Mapping the global geographic potential of Zika virus spread

Abdallah M. Samy[1, 2*], Stephanie M. Thomas[3], Ahmed Abd El Wahed[4], Kevin P. Cohoon[5], A. Townsend Peterson[1]

Author affiliations:

[1]University of Kansas, Lawrence 66045, USA

[2]Ain Shams University, Abbassia, Cairo 11566, Egypt

[3]University of Bayreuth, Universitaetsstrasse 30, D-95447 Bayreuth, Germany

[4]Georg-August University, Burckhardtweg 2, 37077, Goettingen, Germany

[5]Mayo Clinic, Rochester, Minnesota 55902, USA.

Supplementary Materials: Modeling Methods

Occurrences of Zika virus:

Records of the occurrence of Zika virus (ZIKV) were compiled from the Brazilian Ministry of Health (http://portalsaude.saude.gov.br), ProMED-mail (http://www.promedmail.org), and HealthMap (http://www.healthmap.org). Records were filtered to remove duplicate coordinates, and inspected to remove any cases that had been corrected or retracted. The data set for model calibration includes records from Mexico, Central America, and South America, as only in these regions had data sufficiently dense for rigorous model calibration. We randomly divided these records into two equal sets: one to be used for model calibration and the other for model evaluation.

Model Covariates:

Several suites of environmental variables were used as independent variables in our models, by which to characterize environmental variation across the calibration area and globally. These variables were considered as potential drivers of both direct and indirect effects on emergence of mosquito-borne diseases (Peterson 2013). These variables include daytime and nighttime land surface temperature and enhanced vegetation index (EVI) values for January 2012 – July 2015 derived from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery (source: http://goo.gl/fO1Ehh). EVI approximates photosynthetic mass (termed colloquially ‘greenness’), and offers a proxy measure of soil moisture, an important factor in determining larval mosquito habitats (Estallo et al. 2008; Nihei et al. 2014).

Precipitation, aridity, and water stress are also important potential factors in the distribution of ZIKV, in light of their major roles in determining mosquito breeding sites. Effects of precipitation, aridity, and soil water stress on breeding habitats for ZIKV vector species (*Aedes* spp.) have been examined previously (Alto& Juliano 2001; Kalra et al. 1997; Mogi et al. 2015; Wu et al. 2009). An important result is that drought drives human need for water storage in containers, leading to elevated mosquito populations and increased biting rates (Kalra et al. 1997); indeed, Brazil has faced recent drought, which caused increased water storage (Sena et al. 2014). To account for these factors, we used data for (1) maximum and minimum monthly precipitation data from WorldClim (http://www.worldclim.org) and (2) aridity and soil water stress data layers from the Consortium for Spatial Information (CGIAR-CSI; http://www.cgiar-csi.org/), respectively. For soil water stress, we calculated composite indices as maxima and minima across all of the monthly estimates.

Land cover has been identified as key in breeding habitats of *Aedes* spp., the likely prime vectors of ZIKV (Vanwambeke et al. 2007). We thus used a global land cover layer available from WorldGrids (http://worldgrids.org/doku.php). We accounted for elevational gradients identified as influential factors in disease dynamics and mosquito abundance by means of a layer summarizing slope (Dhimal et al. 2015; Lozano-Fuentes et al. 2012).

Disease transmission and spread dynamics, however, are not dependent solely on environmental factors; other sets of variables also play major roles (Koyadun et al. 2012; Teurlai et al. 2015). Hence, we used grids of human population density, nighttime lights, and accessibility via

transportation. Population density grids were drawn from the Gridded Population of the World, version 4 (GPWv4), collection released recently via (http://beta.sedac.ciesin.columbia.edu/data/collection/gpw-v4). Nighttime satellite imagery (year 2013) was used as a proxy for real poverty (Noor et al. 2008; Wang et al. 2012), and was obtained from NOAA-Defense Meteorological Satellite Program (http://goo.gl/uru9Pd). Accessibility was summarized in terms of travel time by land or sea (Nelson 2008), as the connectivity between population sites is an important variable in potential distributions of emerging diseases (Cliff& Haggett 2004); this layer was developed by the European Commission and World Bank (http://goo.gl/zc6EqH). Finally, we used two grids representing the environmental suitability for *Aedes aegypti* and *Ae. albopictus* as potential vectors for ZIKV derived from a recent detailed analysis (Kraemer et al. 2015); we developed parallel models based on another recent analysis of the same two species (Campbell et al. 2015). All grids were resampled to a spatial resolution of 5 x 5 km in ArcMap 10.3, in light of the global distribution of the species.

Ecological Niche Modeling:

We approximated the ZIKV fundamental ecological niche via ecological niche modeling (ENM) via the maximum entropy algorithm implemented in Maxent, version 3.3 (Phillips et al. 2006). Peterson *et al*. defined the fundamental niche as “the set of environmental conditions required for the species to maintain populations without immigrational subsidy” (Peterson et al. 2011). ENM relates known occurrences of species to the set of environmental variables in a maximum entropy, evolutionary-computing environment to approximate this set of environmental conditions associated with maintenance of populations (Peterson et al. 2011).

We calibrated ENMs across Mexico, Central America, and South America, where ZIKV occurrence data were sufficiently dense for rigorous model calibration (Owens et al. 2013); models were projected worldwide for interpretation. To explore and understand contributions of different suites of variables to shaping the distribution of ZIKV, we used different combinations of environmental variables, socioeconomic variables, and accessibility (see Neerinckx et al. 2008). These explorations illuminate the roles of possible drivers of ZIKV transmission beyond just climate (Kilpatrick& Randolph 2012; Weaver 2013). A full elaboration of combinations of drivers that we explored is presented in Table 1.

