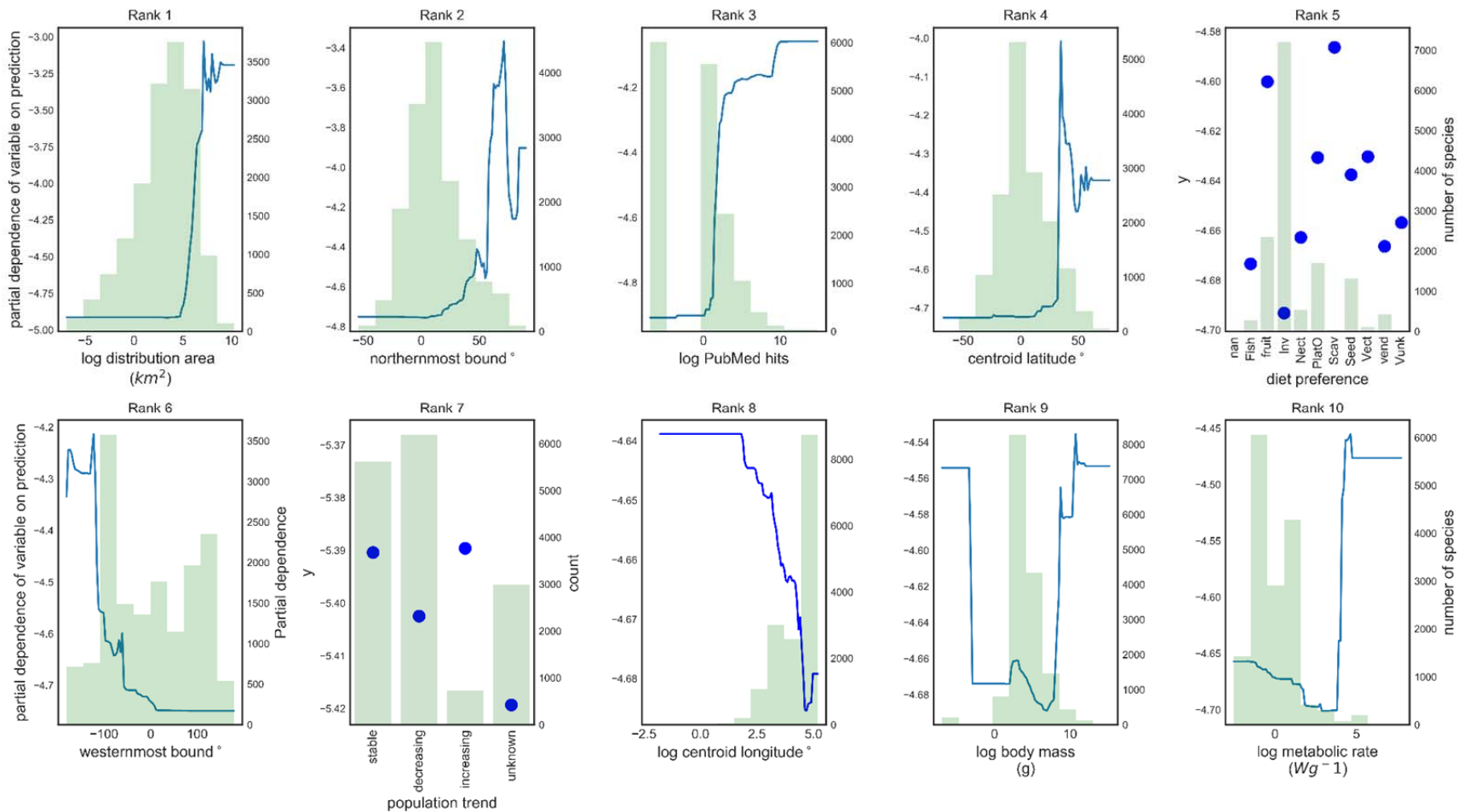


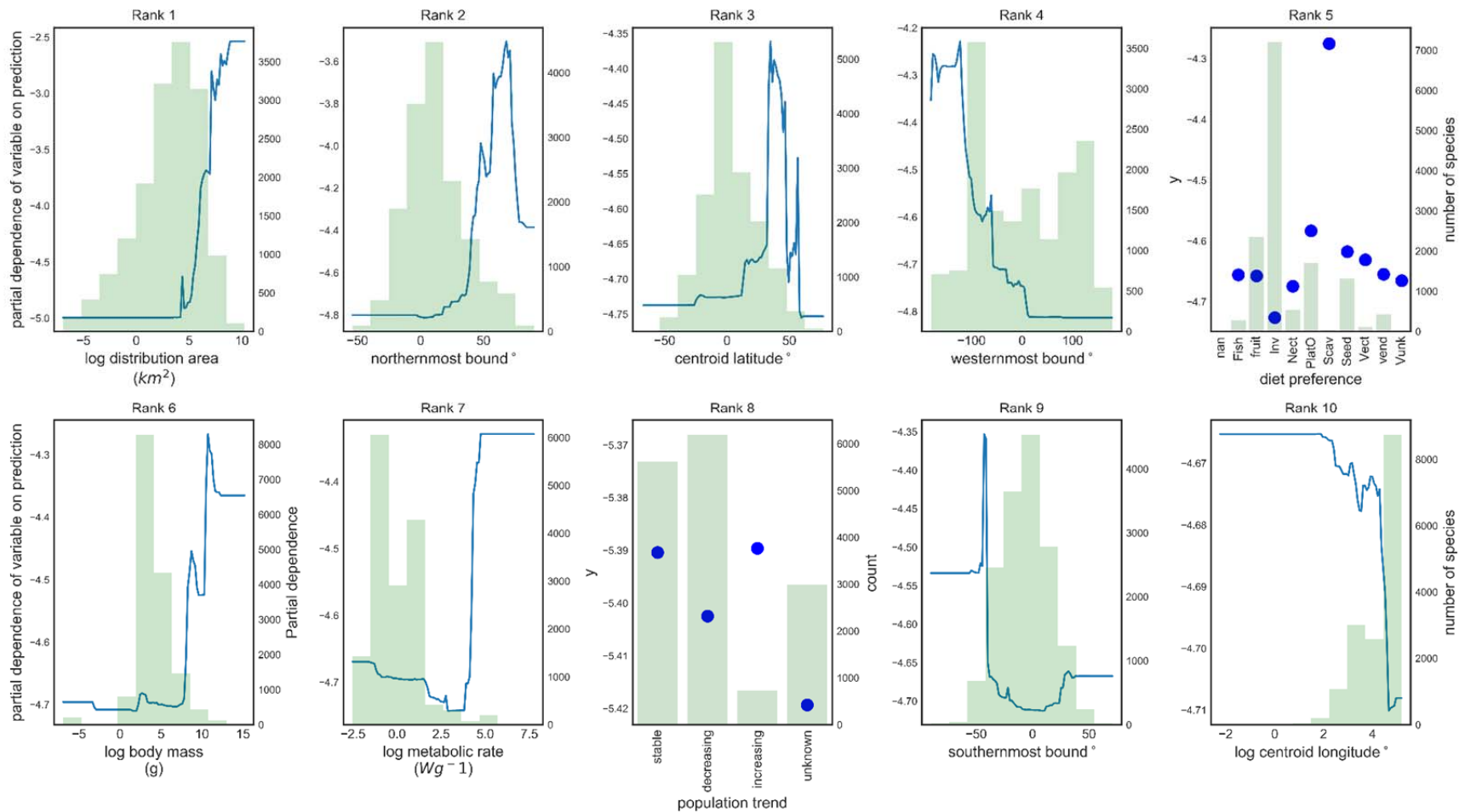
- 1 *Predicting wildlife reservoirs and global vulnerability to zoonotic*
- 2 *Flaviviruses*
- 3 *Pandit et al.*



