp. 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 breadth (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 breadth 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 Supplement 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 and a learning rate of 0.001. 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. 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).

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 our trained model on. 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 designations between *Flaviviridae*) (Fig. 1).

Our ensemble of BRT models trained on common virus and vector traits predicted mosquito vector-virus pairs in the test dataset with high accuracy ($AUC = 0.92 \pm 0.02$). The most important variable in predicting a vector-virus pair was the subgenus of the mosquito species, followed by the continental range of the mosquito species, and the number of viruses vectored by a mosquito species (Table 2). Unsurprisingly, this suggests that 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 and geographic range.

When applied to the 180 potential mosquito-Zika pairs, the model predicted thirty-five vectors to be ranked above the threshold, 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.

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 historically, Zika virus infection has 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. Our findings suggest that many additional mosquito species may be competent vectors of Zika virus, translating to a larger geographic area and greater human population at risk of infection.

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 × 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 species 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.

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 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 human populations in both rural and urban areas may be at greater risk of Zika transmission than previously suspected due to the presence of alternative vector species.

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). 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.

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 divide of hemorrhagic vs. encephalitic. 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.

Considering our predictions of potential vector species and the wider geographic area at possible risk for transmission, the current response to Zika virus in the United States appears limited in scope. 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. *Cx. quinquefasciatus*, for example, is a crepuscular biter (Farajollahi et al. 2011), while *Ae. aegypti* prefers to bite during the day (Yasuno and Tonn 1970). Additionally, their 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). If additional Zika virus vectors are 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 mid-West that fall within the range of the additional vector species predicted by our model. An expansion of public health efforts to recognize the potential threat of these predicted vectors is vital to preventing a public health emergency following the potential establishment of Zika virus in the United States.

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 .

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. 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.

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.

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., 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.

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.

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.

Yasuno, M., and R. J. Tonn. 1970. A study of biting habits of Aedes aegypti in Bangkok, Thailand. Bulletin of the World Health Organization **43**:319–325.

426 # Tables

Table 1: Predicted vectors of Zika, as reported by our model. Mosquito species endemic to the continental United States are bolded.