For each combination, we ran Maxent using 100 bootstrap replicates. The median of the outputs of the replicates was used as a best final estimate in subsequent analyses. Final models were thresholded based on a maximum allowable omission error rate of 5% [E = 5%; (Peterson et al. 2008)], in effect assuming that $\leq$ 5% of occurrence data would have sufficient error in geolocation that variable values might be misrepresented.

Model predictions were evaluated for statistical significance based on predictions among random subsets of 50% of available data. In effect, these tests assess whether a model based on a certain amount of occurrence information across a region will be able to anticipate the next set of occurrences across that same landscape; while not an ideal test of predictive ability when models are transferred to other regions, it is at present the only test available to us. We used partial ROC statistics, which avoid well-known problems with traditional ROC approaches (Lobo et al. 2008). Partial ROC statistics were calculated using the PartialROC function in the ENMGadgets package in R software version 3.2.0 (R Devlopment Core Team 2015), specifying the same E = 5%, a 50% bootstrap resampling, and 100 random iterations (Lobo et al. 2008).

For visualization, we combined two of these thresholded models (Model 3 and Model 4) to illustrate differences between predictions based on different possible drivers of ZIKV transmission. Interesting contrasts emerge from differences in suitability based on all of the environmental dimensions, areas identified as suitable based on climate and presence of vector species, and suitability in terms of human socioeconomic variables and accessibility, as is exemplified in the Figure in the main paper. GIS-readable grids (GeoTIFF format) and KML files are available via Figshare (***LINK TO BE PROVIDED UPON ACCEPTANCE***) and the University of Kansas's ScholarWorks (***LINK TO BE PROVIDED UPON ACCEPTANCE***).

References:

Alto BW, Juliano SA 2001. Precipitation and temperature effects on populations of *Aedes albopictus* (Diptera: Culicidae): implications for range expansion. *Journal of medical entomology*, 38 (5), 646-656.

Campbell LP, Luther C, Moo-Llanes D, Ramsey JM, Danis-Lozano R, Peterson AT 2015. Climate change influences on global distributions of dengue and chikungunya virus vectors. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370 (1665).

Cliff A, Haggett P 2004. Time, travel and infection. *British medical bulletin*, 69, 87-99.

Dhimal M, Gautam I, Joshi HD, O'Hara RB, Ahrens B, Kuch U 2015. Risk factors for the presence of chikungunya and dengue vectors (*Aedes aegypti* and *Aedes albopictus*), their altitudinal distribution and climatic determinants of their abundance in central Nepal. *PLoS Negl Trop Dis*, 9 (3), e0003545.

Estallo EL, Lamfri MA, Scavuzzo CM, Almeida FF, Introini MV, Zaidenberg M, Almiron WR 2008. Models for predicting *Aedes aegypti* larval indices based on satellite images and climatic variables. *J Am Mosq Control Assoc*, 24 (3), 368-376.

Kalra NL, Kaul SM, Rastogi RM 1997. Prevalence of *Aedes aegypti* and *Aedes albopictus* vectors of DF/DHF in north, northeast and central India. *Dengue Bulletin* 21, 84–92.

Kilpatrick AM, Randolph SE 2012. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet*, 380 (9857), 1946-1955.

Koyadun S, Butraporn P, Kittayapong P 2012. Ecologic and sociodemographic risk determinants for dengue transmission in urban areas in Thailand. *Interdisciplinary perspectives on infectious diseases*, 2012, 907494.

Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Elyazar IR, Teng HJ, Brady OJ, Messina JP, Pigott DM, Scott TW, Smith DL, Wint GR, Golding N, Hay SI 2015. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *ELife*, 4, e08347.

Lobo JM, Jiménez-Valverde A, Real R 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17 (2), 145-151.

Lozano-Fuentes S, Hayden MH, Welsh-Rodriguez C, Ochoa-Martinez C, Tapia-Santos B, Kobylinski KC, Uejio CK, Zielinski-Gutierrez E, Monache LD, Monaghan AJ, Steinhoff DF, Eisen L 2012. The dengue virus mosquito vector Aedes aegypti at high elevation in Mexico. *Am J Trop Med Hyg*, 87 (5), 902-909.

Mogi M, Armbruster P, Tuno N, Campos R, Eritja R 2015. Simple Indices Provide Insight to Climate Attributes Delineating the Geographic Range of Aedes albopictus (Diptera: Culicidae) Prior to Worldwide Invasion. *Journal of medical entomology*, 52 (4), 647-657.

Neerinckx SB, Peterson AT, Gulinck H, Deckers J, Leirs H 2008. Geographic distribution and ecological niche of plague in sub-Saharan Africa. *International journal of health geographics*, 7, 54.

Nelson A 2008. Estimated travel time to the nearest city of 50,000 or more people in year 2000. Global Environment Monitoring Unit - Joint Research Centre of the European Commission, Ispra Italy. Available at http://forobs.jrc.ec.europa.eu/products/gam/

Nihei N, Komagata O, Mochizuki K, Kobayashi M 2014. Geospatial analysis of invasion of the Asian tiger mosquito Aedes albopictus: competition with Aedes japonicus japonicus in its northern limit area in Japan. *Geospatial health*, 8 (2), 417-427.

Noor AM, Alegana VA, Gething PW, Tatem AJ, Snow RW 2008. Using remotely sensed night-time light as a proxy for poverty in Africa. *Population Health Metrics*, 6, 5-5.