4

5 *Supplementary Figure 1. Trait profile of zoonotic Flavivirus-positive species. Partial dependence plots of the top 10 predictor traits from the generalized boosted regression trees*
 6 *model, arranged in order of variable importance. Lines (continuous variables) and dots (categorical variables) indicate the partial dependence of the variable on the prediction*
 7 *and the histogram indicates the distribution of the variable in the model data. Note: Diet preference: Fish: Fish; fruit: Fruit/drupes; Inv: Invertebrates including aquatic*
 8 *invertebrates; Nect: Nectar pollen, plant exudates, gums; PlantO: Other plant material; Seed: Seeds; Vect: Reptiles, snakes, amphibians; vend: Mammals, Birds; Vunk:*
 9 *Vertebrates-general or unknown*

10



11

12

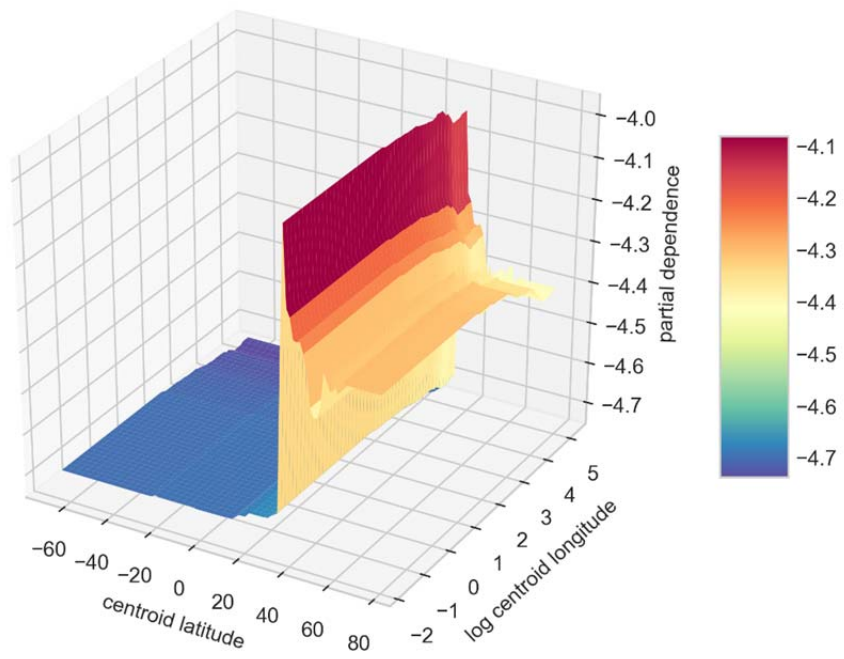
13

14

15

16

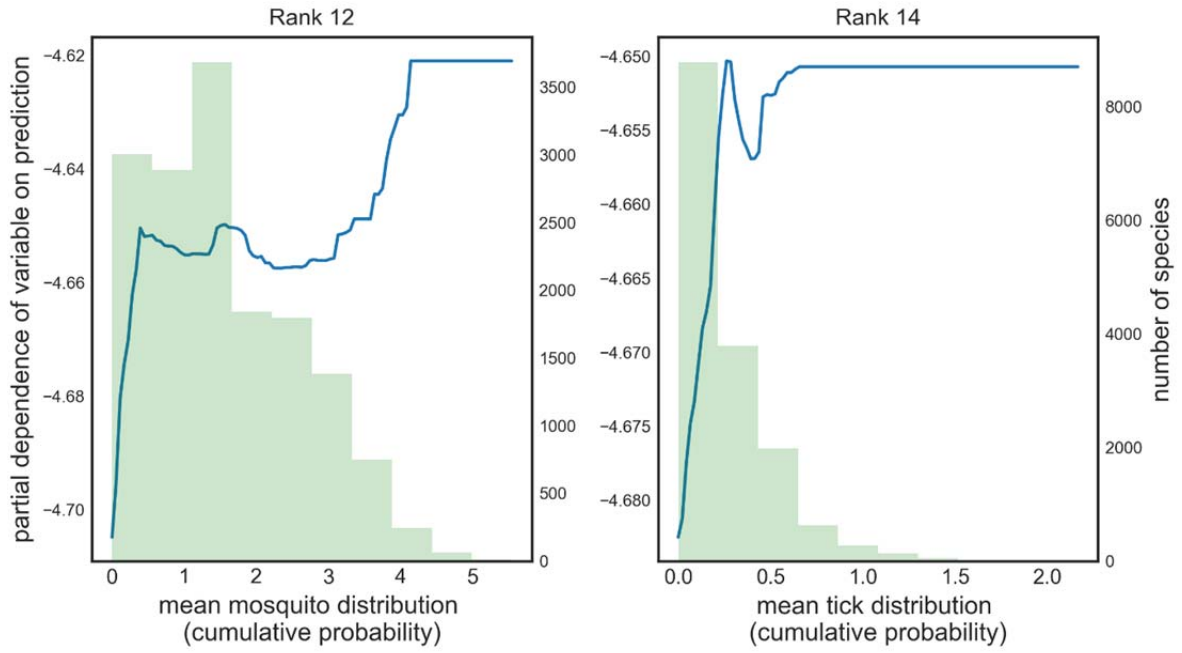
Supplementary Figure 2. Trait profile of zoonotic *Flavivirus*-positive species using a model excluding the variable PubMed hits. Partial dependence plots of the top 10 predictor traits from the generalized boosted regression trees model, arranged in order of variable importance. Lines (continuous variables) and dots (categorical variables) indicate the partial dependence of the variable on the prediction and the histogram indicates the distribution of the variable in the model data. Note: Diet preference: Fish: Fish; fruit: Fruit/drupes; Inv: Invertebrates including aquatic invertebrates; Nect: Nectar pollen, plant exudates, gums; PlantO: Other plant material; Seed: Seeds; Vect: Reptiles, snakes, amphibians; vend: Mammals, Birds; Vunk: Vertebrates-general or unknown



17

18 *Supplementary Figure 3. Partial dependence and interaction of latitude and longitude centroids of vertebrate species for being*
19 *flaviviral host in the generic Flavivirus model.*

