# An ecological assessment of the pandemic threat of Zika virus

Colin J. Carlson[1*†], Eric R. Dougherty[1†], Wayne Getz[1,2]

**Affiliations:**

[1]Department of Environmental Science, Policy, and Management, University of California, Berkeley, 130 Mulford Hall, Berkeley, CA 94720, USA

[2] School of Mathematical Sciences, University of KwaZulu-Natal, South Africa.

*Corresponding author: cjcarlson@berkeley.edu.

†These authors contributed equally to this work.

**Keywords**

Zika viral disease (ZIKV), flavivirus, climate change, ecological niche modeling, species distribution modeling

**Summary**

The current outbreak of Zika virus poses a threat of unknown magnitude to human health[1]. While the range of the virus has been cataloged growing slowly over the last 50 years, the recent explosive expansion in the Americas indicates that the full potential distribution of Zika remains uncertain[2-4]. Moreover, most current epidemiology relies on its similarities to dengue fever, a phylogenetically closely related disease of unknown similarity in spatial range or ecological niche[5,6]. Here we compile the first spatially explicit global occurrence dataset from Zika viral surveillance and serological surveys, and construct ecological niche models to test basic hypotheses about its spread and potential establishment. The hypothesis that the outbreak of cases in Mexico and North America are anomalous and outside the ecological niche of the disease, and may be linked to El Nino or similar climatic events, remains plausible at this time[7]. Comparison of the Zika niche against the known distribution of dengue fever suggests that Zika is more constrained by the seasonality of precipitation and diurnal temperature fluctuations, likely confining the disease to the tropics outside of pandemic scenarios. Projecting the range of the diseases in conjunction with vector species (*Aedes africanus*, *Ae. aegypti*, and *Ae. albopictus*) that transmit the pathogens, under climate change, suggests that Zika has potential for northward expansion; but, based on current knowledge, Zika is unlikely to fill the full range its vectors occupy. With recent sexual transmission of the virus known to have occurred in the United States, we caution that our results only apply to the vector-borne aspect of the disease, and while the threat of a mosquito-carried Zika pandemic may be overstated in the media, other transmission modes of the virus may emerge and facilitate naturalization worldwide.

44 **Main Text**

45 Following a twenty-fold upsurge in microcephalic newborns in Brazil tentatively linked to Zika
46 virus (ZIKV), the World Health Organization has declared an international health emergency[1].
47 Despite being profiled for the first time in 1947[8], Zika remains poorly characterized at a global
48 scale. Thus, the present pandemic expansion in the Americas poses a threat of currently unknown
49 magnitude. Closely related to dengue fever, Zika conventionally presents as a mild infection,
50 with 80% of cases estimated to be asymptomatic[9]. The cryptic nature of infection has resulted in
51 sporadic documentation of the disease and rarely includes spatially explicit information beyond
52 the regional scale[1-4]. This greatly limits the confidence with which statistical inferences can be
53 made about the expansion of the virus. With an estimated 440,000-1,300,000 cases in Brazil in
54 2015[9], and continuing emergence of new cases in Central America and, most recently, the United
55 States, assessing the full pandemic potential of the virus is an urgent task with major
56 ramifications for global health policy.

57

58 Current evidence portrays the global spread of ZIKV as a basic diffusion process facilitated by
59 human and mosquito movement, a hypothesis supported by the frequency of infected traveler
60 case studies in the Zika literature[10-13]. Tracing phylogenetic and epidemiological data has
61 revealed the expansion of ZIKV has occurred in a stepwise process through the South Pacific,
62 moving the disease from Southeast Asia into French Polynesia and the Philippines, and
63 subsequently to Easter Island[1-4]. ZIKV is conjectured to have dispersed into South America as
64 recently as three years ago from the last of those locations, and the virus is not presumed to be at
65 a biogeographic equilibrium in the Americas. With cases in the ongoing outbreak in Colombia,
66 El Salvador, Guatemala, Paraguay, and Venezuela, and by November of last year, as far north as