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

Figures

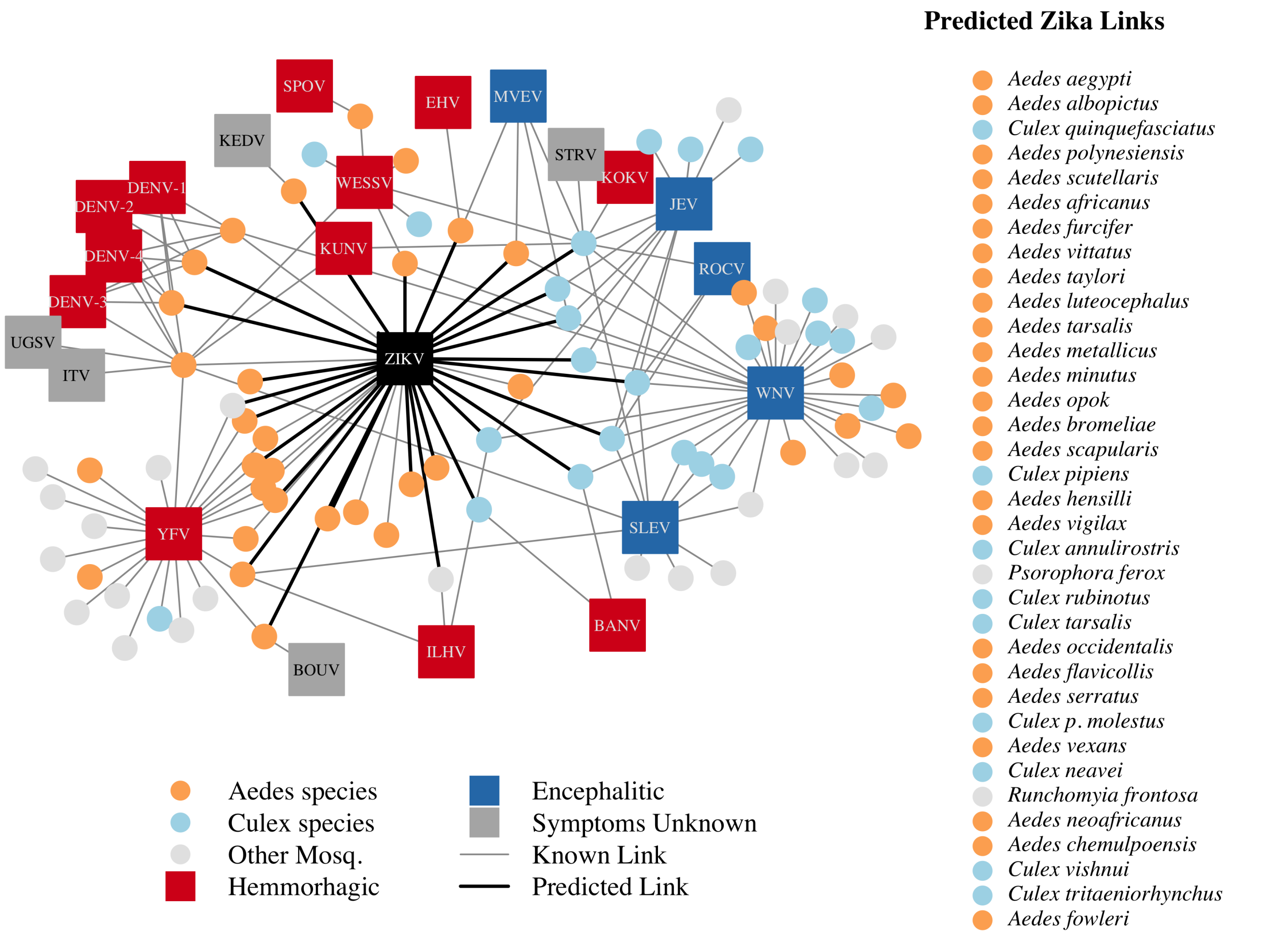

Figure 1: **A network diagram of mosquito vectors (circles) and their flavivirus pairs (rectangles).** The *Culex* mosquitoes (light blue) and primarily encephalitic viruses (blue) are more clustered than the *Aedes* (orange) and hemmorhagic viruses (red). Notably, West Nile Virus is vectored by both *Aedes* and *Culex* species. Predicted vectors of Zika are shown by bolded links in black. The inset pictures the predicted vectors of Zika and their species name, 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).