20

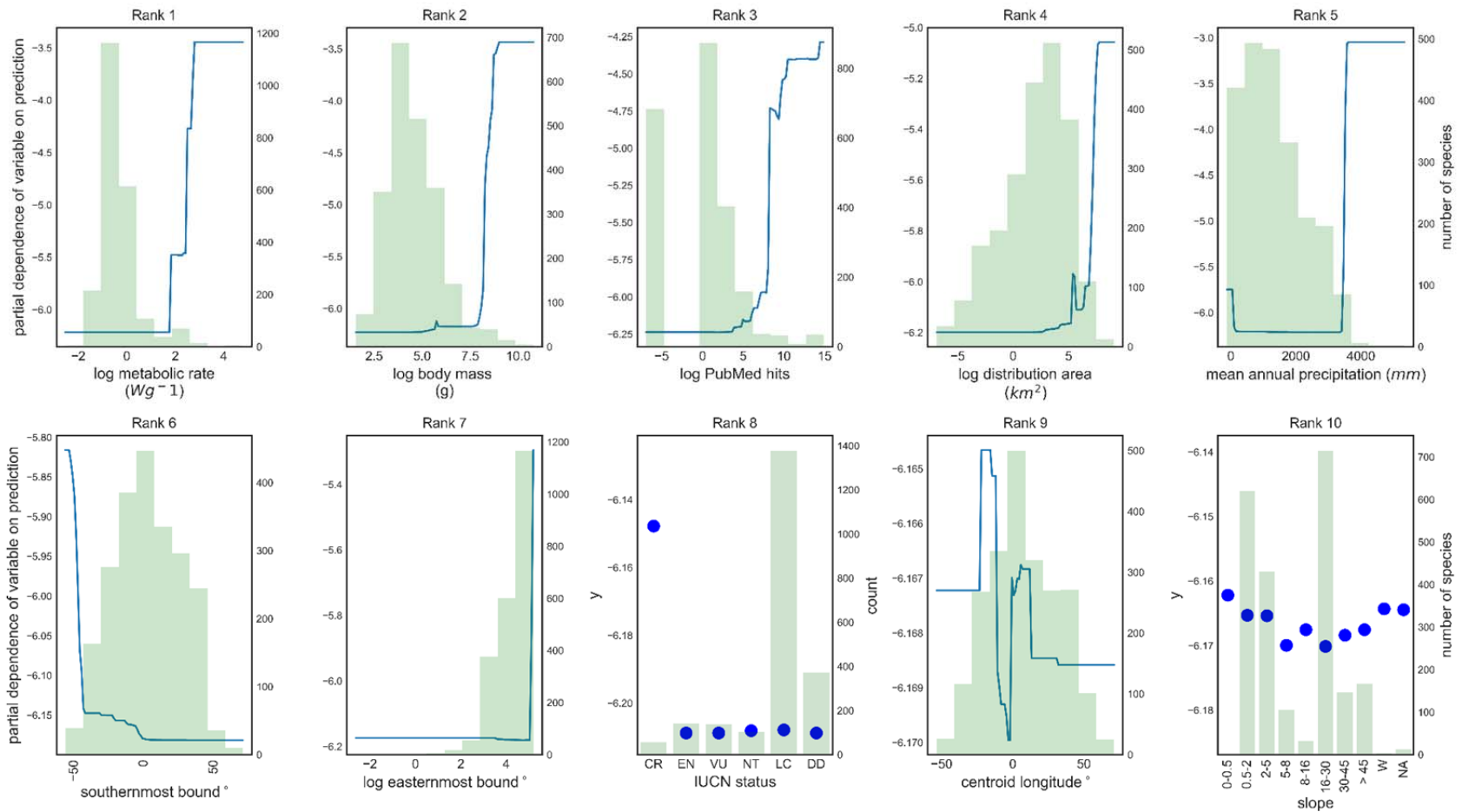


21

22

23

Supplementary Figure 4. Partial dependence of variables related to the distributions of mosquito and tick vectors for the generic *Flavivirus* model.