67 Mexico, Puerto Rico, and the continental United States, the full potential distribution of the
68 disease remains unknown. Moreover, alternative explanations for the disease's expansion remain
69 unconsidered; most notably, the role of climate change in Zika's expansion is uncertain[7].
70
71 We present three competing hypotheses that describe the path of expansion that Zika could take,
72 based on evaluations of the ecological niche of the virus within and outside of its vectors. If the
73 Zika niche is indistinguishable from that of its *Aedes* vectors (as is essentially the case for
74 dengue fever[14]), future range expansions should match mosquito ranges. On the other hand, if
75 Zika has a transmission niche that is constrained by climatic factors within the ranges of its
76 mosquito vectors, its range may be much more limited—with, as we show below, possible
77 confinement to the tropics. In this case, the expansion of Zika into North America represents one
78 of two hypothesized processes: a steady range expansion driven by climatic shifts, or an
79 anomalous event driven by human dispersal or extreme weather events. To test these hypotheses,
80 we present the first spatially explicit database of Zika occurrences from the literature and an
81 ecological niche model[15] using that data to map the potential distribution of the virus.
82
83 Our dataset includes 64 of the known occurrences of the disease – a combination of clinical
84 cases and seropositivity surveys in humans and mosquitoes. Of these, 60 points from outside the
85 current outbreak are used in our model to determine the expected distribution in the Americas
86 based on the niche in areas where the virus is established (rather than potentially transient).
87 Spanning seven decades, these data have not previously been compiled nor explicitly geo-
88 referenced, and emergency modeling efforts for diseases of special concern often work from
89 limited data. Previously published sensitivity analyses unequivocally suggest that accuracy of the

90 modeling methods we employ plateaus at or near 50 points, justifying the use of a dataset of this
91 size[16-18]. Ensemble modeling also vastly improves the predictive power with datasets of this sort
92 (Extended Data Fig. 1-5), and reduces the associated error. Our final model combines seven
93 methods with a variable set chosen from bioclimatic variables and a vegetation index to
94 minimize predictor covariance. The ensemble model performs very well (AUC = 0.993; Fig. 1),
95 and strongly matches most occurrences including the hotspots of Brazilian microcephaly. It also
96 predicts additional regions where Zika is as yet unrecorded, but where further inquiry may be
97 desired (in particular, Southern Sudan and the northern coast of Australia). Our model indicates
98 that certain occurrences, like the 1954 report from Egypt and almost all North American cases,
99 are likely outside the stable transmission niche[19] (i.e., persistent over time) of the virus.
100 Moreover, we note that visual presentation of cases at the country level may make the range of
101 the virus appear far larger than our models suggest (see Fig. 1). Projecting niche models to the
102 year 2050 suggests that expansion of Zika's niche outside the tropics is an unlikely scenario,
103 independent of vector availability (Fig. 2d). However, significant westward expansion in South
104 America and eastward expansion in Africa implies that Zika may continue to emerge in the
105 tropics.
106
107 Given the public health crisis posed by Zika, and the potential costs associated with
108 underpredicting the extent of the current outbreak, we pay special attention to evaluating the
109 sensitivity of our models to variations in our preliminary dataset. Geographical data on cases in
110 the Americas are lacking, and the routes and drivers of transmission involved in that outbreak are
111 uncertain, preventing cross validation of models of the current outbreak with our Old World
112 model. But, in light evidence that African and Asian strains of the virus may be ecologically

113 distinct, we present models trained on each continent and projected globally, as a basic
114 sensitivity analysis (Extended Data Fig. 7). The two models cross-validate poorly; driven by both
115 the 50% reduction in sample size and the higher degree of aggregation of Asian occurrences, the
116 two projected distributions are severely different. But despite the major over-prediction of the
117 Asian model and the overfitting of the African model, we emphasize that neither extreme
118 scenario predicts any greater range in the Americas. Moreover, despite low transferability
119 between continents, both sub-models are well matched by our aggregated model in their native
120 range, further supporting the accuracy and predictive power of our global projection.
121
122 Recently published work by Bogoch *et al.*[6] uses an ecological niche model for dengue as a proxy
123 for the potential full distribution of ZIKV in the Americas, presenting findings in terms of
124 potential seasonal vs. full-year transmission zones. While that approach has been effectively
125 validated for dengue transmission in mosquitoes, using a model of one disease to represent the
126 potential distribution of another emerging pathogen is only a placeholder, and is particularly
127 concerning given the lack of evidence in our models that ZIKV and dengue have a similar niche
128 breadth[20]. To evaluate the similarity of Zika and dengue, we built another niche model using the
129 dengue occurrence database compiled by Messina *et al.*[21]. Comparing the two niche models
130 reveals that the two niches are significantly different (Schoener's $D = 0.256$; $p < 0.02$; Extended
131 Data Fig. 6). While the two occupy a similar region of global climate space, Zika is more strictly
132 tropical than dengue, occupying regions with higher diurnal temperature fluctuations and
133 seasonality of precipitation (Fig. 2a). Moreover, our future projections for dengue (which
134 strongly agree with previously published ones[22]) show an expansion out of the tropics that is not