Owens HL, Campbell LP, Dornak LL, Saupe EE, Barve N, Soberón J, Ingenloff K, Lira-Noriega A, Hensz CM, Myers CE, Peterson AT 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, 263, 10-18.

Peterson A, Soberón J, Pearson R, Anderson R, Martínez-Meyer E, al. e 2011. Ecological Niches and Geographic Distributions Princeton: Princeton University.

Peterson AT 2014. Mapping Disease Transmission Risk: Enriching Models Using Biogeography and Ecology *Johns Hopkins University Press, Baltimore, Maryland, USA*.

Peterson AT, Papeş M, Soberón J 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, 213 (1), 63-72.

Phillips SJ, Anderson RP, Schapire RE 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190 (3-4), 231-259.

Sena A, Barcellos C, Freitas C, Corvalan C 2014. Managing the Health Impacts of Drought in Brazil. *International Journal of Environmental Research and Public Health*, 11 (10), 10737-10751.

R Devlopment Core Team 2015. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austia. Available at http://www.R-project.org.

Teurlai M, Menkes CE, Cavarero V, Degallier N, Descloux E, Grangeon JP, Guillaumot L, Libourel T, Lucio PS, Mathieu-Daude F, Mangeas M 2015. Socio-economic and Climate Factors Associated with Dengue Fever Spatial Heterogeneity: A Worked Example in New Caledonia. *PLoS Negl Trop Dis*, 9 (12), e0004211.

Vanwambeke SO, Somboon P, Harbach RE, Isenstadt M, Lambin EF, Walton C, Butlin RK 2007. Landscape and land cover factors influence the presence of Aedes and Anopheles larvae. *Journal of medical entomology*, 44 (1), 133-144.

Wang W, Cheng H, Zhang L 2012. Poverty assessment using DMSP/OLS night-time light satellite imagery at a provincial scale in China. *Advances in Space Research*, 49 (8), 1253-1264.

Weaver SC 2013. Urbanization and geographic expansion of zoonotic arboviral diseases: mechanisms and potential strategies for prevention. *Trends in Microbiology*, 21 (8), 360-363.

Wu PC, Lay JG, Guo HR, Lin CY, Lung SC, Su HJ 2009. Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. *The Science of the total environment*, 407 (7), 2224-2233.

Table 1: Combinations of environmental variables used to calibrate the ecological niche models for ZIKV in this studies. Check marks indicate that the variable was used in the model; X's indicate variables that were not used in the model.

Variable	Model 1	Model 2	Model 3	Model 4
Population density 2015	X	√	X	√
Daytime temperature	√	√	√	√
Nighttime temperature	√	√	√	√
Maximum monthly precipitation	√	√	√	√
Minimum monthly precipitation	√	√	√	√
Enhanced vegetation index	√	√	√	√
Maximum soil water stress	√	√	√	√
Minimum soil water stress	√	√	√	√
Aridity	√	√	√	√
Elevation	√	√	√	√
Land-cover	X	√	X	√
***Aedes aegypti* abundance**	X	X	√	√
***Aedes albopictus* abundance**	X	X	√	√
Nighttime lights	X	√	X	√
Accessibility	X	√	X	√

Summary of Figures

Figure 1. Map of available ZIKV occurrence data in the Americas.

Figures 2-5. Maps of estimated ZIKV global potential distribution based on different combinations of environmental variables.

Figure 6. Overlay of known global occurrences of ZIKV on our ENM predictions (Model 1), as a further corroboration of the predictive ability of these models.

Figures 7, 8, and 9. Close-ups of North America, Europe, and Asia and Australia to provide additional detail to the figure in the main paper.