24

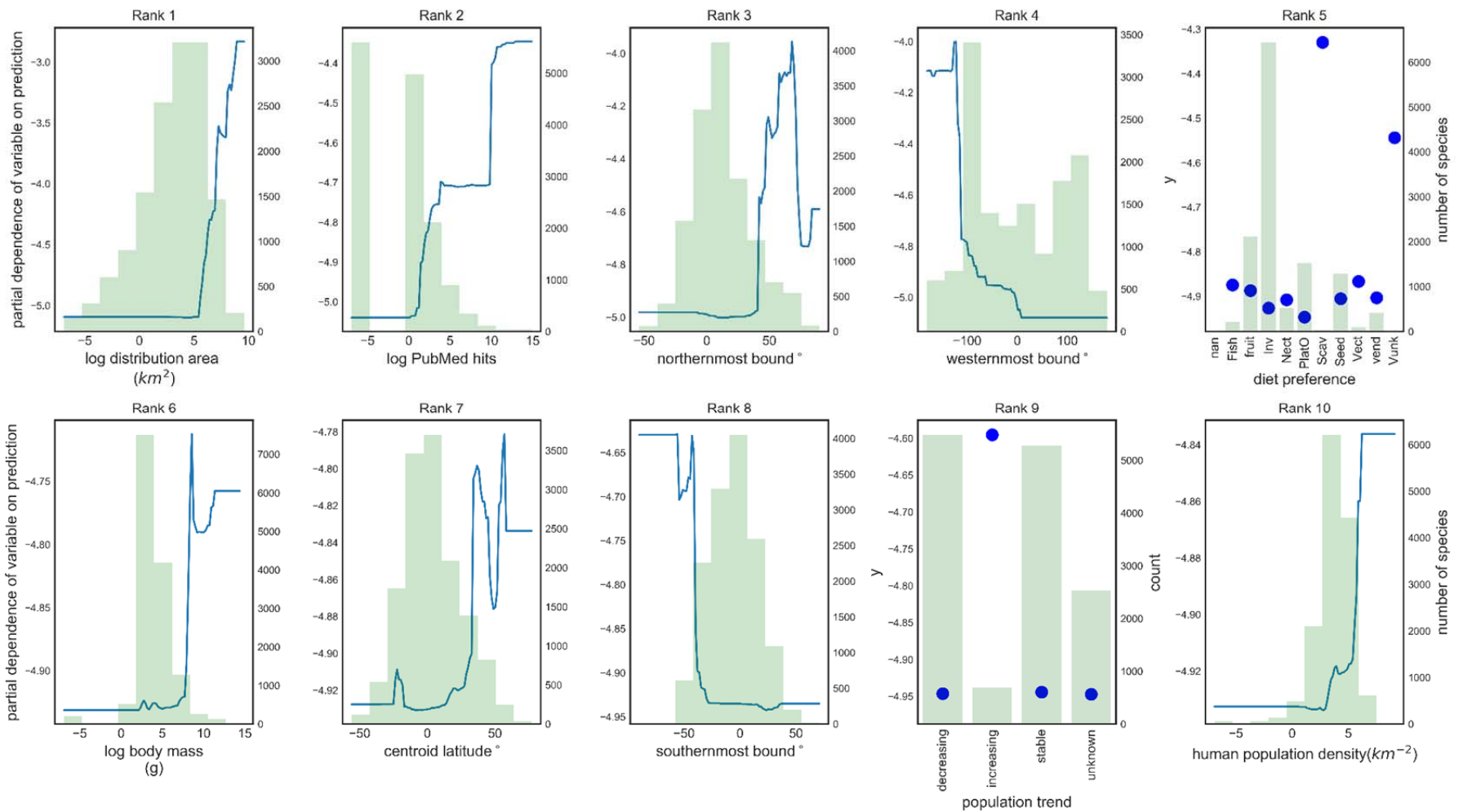
25

26

27

28

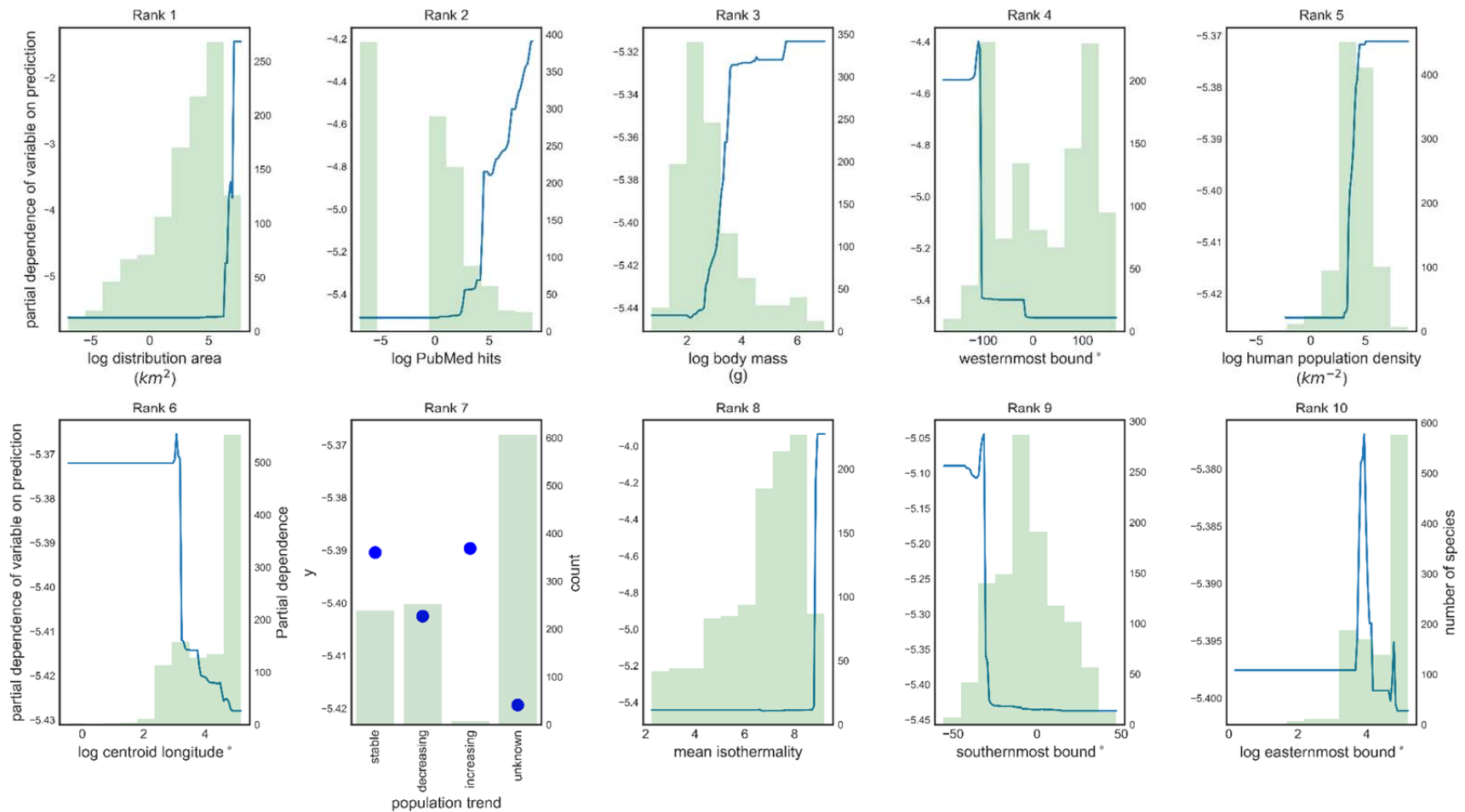
Supplementary Figure 5. Trait profile of YFV- and ZIKV-positive species. Partial dependence plots of the top 10 predictor traits from the generalized boosted regression trees model, arranged in order of variable importance. Lines (continuous variables) and dots (categorical variables) indicate the partial dependence of the variable on the prediction and the histogram indicates the distribution of the variable in the model data.