135 shared with Zika (Fig. 2, 3). These results call into question the applicability of dengue niche
136 models used to project a significant future range for Zika in North America[6].
137
138 Given the ecological nonequivalence of Zika and dengue, and the occurrence of Zika cases
139 outside our predicted suitable range for the virus, the 2016 Zika outbreak may be in ephemeral
140 rather than stable parts of the Zika transmission niche due to anomalous climatic conditions.
141 Specifically, El Niño Southern Oscillation (ENSO) events drive outbreaks of dengue in the
142 Americas and in Southeast Asia[23], and we conjecture that the 2016 ENSO event may be
143 responsible in large part for the severity of the ZIKV outbreak in North and Central America, a
144 hypothesis also raised by Paz *et al.*[7] in response to Bogoch *et al.*[6]. While wind-dispersed
145 mosquitoes carrying infections can be responsible for the introduction of diseases to new
146 regions[24], reported cases in the United States have all been contracted sexually or while traveling
147 abroad to regions with endemic outbreaks, further supporting the tropical constraint hypothesis.
148 However, the rapid expansion during the current outbreak beyond the boundaries of the stable
149 transmission niche indicates that regions outside our modeled range may support transmission
150 during anomalous periods of climatic flux, but will not necessarily enable naturalization of the
151 pathogen in the future. This highlights one of the most important limitations of this work, as
152 ecological niche models relate occurrence to climate, while disease drivers may operate at the
153 temporal scale of weather.
154
155 While the potential for rare, weather-driven outbreaks should not be overlooked, our models
156 imply that it is premature to expect Zika naturalization as an eventuality in North America.
157 Without more definitive information on the basic biology of Zika, however, the confidence with

158 which niche models can forecast pandemics is limited. In particular, we draw attention to recent
159 evidence suggesting Zika persistence may depend on wildlife reservoirs in addition to human
160 hosts and mosquitoes. Primates have been suggested as the primary candidate clade, because the
161 Zika flavivirus was first isolated in a rhesus macaque in the Zika Forest in Uganda. But as rhesus
162 macaques do not occur on the African continent, and were captive there for inoculation
163 experiments, the primate reservoir hypothesis remains unsupported. A 2015 case of an Australian
164 presumed to have contracted Zika from a monkey bite while traveling in Indonesia, however,
165 indicates that primates may transmit the virus directly[12]. Additionally, antibodies against Zika
166 have been observed in several rodent and livestock species in Pakistan[25], as well as several large
167 mammal species, including orangutans, zebras, and elephants[26]. The potential for any North
168 American wildlife species to play host to Zika is, at the present time, entirely unknown, and the
169 infection of alternate hosts could potentially support new regions of stable transmission.
170
171 From the limited data in existence, we conclude that the global threat of a specifically vector-
172 borne Zika pandemic, though devastating, may be limited to the tropics. However, sexual
173 transmission of Zika infections may still facilitate a significant outbreak in the United States and
174 other previously unsuitable regions, particular under evolutionary processes that select for the
175 most directly transmissible strains of pathogens[27]. A case of sexual transmission in Texas has
176 been suspected in the 2016 outbreak, and two previous reports of likely sexual transmission of
177 ZIKV originate from 2011 and 2015[3,28], though these seem to have been overlooked in most
178 press coverage, which has presented the case of sexual transmission in Texas as a novel facet to
179 the disease. Even if the Zika cases in the United States represent a rare spillover outside of the
180 mosquito-borne viral niche, sexual transmission could create a new, unbounded niche in which

181 the virus could spread. We draw attention to the potential parallels with simian and human
182 immunodeficiency virus (SIV/HIV), for which a sexually transmitted pandemic has
183 overshadowed the zoonotic origin of the disease[29]. With Zika's asymptomatic presentation and
184 the overall confusion surrounding its basic biology and transmission modes, we caution that its
185 potential for a sexually-transmitted global pandemic cannot be overlooked in the coming months.
186
187

188 **Supplementary Information**

189 Supplementary Information is linked to the online version of the paper at
190 www.nature.com/nature.
191

192 **Acknowledgements**

193 C.J.C. thanks Fausto Bustos for feedback on initial ideas presented in the paper, and Kevin
194 Burgio for extensive methodological support, training and mentorship. We also thank two
195 anonymous reviewers for their aid in strengthening the manuscript.
196

197 **Author Contributions**

198 C.J.C. and E.R.D. collected the data, ran the models and wrote the first draft. All authors edited
199 and approved the final text submitted for review.
200

## Author Information

Data presented in the paper are available in Table S1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.J.C. (cjcarlson@berkeley.edu)

## References

[1] Attar, N. Zika virus circulates in new regions. *Nature Rev. Microbiol.* **14**, 62 (2016).

[2] Musso, D., Nilles, E.J. & Cao-Lormeau, V.M. Rapid spread of emerging Zika virus in the Pacific Area. *Clin. Microbiol. Infec.* **20**, O595-O596 (2014).

[3] Musso, D., Cao-Lormeau, V.M., & Gubler, D.J. Zika virus: following the path of dengue and chikungunya? *Lancet* **386**, 243-244 (2015).

[4] Gatherer, D. Zika virus: a previously slow pandemic spreads rapidly through the Americas? *J. Gen. Virol.,* in press.