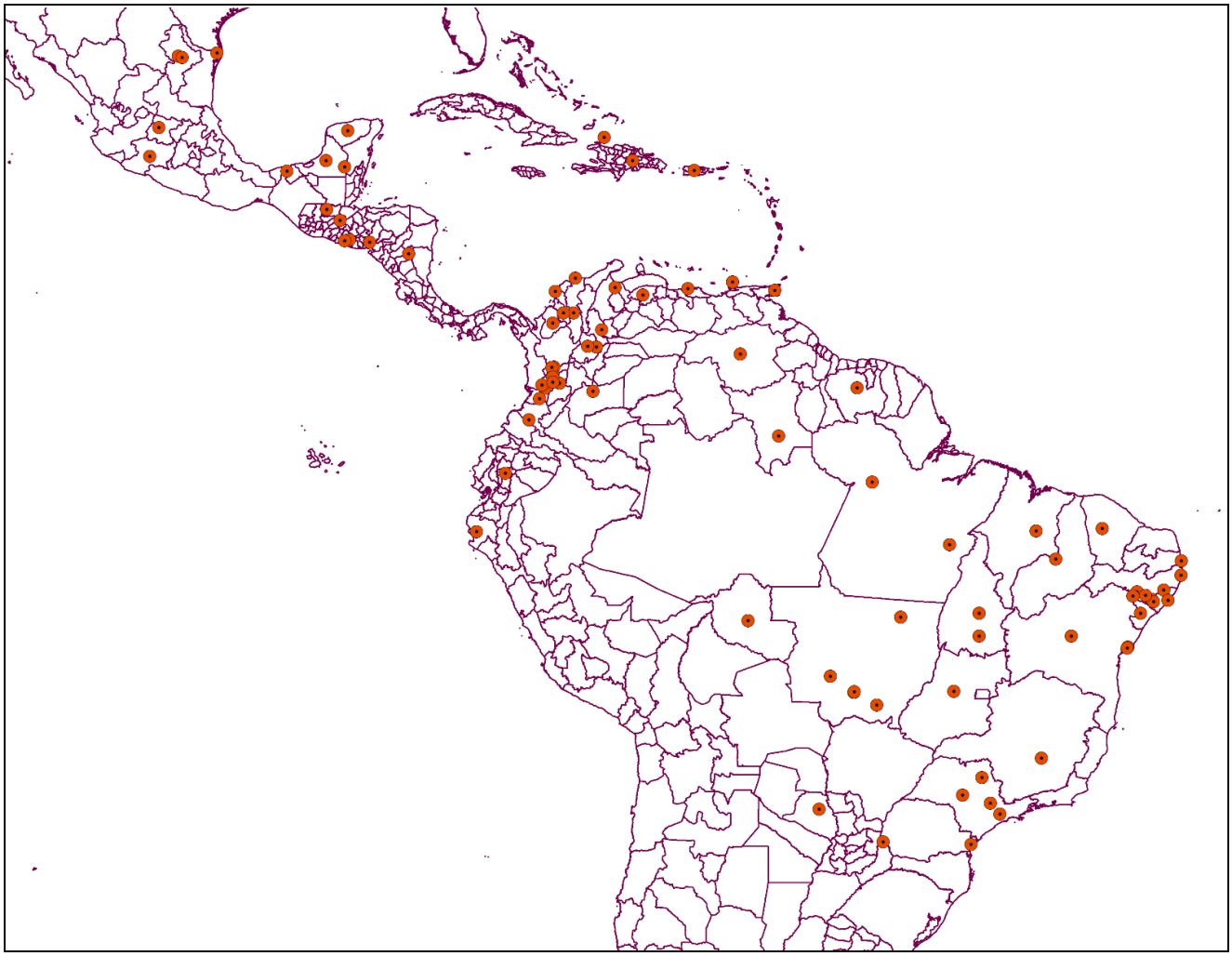

Figure 1: ZIKV occurrence records (dotted circles) used in model calibration across Mexico, Central America, and South America.

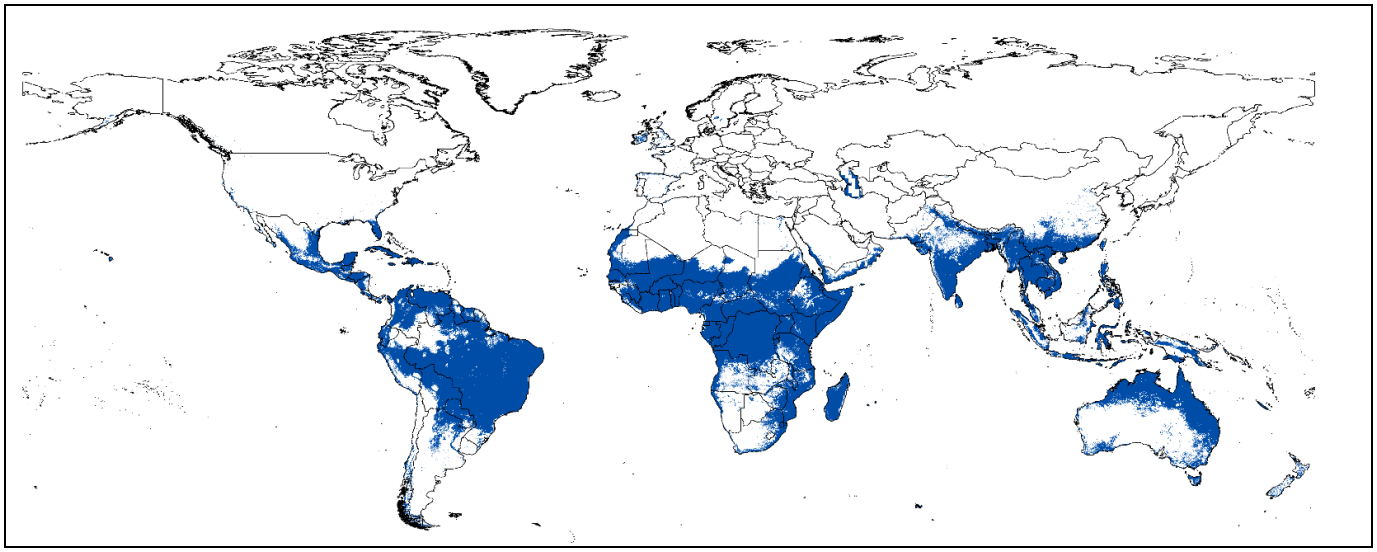

Figure 2: Potential geographic distribution of ZIKV (in blue) based on analysis of only environmental variables (Model 1), based on a threshold that admits a maximum of 5% omission error.