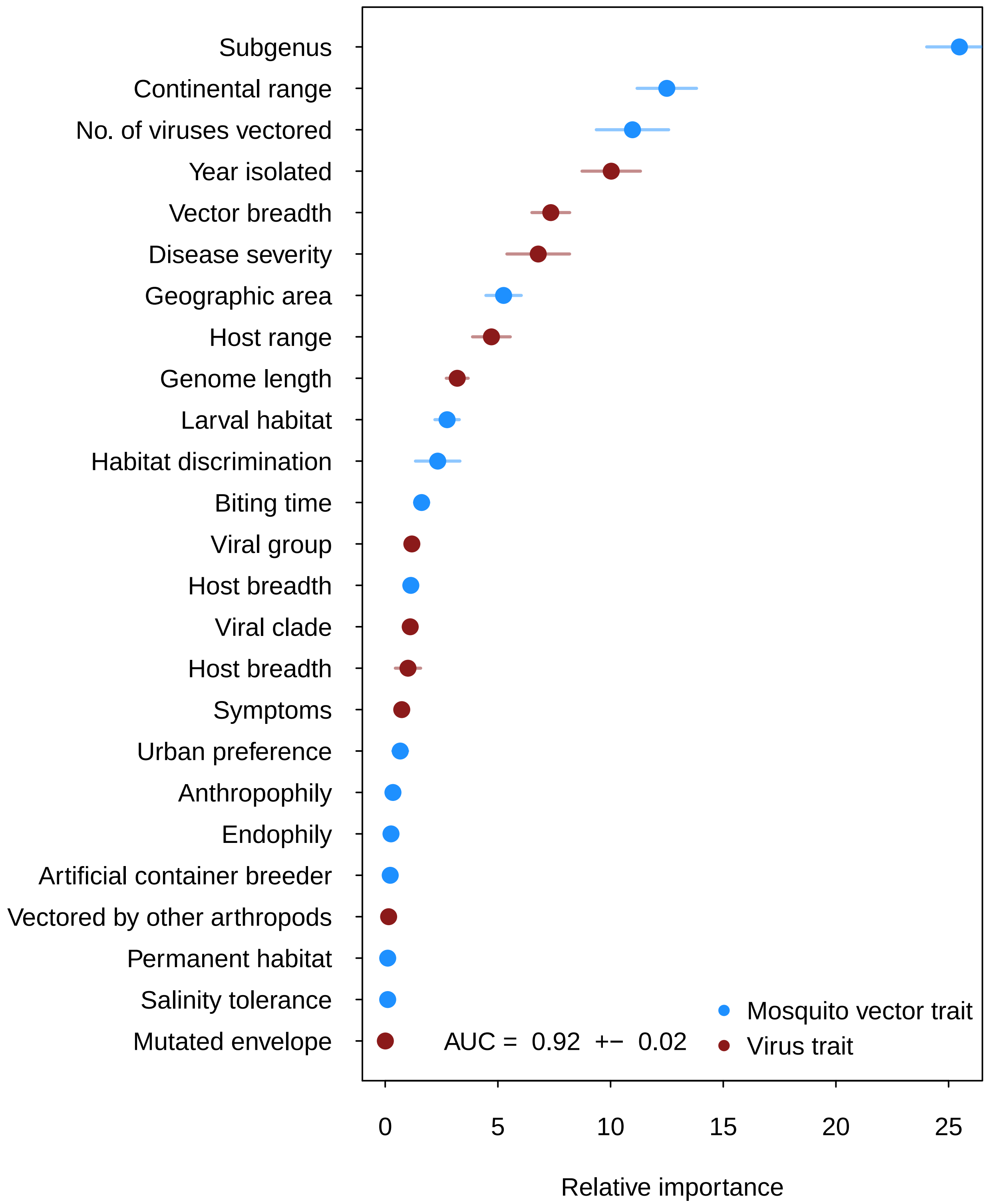

Figure 2: **Variable importance by permutation, averaged over 25 models.** Because some categorical variables were treated as binary by our model (i.e. continental range), the relative importance of each binary variable was summed to result in the overall importance of the categorical variable. Error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

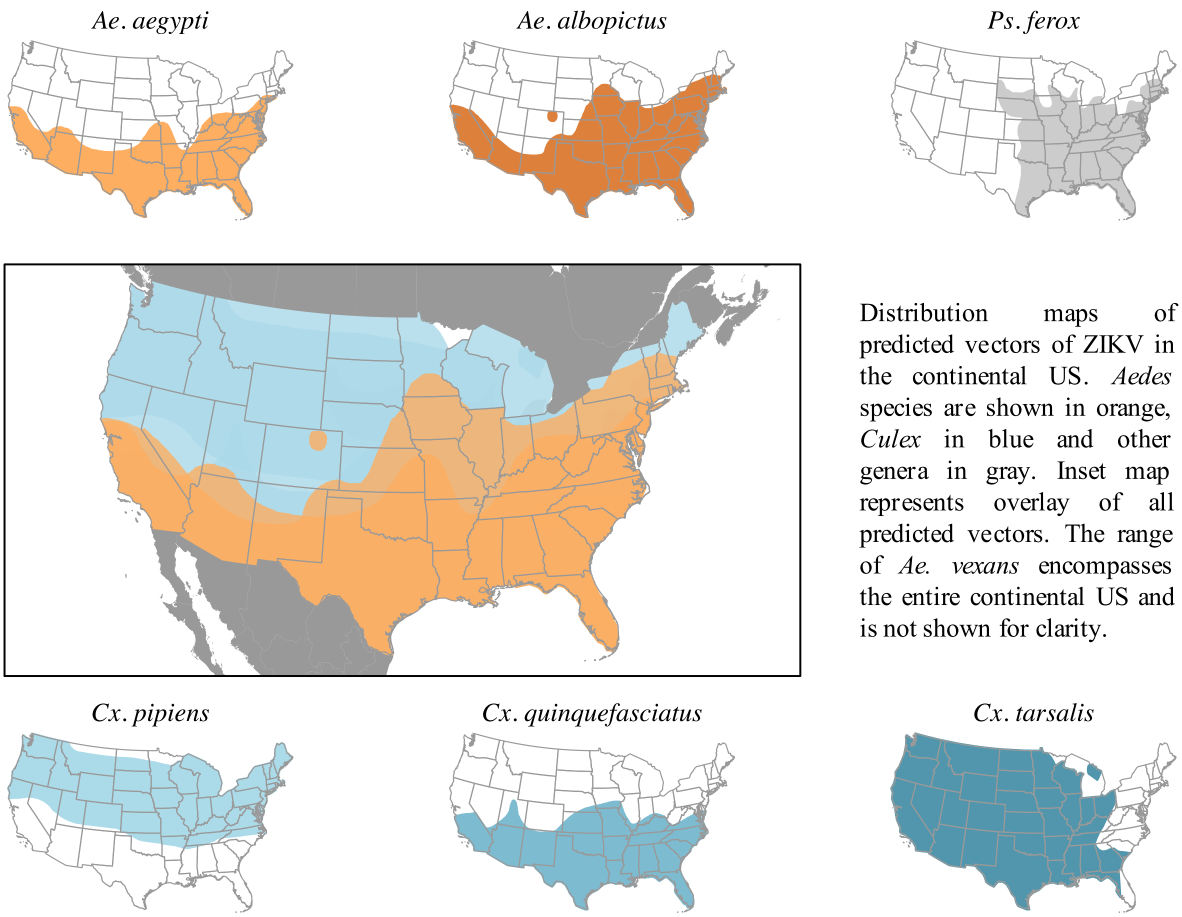

Figure 3: **Distribution maps of predicted vectors of Zika virus in the continental US.** Maps of *Aedes* species are based on Centers for Disease Control and Prevention (2016). All other species' distributions are adapted from Darsie and Ward (2005).