[5] Kelser, E.A. Meet dengue's cousin, Zika. *Microbes Infect.,* in press

[6] Bogoch, I.I., *et al.* Anticipating the international spread of Zika virus from Brazil. *Lancet* **387**, 335-336 (2016).

[7] Paz, S. & Semenza, J.C. El Niño and climate change–contributing factors in the dispersal of Zika virus in the Americas? *Lancet,* in press.

[8] Dick, G.W.A., Kitchen, S.F. & Haddow, A.J. Zika virus (I). Isolations and serological specificity. *Trans. R. Soc. Trop. Med. Hyg.* **46**, 509-520 (1952).

[9] Hennessey, M. Zika virus spreads to new areas—region of the Americas, May 2015-January 2016. *CDC Morbidity and Mortality Report* **65**, 55-58 (2016).

[10] Summers, D.J., Acosta, R.W. & Acosta, A. M. Zika virus in an American recreational traveler. *J. Travel Med.* **22**, 338-340 (2015).

[11] Fonseca, K. Case report: first case of Zika virus infection in a returning Canadian traveler. *Am. J. Trop. Med. Hyg.* **91**, 1035-1038 (2014).

[12] Kwong, J.C., Druce, J.D. & Leder, K. Case report: Zika virus infection acquired during brief travel to Indonesia. *Am. J. Trop. Med. Hyg.* **89,** 516-517 (2013)

229 [13] Zammarchi, L. Zika virus infection in a traveler returning to Europe from Brazil, March
230 2015. *Euro. Surveill.* **20**, pii-21153 (2015).
231 [14] Lambrechts, L., Scott, T. W. & Gubler, D. J. Consequences of the expanding global
232 distribution of Aedes albopictus for dengue virus transmission. *PLoS Negl. Trop. Dis.* **4**, e646
233 (2010).
234 [15] Peterson, A.T. Ecological niche modeling and spatial patterns of disease transmission.
235 *Emerg. Infect. Dis.* **12**, 1822-1826 (2006).
236 [16] Stockwell, D.R.B. & Peterson, A.T. Effects of sample size on accuracy of species
237 distribution models. *Ecol. Mod.* **148**: 1-13 (2002).
238 [17] Hernandez, P.A., *et al.* The effect of sample size and species characteristics on performance
239 of different species distribution modeling methods. *Ecog.* **29**:773-785 (2006).
240 [18]Wisz, M.S., *et al*. Effects of sample size on the performance of species distribution models.
241 *Div. Dist.* **14**: 763-773 (2008).
242 [19] Pigott, D.M., *et al.* Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* **3**,
243 e04395 (2014).
244 [20] Christofferson, R.C. Zika virus emergence and expansion: Lessons learned from dengue and
245 chikungunya may not provide all the answers. *Am. J. Trop. Med. Hyg. (in press).*
246 [21] Messina, J.P., *et al.* A global compendium of human dengue virus occurrence. *Sci. Data* **1**,
247 140004 (2014).
248 [22] Messina, J. P., *et al*. The many projected futures of dengue. *Nature Rev. Microbiol.* ***13***, 230-
249 239.(2015).
250 [23] Johansson, M.A., Cummings, D.A.T. & Glass, G.E. Multiyear climate variability and
251 dengue – El Niño southern oscillation, weather, and dengue incidence in Puerto Rico,
252 Mexico, and Thailand: a longitudinal data analysis. *PLoS Medicine* **6**, e1000168 (2009).
253 [24] Ritchie, S.A. & Rochester, W. Wind-blown mosquitoes and introduction of Japanese
254 encephalitis into Australia. *Emerg. Infect. Dis.* **7**, 900 (2001).
255 [25] Darwish, M. A., Hoogstraal, H., Roberts, T. J., Ahmed, I. P., & Omar, F. A sero-
256 epidemiological survey for certain arboviruses (Togaviridae) in Pakistan. *T. Roy. Soc. Trop.*
257 *Med. H* **77**, 442-445 (1983).
258 [26] Ioos, S., *et al*. Current Zika virus epidemiology and recent epidemics. *Med. maladies*
259 *infect.* **44**, 302-307 (2014).

[27] Neumann, G., Noda, T. & Kawaoka, Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **459**, 931-939 (2009).

[28] Foy, B. D., *et al*. Probable non-vector-borne transmission of Zika virus, Colorado, USA. *Emerg. Infect. Dis.* **17**, 880-2 (2011).

[29] Gao, F., *et al.* Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397**, 436-441 (1999).