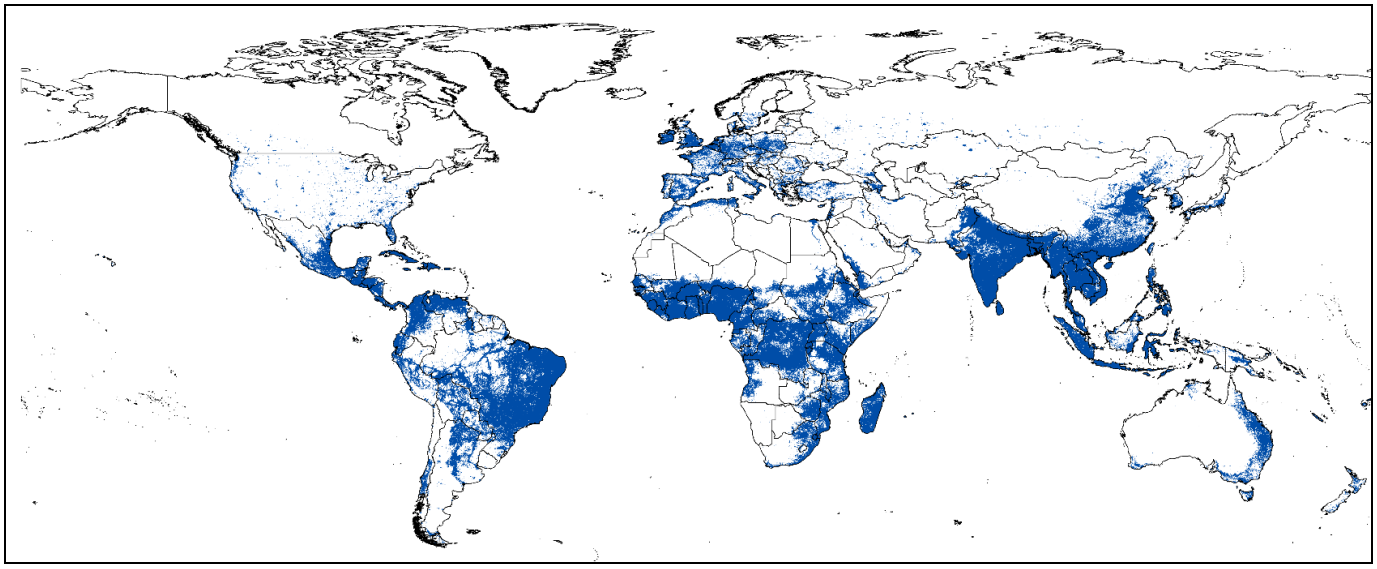

Figure 3: Potential geographic distribution of ZIKV (in blue) based on analysis of all environmental, socioeconomic, poverty proxy, and accessibility variables (Model 2), based on a threshold that admits a maximum of 5% omission error.

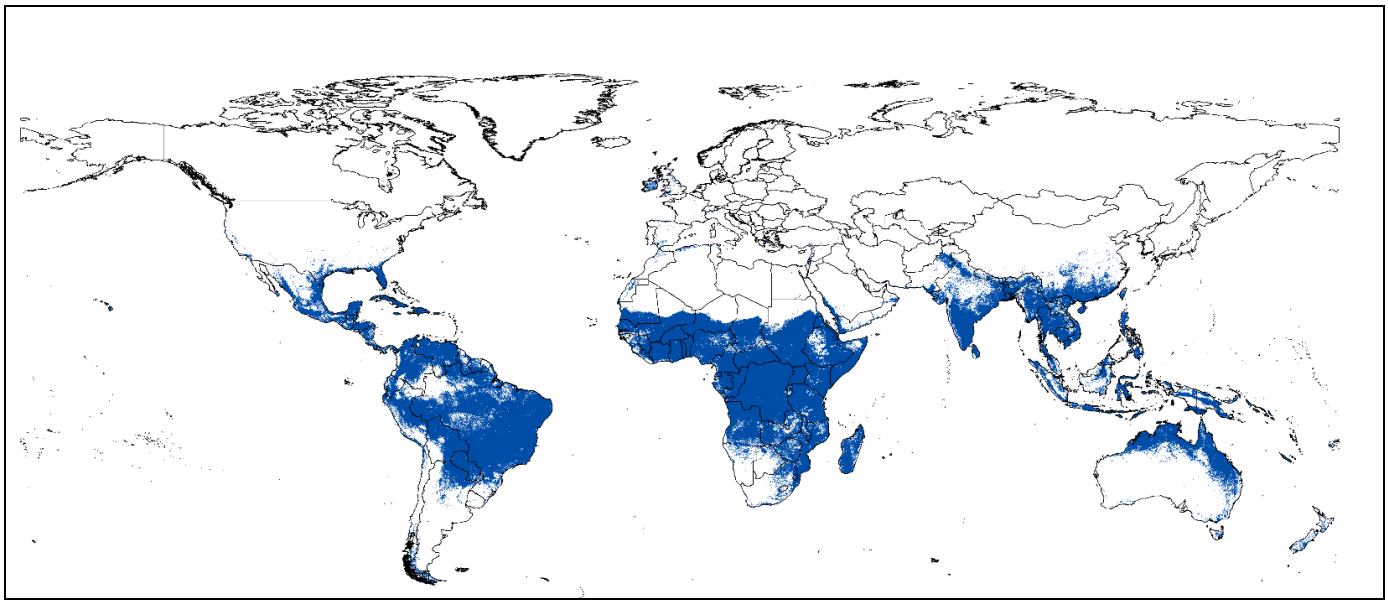

Figure 4: Potential geographic distribution of ZIKV (in blue) based on analysis of environmental variables and mosquito density (Model 3), based on a threshold that admits a maximum of 5% omission error.