29

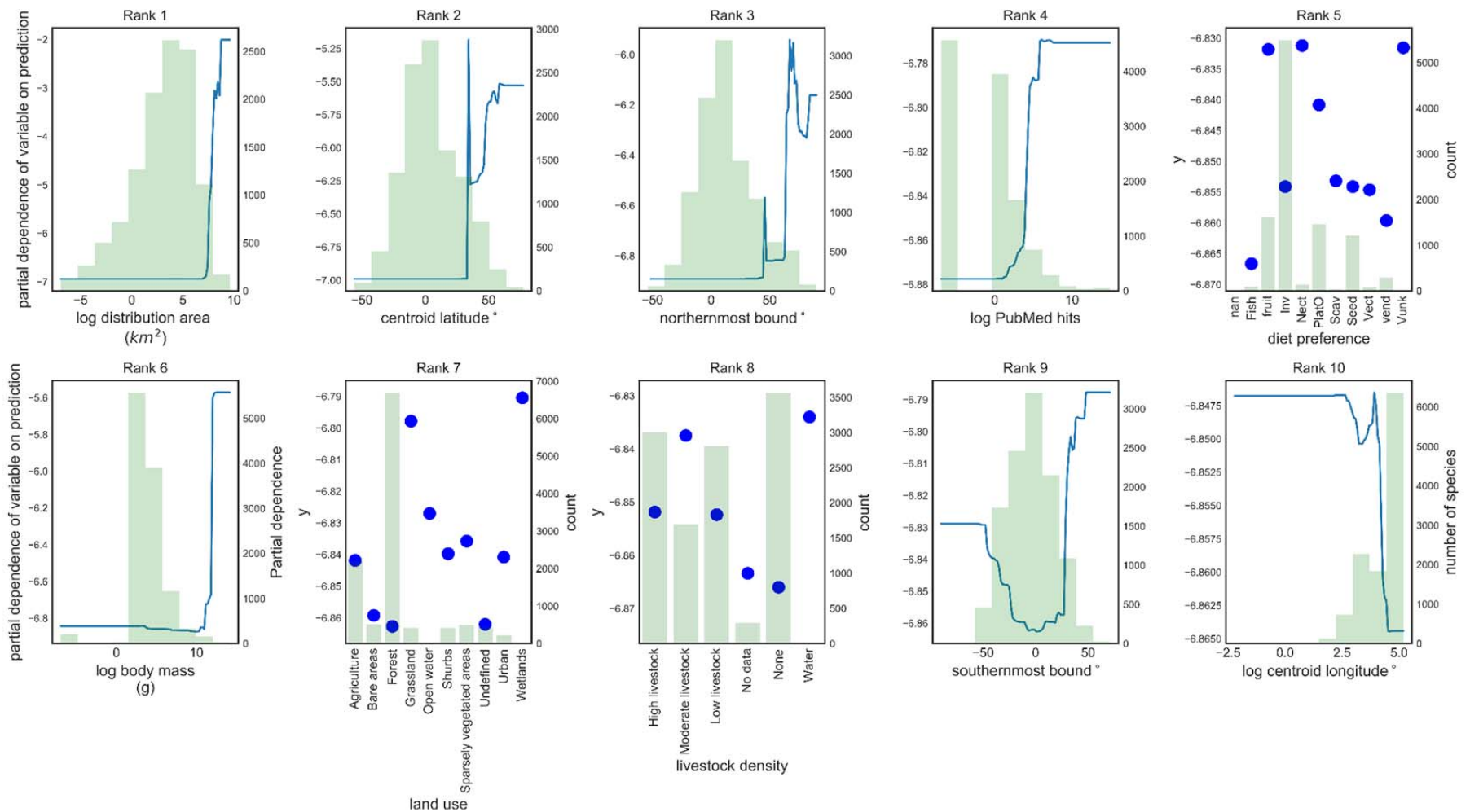
30 *Supplementary Figure 6. Trait profile of WNV, SLEV, and USUV positive species: Partial dependence plots of the top 10 predictor traits from the generalized boosted regression*
 31 *trees model, arranged in order of variable importance. Lines (continuous variables) and dots (categorical variables) indicate the partial dependence of the variable on the*
 32 *prediction and the histogram indicates the distribution of the variable in the model data. Note: Diet preference: Fish: Fish; fruit: Fruit/drupes; Inv: Invertebrates including aquatic*
 33 *invertebrates; Nect: Nectar pollen, plant exudates, gums; PlantO: Other plant material; Seed: Seeds; Vect: Reptiles, snakes, amphibians; vend: Mammals, Birds; Vunk:*
 34 *Vertebrates-general or unknown*

35



36

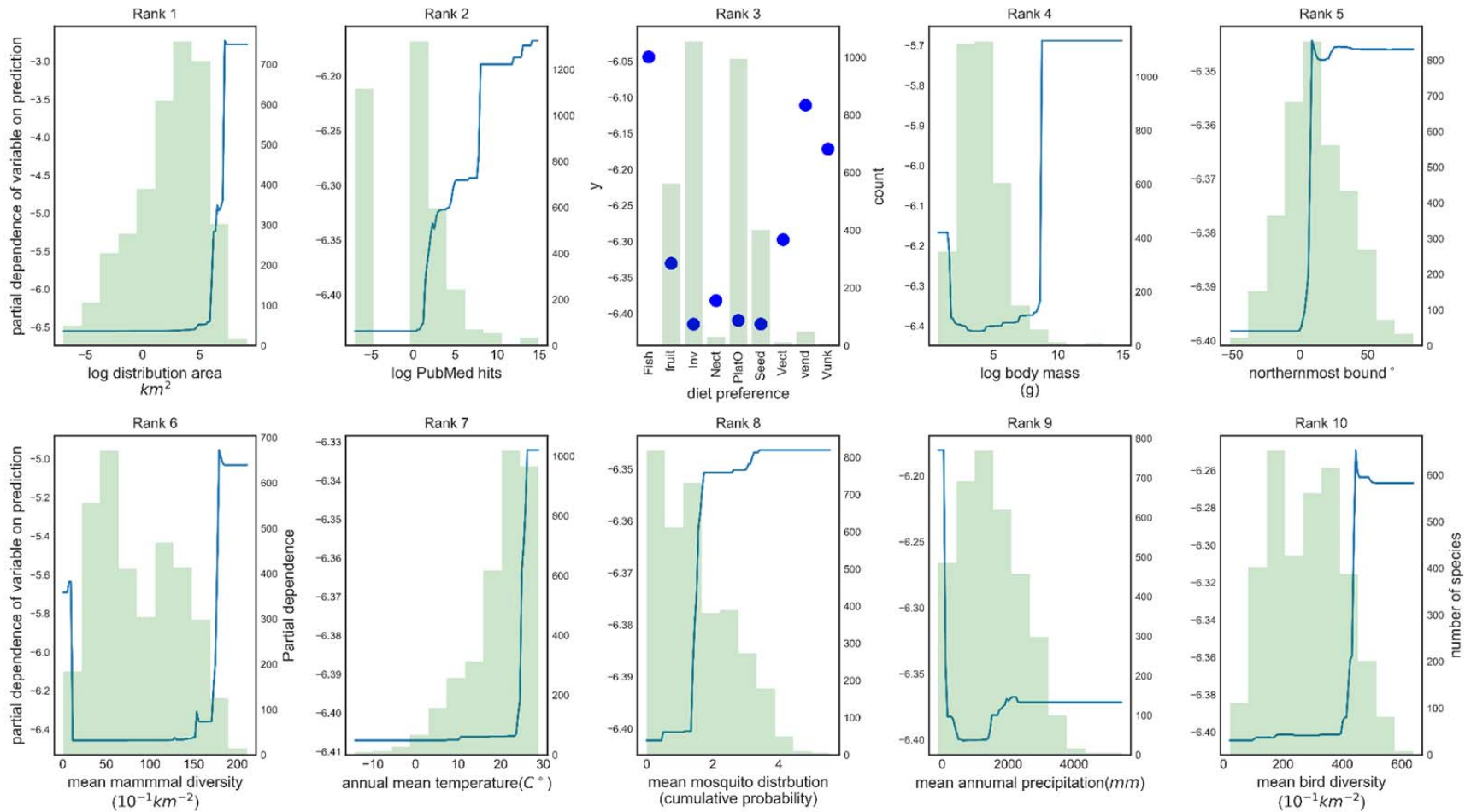
37 *Supplementary Figure 7. Trait profile of RBV, ENTV, and DBV positive species. Partial dependence plots of the top 10 predictor traits from the generalized boosted regression trees*
 38 *model, arranged in order of variable importance. Lines (continuous variables) and dots (categorical variables) indicate the partial dependence of the variable on the prediction*
 39 *and the histogram indicates the distribution of the variable in the model data.*