## Methods

Occurrence data for Zika virus was compiled from the literature from studies dating as far back as the original discovery of the virus in Zika Forest, Uganda in 1947. Special attention was paid to correctly attributing cases of travelers to the true source of infection. Locality data was extracted from a set of clinical and survey papers, and georeferenced using a combination of Google Maps for hospitals and the Tulane University GEOLocate web platform for the remainder[1], which allows for the attribution of an uncertainty radius to points only identified to a regional level. Sixty four points were used in the final models presented in our paper after limiting our results to only locations that could be estimated within 75 km (Extended Data: Table S1). To our knowledge, this spatially explicit database is the most inclusive dataset currently in the literature.

Occurrence data for the other species included in our study were compiled from the literature. For *Aedes africanus*, we used a dataset of 99 points downloaded from the Global Biodiversity Informatics Facility (www.gbif.org). GBIF's coverage of *Aedes aegypti* and *Aedes albopictus* was however deemed to be lacking; occurrences for those species was taken from the previously published work of Kraemer *et al.*[2,3] Messina *et al.*'s database was used for dengue[4], as it has been previously published in *Scientific Data* and used with great success to generate a global distribution model.[5] Both of these datasets were reduced down to point-only data (i.e., polygons of occurrence were excluded), leaving 5,216 points for dengue and 13,992 and 17,280 points for *Ae. aegypti* and *Ae. albopictus* respectively.

289 Due to the potentially transient nature of the New World distribution of Zika virus, our model
290 uses presence and 1000 randomly selected pseudo-absence points from the Eurasian, African,
291 and Australian regions where the virus is established. We used the WorldClim data set
292 BIOCLIM at 2.5 arcminute resolution to provide all but one of our climate variables.[6] The
293 BIOCLIM features 19 variables (BIO1-BIO19) that summarize trends and extremes in
294 temperature and precipitation at a global scale. Given the relevance of the normalized difference
295 vegetation index (NDVI) in previous studies of dengue and as a predictor of vector mosquito
296 distributions[7], we downloaded monthly average NDVI layers for each month in 2014 from the
297 NASA Earth Observations TERRA/MODIS data portal[8], and averaged those twelve layers to
298 provide a single mean NDVI layer. Species distribution models were executed using the
299 BIOMOD2 package in R 3.1.1, which produces ensemble species distribution models using ten
300 different methods: general linear models (GLM), general boosted models or boosted regression
301 trees (GBM), general additive models (GAM), classification tree analysis (CTA), artificial neural
302 networks (ANN), surface range envelope (SRE), flexible discriminant analysis (FDA), multiple
303 adaptive regression splines (MARS), random forests (RF), and maximum entropy (MAXENT).[9]
304 Models were run individually for Zika (ZIKV), dengue (DENV), *Ae. aegypti, Ae. albopictus,* and
305 *Ae. africanus*. For Zika, models trained on Old World environmental data were used to establish
306 the potential distribution of the virus in the Americas under climatic conditions captured by
307 WorldClim data, which represent an expected range of variability that do not incorporate
308 anomalous events like 2015 El Niño Southern Oscillation.
309
310 To address colinearity in the environmental variable set, we produced a correlation matrix for
311 our 20 variables, and identified each pair with a correlation coefficient > 0.8. For each species,

312 we ran a single ensemble model with all ten methods and averaged the variable importance for
313 our 20 predictors across the methods (See Table S2-S6). In each pair we identified the variable
314 with the greater contribution, and we produced species-specific reduced variable sets used in the
315 final published models by eliminating any covariates that universally performed poorer than their
316 pairmate. Based in this criteria, we excluded the following variables for each species to reduce
317 colinearity:

318 - ZIKV: BIO8, BIO9, BIO14, BIO18
319 - DENV: BIO3, BIO5, BIO12, BIO17
320 - *Ae. aegypti:* BIO6, BIO8, BIO12, BIO17
321 - *Ae. africanus:* BIO5, BIO6, BIO12, BIO17
322 - *Ae. albopictus:* BIO8, BIO9, BIO16, BIO17

323 The AUC of every model run with reduced variable sets is presented in Table S7. We found no
324 significant correlation between NDVI and any individual BIOCLIM variable, so NDVI was
325 included in every model of current distributions. We ran five iterations of each reduced variable
326 set model and eliminated any prediction methods from the ensemble with an AUC of lower than
327 0.95, so that the final model had only included the best predicting models. This was found to
328 only leave the RF method for DENV, so a cutoff of 0.9 was applied in that case, to keep the
329 ensemble approach constant across datasets. The final models were run with the following
330 methods with ten iterations using an 80/20 training-test split in the final presentation:

331 - ZIKV: GLM, GBM, GAM, CTA, FDA, MARS, RF
332 - DENV: GLM, GBM, GAM, FDA, MARS, RF, MAXENT
333 - *Ae. aegypti:* GLM, GBM, GAM, CTA, ANN, FDA, MARS, RF
334 - *Ae. africanus:* GLM, GBM, GAM, CTA, ANN, FDA, MARS, RF