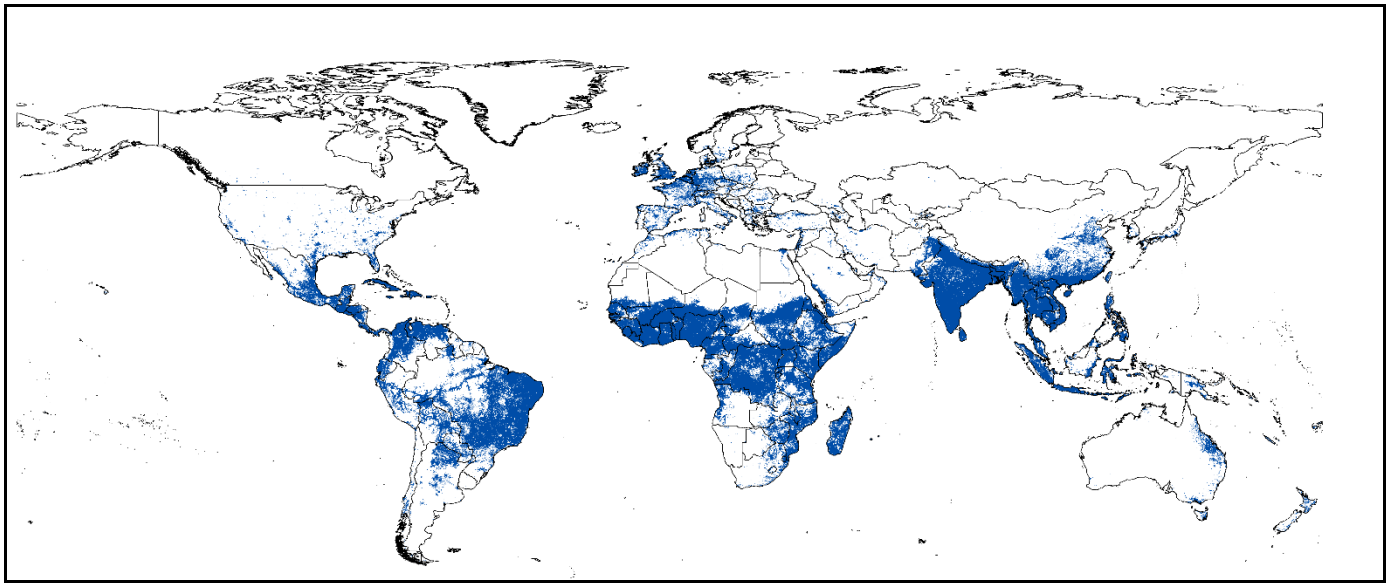

Figure 5: Potential geographic distribution of ZIKV (in blue) based on analysis of all environmental, socioeconomic, mosquito-related, poverty-related, and accessibility variables (Model 4), based on a threshold that admits a maximum of 5% omission error.

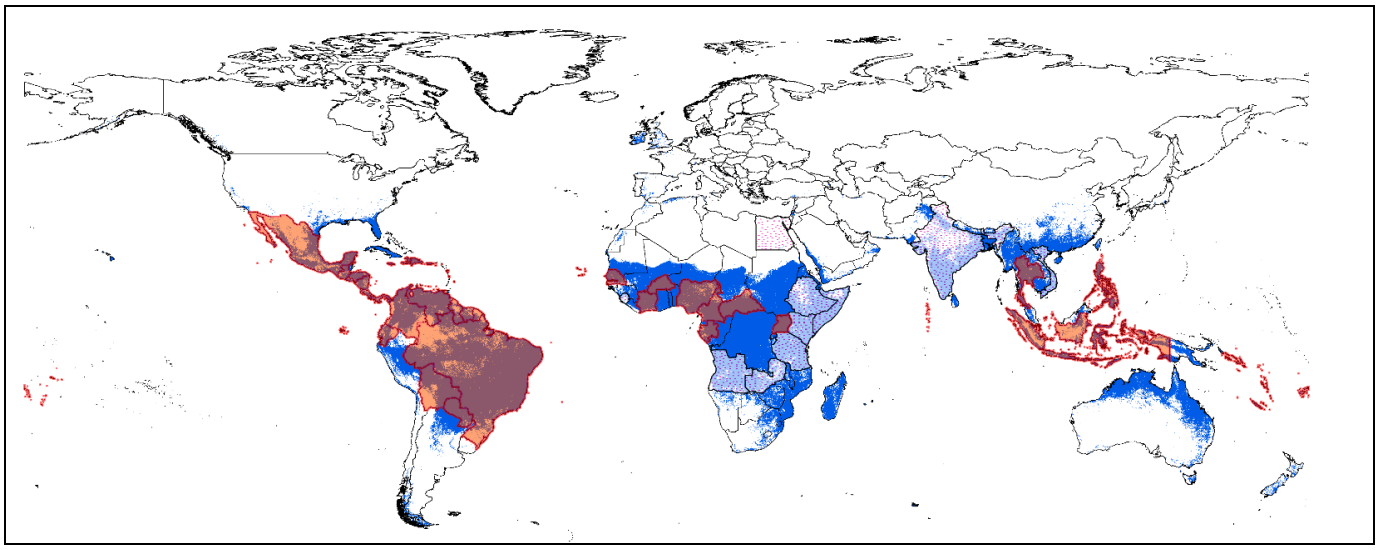

Figure 6: Potential geographic distribution of ZIKV based on environmental variables only (Model 1) in relation to known global distribution of the virus. Note that models were calibrated based on occurrences in Mexico, Central America, and South America, such that occurrences in Africa, Asia, and the Pacific had no contribution to model calibration. Countries reported with authochonous cases (orange shading with brown boundaries) and countries with known seropositive cases (red stippled areas and light-blue shading) are shown in relation to the model predictions (in blue, and purple and light blue where overlapped by the known-positive cases. Note the generally close correspondence between model predictions and known ZIKV occurrences.