40

41 *Supplementary Figure 8. Trait profile of TBEV positive species. Partial dependence plots of the top 10 predictor traits from the generalized boosted regression trees model,*
 42 *arranged in order of variable importance. Lines (continuous variables) and dots (categorical variables) indicate the partial dependence of the variable on the prediction and the*
 43 *histogram indicates the distribution of the variable in the model data. Note: Diet preference: Fish: Fish; fruit: Fruit/drupes; Inv: Invertebrates including aquatic invertebrates;*
 44 *Nect: Nectar pollen, plant exudates, gums; PlantO: Other plant material; Seed: Seeds; Vect: Reptiles, snakes, amphibians; vend: Mammals, Birds; Vunk: Vertebrates-general or*
 45 *unknown*

46

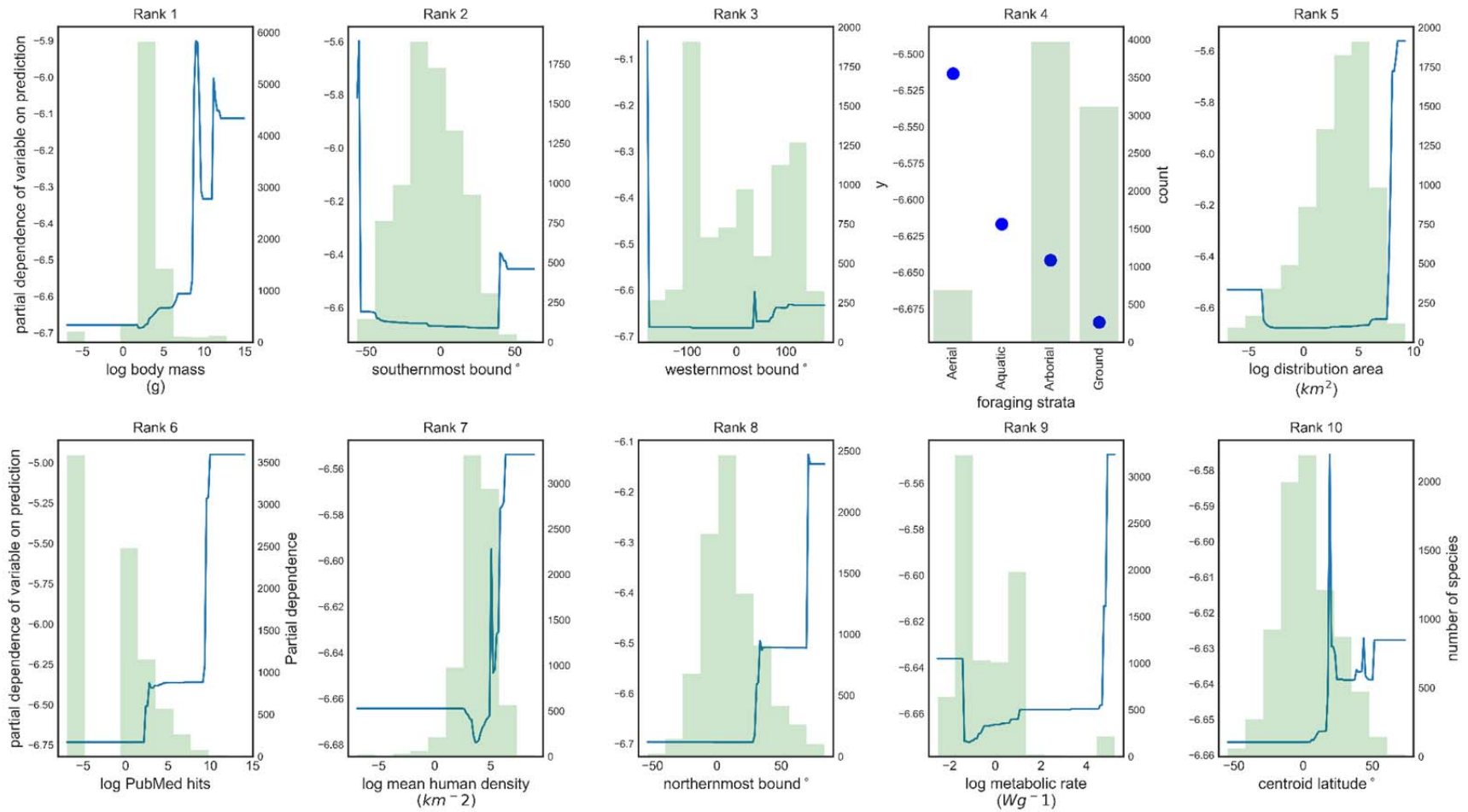


47

48 *Supplementary Figure 9: Trait profile of DENV positive species. Partial dependence plots of the top 10 predictor traits from the generalized boosted regression trees model,*
 49 *arranged in order of variable importance. Lines (continuous variables) and dots (categorical variables) indicate the partial dependence of the variable on the prediction and the*
 50 *histogram indicates the distribution of the variable in the model data. Note: Diet preference: Fish: Fish; fruit: Fruit/drupes; Inv: Invertebrates including aquatic invertebrates;*
 51 *Nect: Nectar pollen, plant exudates, gums; PlantO: Other plant material; Seed: Seeds; Vect: Reptiles, snakes, amphibians; vend: Mammals, Birds; Vunk: Vertebrates-general or*
 52 *unknown*