335 • *Ae. albopictus:* GLM, GBM, GAM, CTA, FDA, MARS, MAXENT, RF

336 The importance of variables of the reduced model set for each are presented in Table S8-S12.

337

338 To project the distribution of the species under climate change, we reran each model with the
339 previously chosen method and variable sets but excluding NDVI, for which we did not feel we
340 could appropriately simulate future values. BioClim forecasts were taken from WorldClim using
341 the Hadley Centre Global Environmental Model v. 2 Earth System climate forecast (HadGEM2-
342 ES) predictions for representative climate pathway 8.5 (RCP85), which, within that model,
343 represents a worst case scenario for carbon emissions and climate warming.[10] All five species'
344 models were retrained on current climate data and projected onto forecasts for the year 2050, the
345 results of which are shown in Figure 3. Finally, to compare the niche of dengue and Zika, we
346 used the R package ecospat, which uses principal component analysis to define the position of
347 species' ecological niche relative to background environmental variation[11,12]. The ecospat
348 analysis was run using the full 64 point database and the full extent of global environmental data,
349 because, while the niche of Zika in the Americas is uncertain, dengue is well established, and
350 measuring its niche required a full background sample. We excluded BIO5 and BIO12 from our
351 analysis as they were included in neither of the final models for the diseases; niche similarity
352 tests were run 100 times with 100 iterations each. The results of that analysis are presented in
353 Figure S1, which shows both the one-directional similarity test and the bidirectional equivalence
354 test.

355

356 Finally, to assess the transferability of our Zika model across environmental space, we conducted
357 a geographic cross validation (GCV) between African and Asian datasets. While under normal

circumstances, a model would be trained on New World data and projected onto the Old World, the lack of data prior to the current outbreak makes such a direct comparison infeasible. However, given the evidence for separate Asian and African strains, a cross-validation between the two was supported, and models trained on those two continents were projected globally to test the performance of the model across geographic regions (Extended Data Fig. 7), and evaluate how sensitive our projections in the Americas are to the environmental covariates sampled. The clustering of points in western India narrows the environmental range sampled by presences, significantly limiting the transferability of the Asian sub-model. In contrast, the African sub-model performs well in new regions, and corresponds well to the global model.

**References**


[1] GEOLocate. http://www.museum.tulane.edu/geolocate/web/WebGeoref.aspx. Accessed 2/7/2016.

[2] Kraemer, M.U.G., *et al.* The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *eLife* **4**, e08347 (2015).

[3] Kraemer, M.U.G., *et al.* The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence. *Sci. Data* **2**, 150035 (2015).

[4] Messina, J.P., *et al.* A global compendium of human dengue virus occurrence. *Sci. Data* **1**, 140004 (2014).

[5] Bhatt, S. The global distribution and burden of dengue. *Nature* **496**, 504-507 (2013).

[6] Hijmans, R. *et al.* Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965-1978 (2005).

[7] Peterson, A.T., Martínez-Campos, C., Nakazawa, Y. & Martínez-Meyer, E. Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. *T. Roy. Soc. Trop. Med. H.* **99**, 647-655 (2005).

[8] NASA Earth Observations TERRA/MODIS NDVI. http://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD13A2_M_NDVI. Accessed 2/7/2016.

[9] Thuiller, W., Lafourcade, B., Engler, R. & Araujo, M.B. BIOMOD: a platform for ensemble forecasting of species distributions. *Ecography* **32**, 369-373 (2009).

[10] Jones, C., *et al.* Twenty-first-century compatible CO2 emissions and airborne fraction simulated by CMIP5 Earth System models under four representative concentration pathways. *J. Climatol.* **26**, 4398-4413 (2013).

[11] Broenniman, O., *et al.* Ecospat: spatial ecology: miscellaneous methods. R package version 1.0. http://www.unil.ch/ecospat/home/menuinst/tools-data/tools.html. (2014)

[12] Broenniman, O., *et al.* Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecol. Biogeogr.* **21**, 481-497 (2012).

## Figures and Tables

**Figure 1.** The global distribution of case reports of Zika virus (1947 to February 2016) broken down by country (blue shading) and an ensemble niche model built from occurrence data (red shading). Our model predicts occurrence in part of every shaded country; it is clear that displaying cases at country resolution overstates the distribution of the virus, especially in the Americas (for example, Alaska).