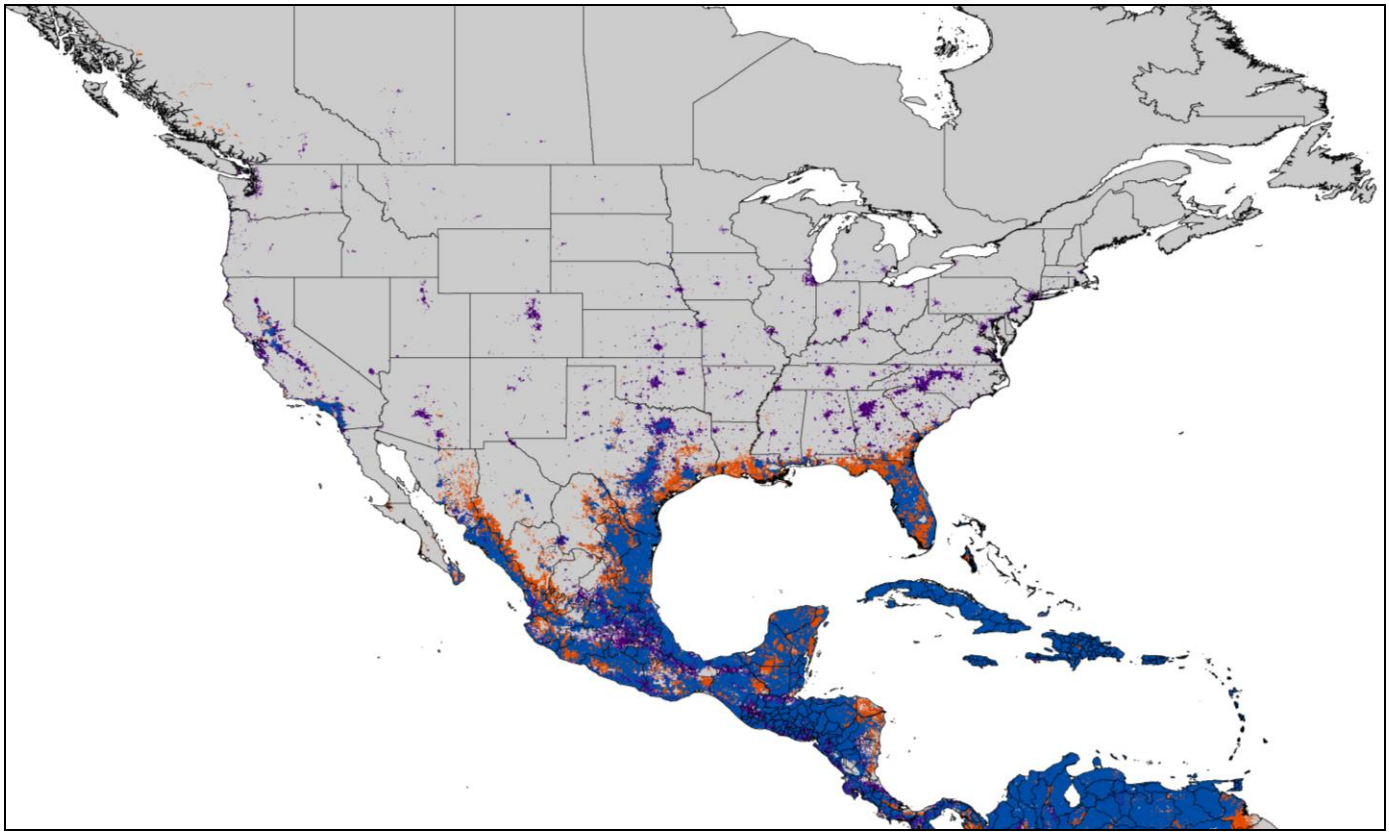

Figure 7. Close-up of North America to provide detail additional to the figure in the main paper.