53

54



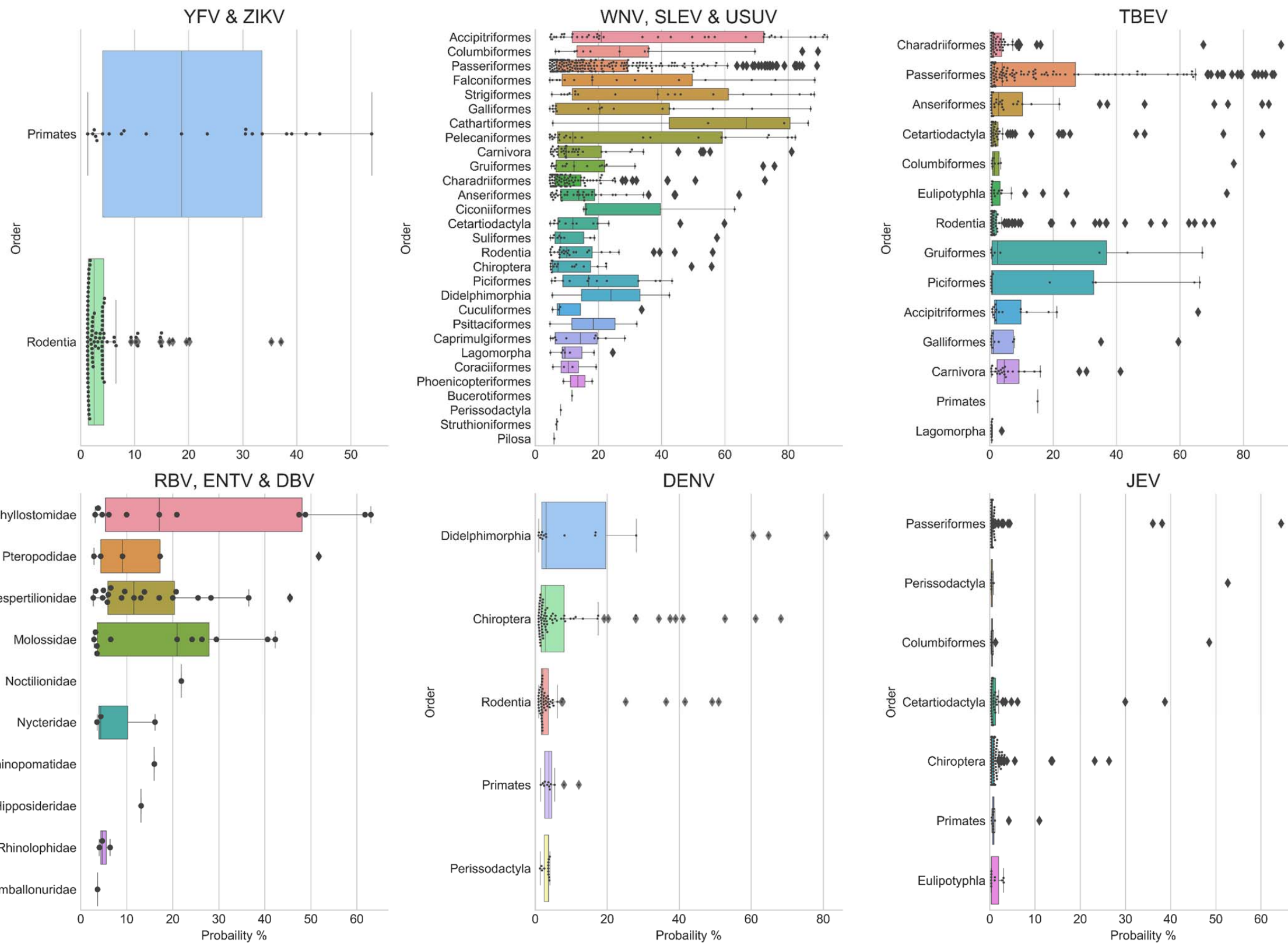
55

56

57

58

Supplementary Figure 10. Trait profile of JEV positive species. Partial dependence plots of the top 10 predictor traits from the generalized boosted regression trees model, arranged in order of variable importance. Lines (continuous variables) and dots (categorical variables) indicate the partial dependence of the variable on the prediction and the histogram indicates the distribution of the variable in the model data.



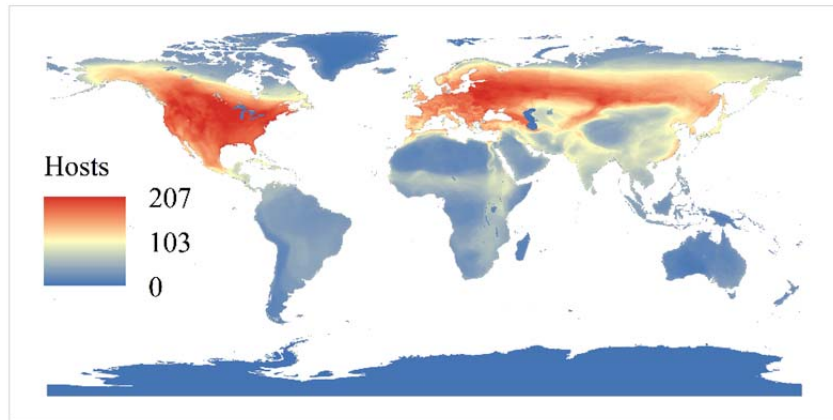
Supplementary Figure 12. Model-predicted probabilities. Boxplots show the distribution of model-predicted probabilities for species from modeled orders; dots show data points for species and diamonds show outlier species.

62

63

64

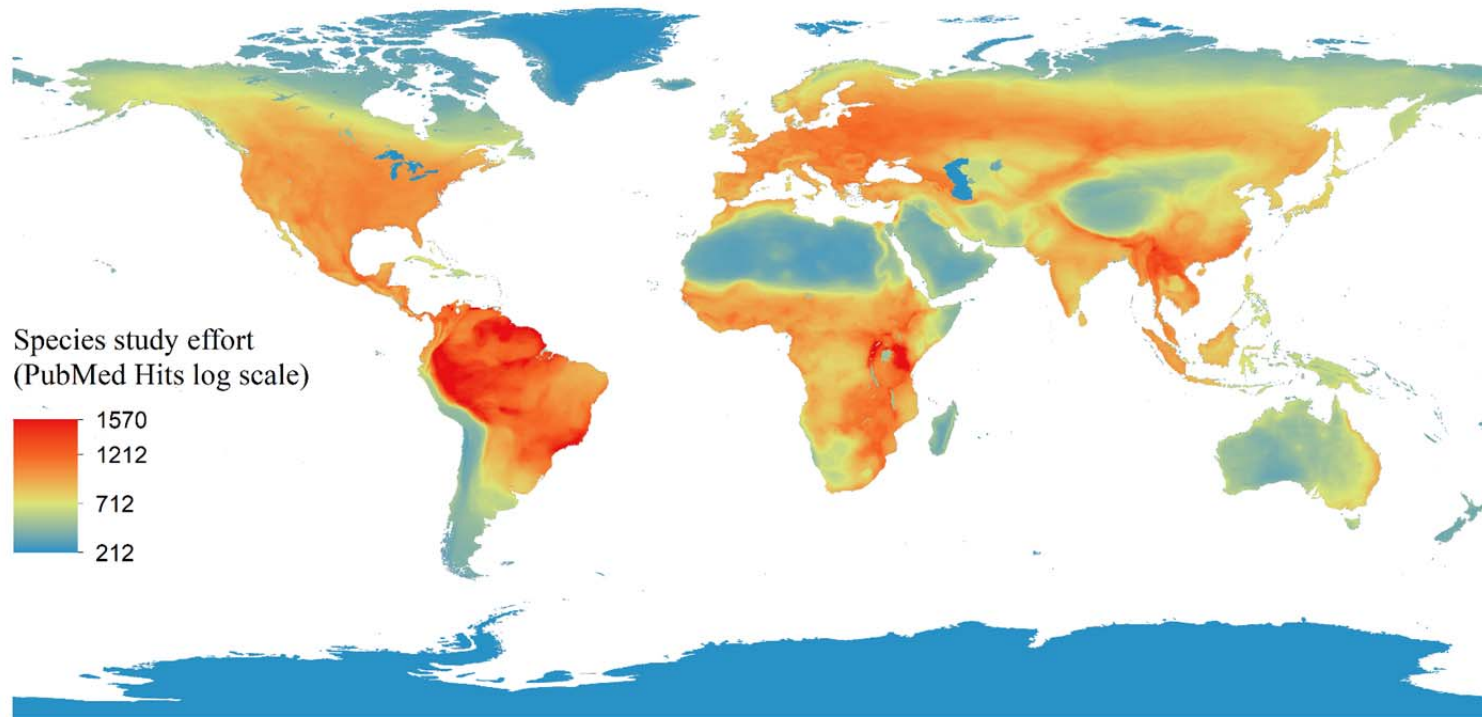
WNV, SELV, USUV (Virus isolation and PCR only)



65

66 *Supplementary Figure 13: Overlapping geographical ranges of model predicted hosts in the 95th percentile of probability for*
67 *WNV, SLEV and USUV by relabeling positive hosts only when they are found positive by PCR or virus isolation. Map was*
68 *generated using species distribution data from IUCN¹, and BirdLife International and NatureServe².*

69



71

72

73

Supplementary Figure 14: Geographical distribution of study effort shown by overlapping geographical ranges of all the mammalian and avian species and their associated studies reported in PubMed Central database. Map was generated using species distribution data from IUCN¹, and BirdLife International and NatureServe².