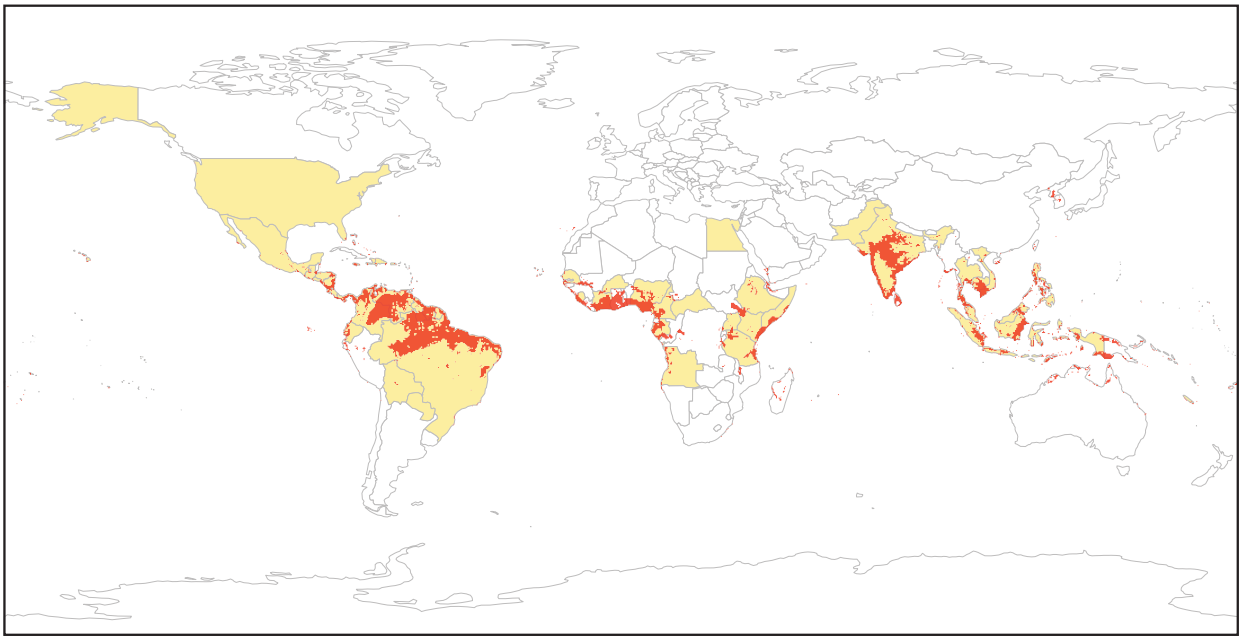


**Figure 2.** The ecological niche of Zika and dengue in principal component space (a). Solid and dashed lines are 100%/50% boundaries for all environmental data Despite apparent overlap in environmental niche space, the dissimilarity between the black shading in each principal component graph indicates statistically significant differences between the niches, evident in the projections of our niche models for dengue (b) and Zika (c).