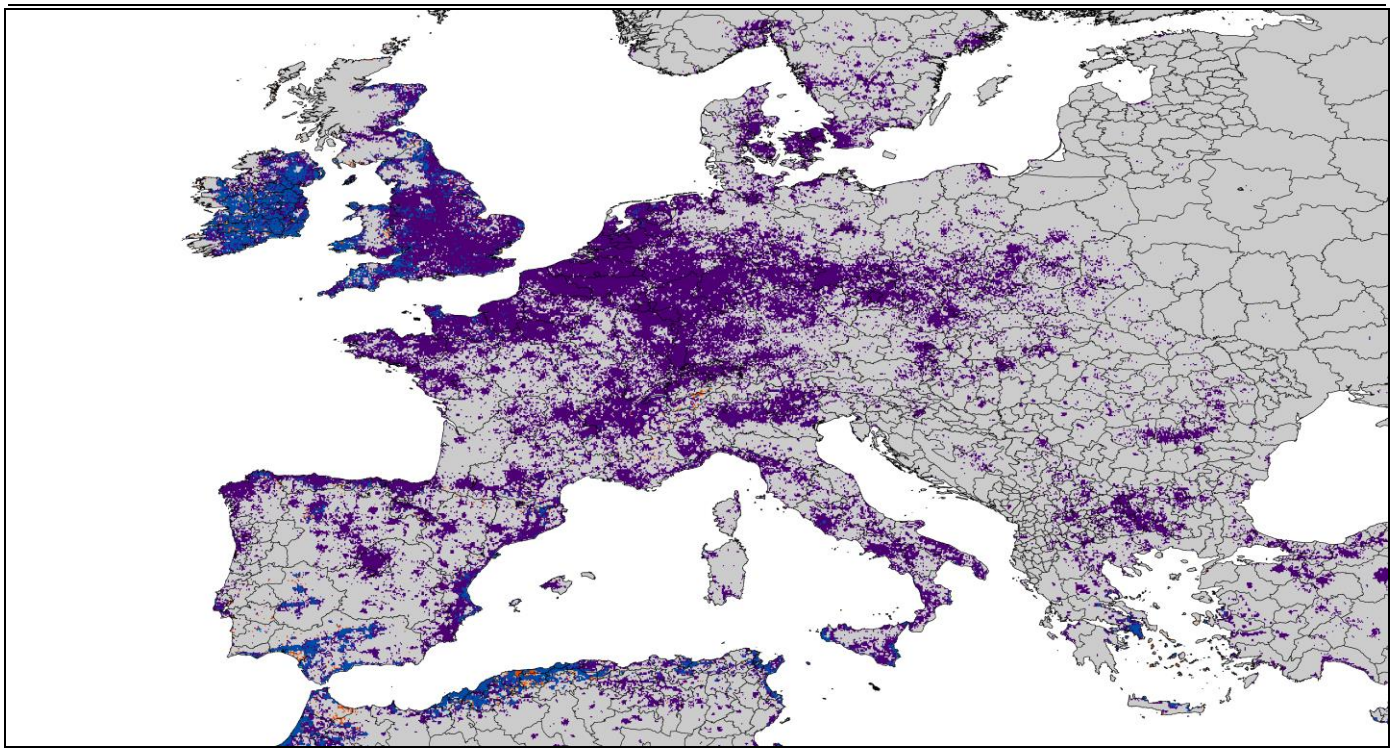

Figure 8. Close-up of Europe to provide detail additional to the figure in the main paper.

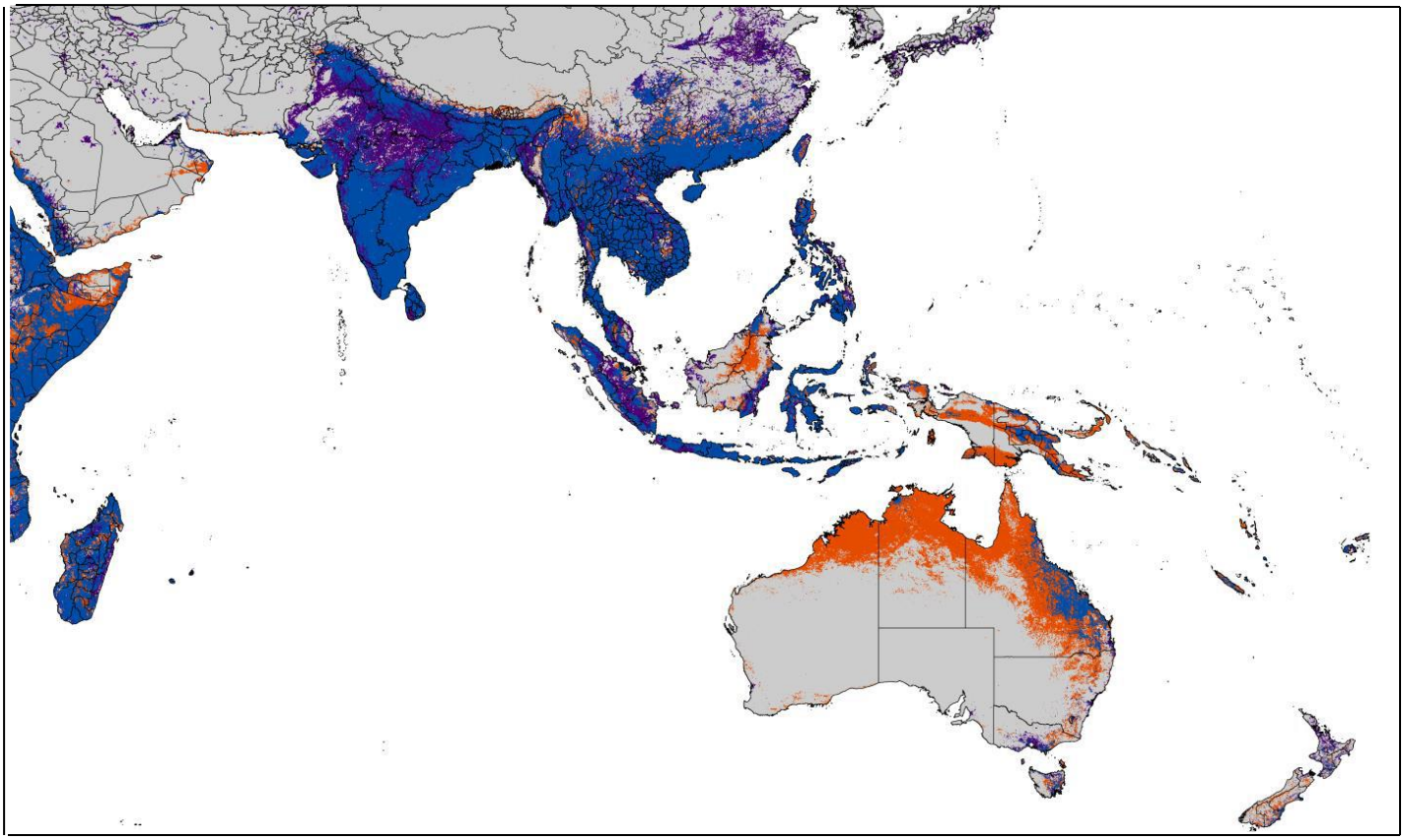

Figure 9. Close-up of Asia and Australia to provide detail additional to the figure in the main paper.