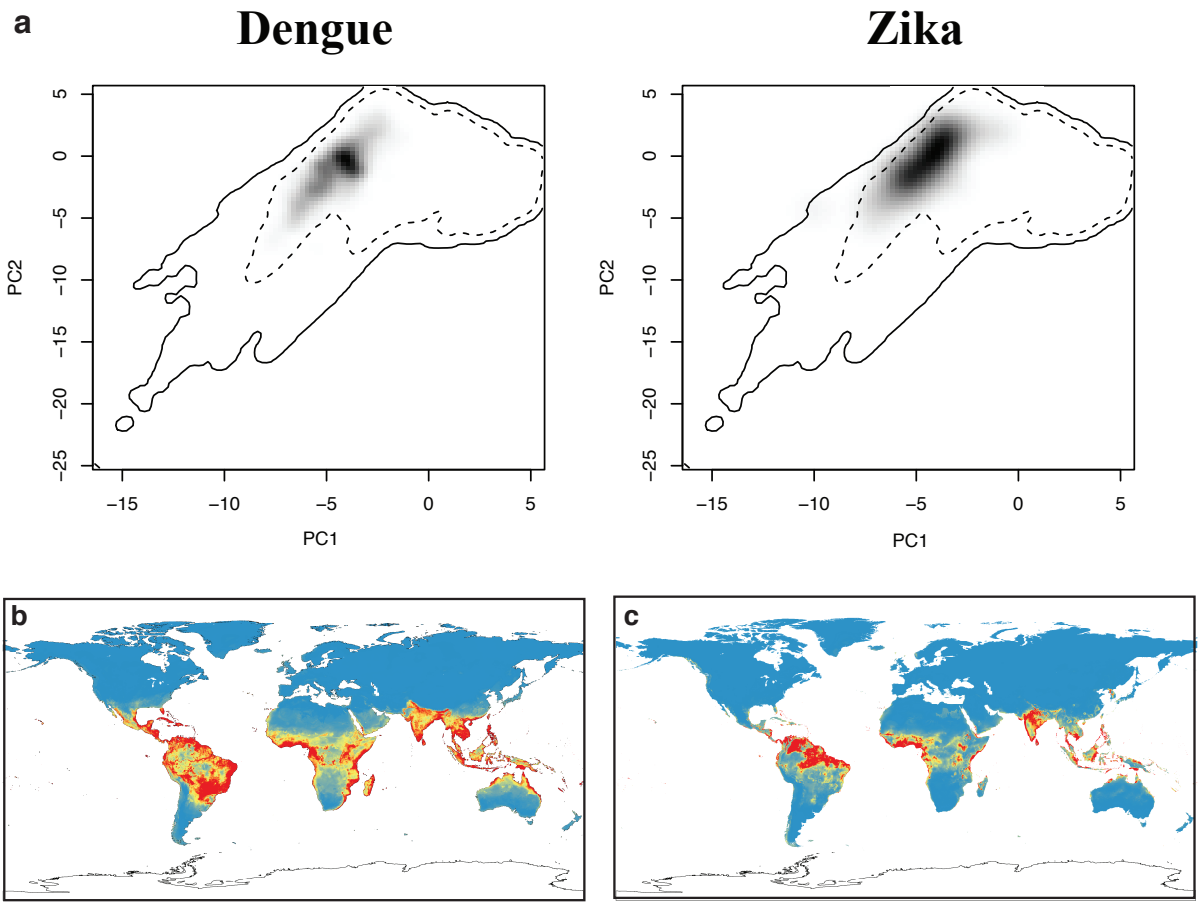
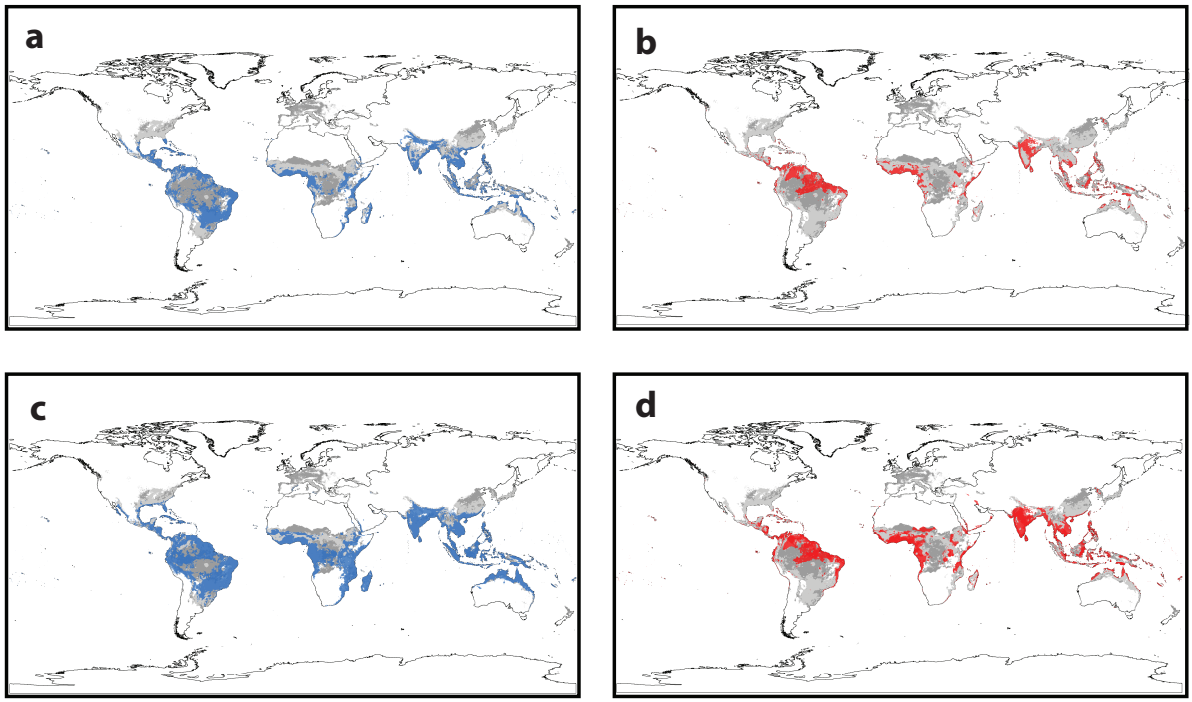


**Figure 3.** The estimated distribution of Zika (red) and dengue (blue) based on current (a, b) and 2050 climate projections (c, d), compared against the current (light grey) and future distribution of all three mosquito vectors (a-d).



**Extended Data:**

**Extended Data Figure 1 |** Final ensemble model for Zika virus

**Extended Data Figure 2 |** Final ensemble model for dengue fever

**Extended Data Figure 3 |** Final ensemble model for *Aedes aegypti*

**Extended Data Figure 4 |** Final ensemble model for *Aedes africanus*

**Extended Data Figure 5 |** Final ensemble model for *Aedes albopictus*

**Extended Data Figure 6 |** Niche overlap between ZIKV and DEV

**Extended Data Figure 7 |** Geographic cross validation