74 *Supplementary Tables*

75

76

77

Supplementary Table 1. Model parameters and the optimal number of trees for different models

<i>Model</i>	<i>Positive samples</i>		<i>Total samples</i>		<i>Optimal Number of trees</i>	<i>cv AUC score (SE)</i>	<i>Holdout AUC</i>
	<i>Training</i>	<i>Holdout</i>	<i>Training</i>	<i>Holdout</i>			
<i>Generic</i>	305	76	12236	3056	9250	0.906 (0.011)	0.923
<i>Generic (without PubMed hits)</i>	305	76	12236	3056	11200	0.894 (0.01)	0.924
<i>Group 1: YFV, ZIKV</i>	11	3	1774	444	1400	0.944 (0.03)	0.975
<i>Group 2: WNV, SLEV, USUV</i>	203	51	11112	2778	5850	0.903 (0.008)	0.938
<i>Group 2: WNV, SLEV, USUV (PCR/virus isolation)</i>	118	30	11197	2799	6450	0.924 (0.011)	0.892
<i>TBEV</i>	62	15	9038	2260	4400	0.978 (0.004)	0.972
<i>Group 3: RBV, ENTV, DBV</i>	13	3	871	218	1700	0.929 (0.031)	0.662
<i>DENV</i>	18	5	2740	685	2200	0.909 (0.041)	0.877
<i>JEV</i>	16	4	6501	1625	1250	0.834 (0.046)	0.992

78 *Note: tree complexity: 5, learning rate: 0.001, bag fraction: 0.5, cv: 10*

79 **Supplementary Table 2. Taxonomic Orders known to have confirmed species for viruses from**
80 **each virus group that are used for predictions in respective models**

Virus Groups	Orders
<i>Group 1: YFV, ZIKV</i>	<i>Primates, Rodentia</i>
<i>Group 2: WNV, SLEV, USUV</i>	<i>Accipitriformes, Passeriformes, Charadriiformes, Strigiformes, Galliformes, Primates, Anseriformes, Pelecaniformes, Chiroptera, Rodentia, Cetartiodactyla, Pilosa, Bucerotiformes, Carnivora, Caprimulgiformes, Falconiformes, Piciformes, Columbiformes, Gruiformes, Cathartiformes, Suliformes, Perissodactyla, Phoenicopteriformes, Cuculiformes, Psittaciformes, Coraciiformes, Ciconiiformes, Struthioniformes, Lagomorpha, Didelphimorphia</i>
<i>Group 2: WNV, SLEV, USUV (PCR/Isolation only)</i>	<i>Accipitriformes, Passeriformes, Charadriiformes, Strigiformes, Galliformes, Primates, Anseriformes, Pelecaniformes, Chiroptera, Rodentia, Cetartiodactyla, Bucerotiformes, Carnivora, Caprimulgiformes, Falconiformes, Piciformes, Columbiformes, Gruiformes, Cathartiformes, Suliformes, Perissodactyla, Phoenicopteriformes, Psittaciformes, Coraciiformes, Ciconiiformes, Struthioniformes, Didelphimorphia</i>
<i>TBEV</i>	<i>Rodentia, Cetartiodactyla, Eulipotyphla, Primates, Lagomorpha, Carnivora, Passeriformes, Piciformes, Galliformes, Charadriiformes, Anseriformes, Gruiformes, Accipitriformes, Columbiformes</i>
<i>Group 3: RBV, ENTV, DBV</i>	<i>Chiroptera</i>
<i>DENV</i>	<i>Chiroptera, Didelphimorphia, Perissodactyla, Primates, Rodentia</i>
<i>JEV</i>	<i>Cetartiodactyla, Columbiformes, Eulipotyphla, Perissodactyla, Chiroptera, Passeriformes, Primates</i>

81 *Note: Blue: avian orders*

82 **Supplementary Table 3. Feature dataset used in the generalized boosted regression model for predicting flavivirus potential hosts**

Model Category	Variable	Variable Type	Description	Source
Conservation Status	<i>IUCN status</i>	<i>categorical</i>		1
	<i>Population Trend</i>	<i>categorical</i>		1
Ecology	<i>Habitat breadth</i>	<i>numeric</i>	<i>Number of habitats</i>	3
	<i>Diet preference</i>	<i>categorical</i>	<i>predominant diet type</i>	4
	<i>Foraging strata</i>	<i>categorical</i>	<i>Foraging Strata</i>	4
	<i>Activity time</i>	<i>categorical</i>	<i>diurnal-nocturnal</i>	4
Physiology	<i>Metabolic rate (W)</i>	<i>numeric</i>	<i>metabolic rate</i>	4
	<i>Body temperature (°C)</i>	<i>numeric</i>	<i>body temperature</i>	4
	<i>Body mass (g)</i>	<i>numeric</i>	<i>body mass</i>	4
Distribution	<i>Centroid (X, Y)</i>	<i>numeric</i>	<i>centroids of distribution shapefile</i>	
	<i>Northernmost bound</i>	<i>numeric</i>	<i>Y maximum value</i>	-
	<i>Southernmost bound</i>	<i>numeric</i>	<i>Y minimum value</i>	-
	<i>Easternmost bound</i>	<i>numeric</i>	<i>X maximum value</i>	-
	<i>Westernmost bound</i>	<i>numeric</i>	<i>X minimum value</i>	-
	<i>Distribution Area (km⁻²)</i>	<i>numeric</i>	<i>area of the distribution shapefile</i>	-

83

84

85 **Supplementary Table 4. Model variables calculated from global raster datasets**

Model Category	Variable	Variable Type	Description	Source	Statistic
Bio-Climatic	<i>Annual precipitation (mm)</i>	<i>numeric</i>		5	<i>mean</i>
	<i>Annual Mean Temperature (°C)</i>	<i>numeric</i>		5	<i>mean</i>
	<i>Isothermality</i>	<i>numeric</i>	<i>comparison of daily difference to the yearly difference</i>	5	<i>mean</i>
	<i>Slope</i>	<i>categorical</i>		6	<i>mode</i>
Diversity	<i>Mammal diversity (n×10 km⁻²)</i>	<i>numeric</i>	<i>mammalian diversity in the range of species</i>	7	<i>mean</i>
	<i>Bird diversity (n×10 km⁻²)</i>	<i>numeric</i>	<i>avian diversity in the range of species</i>	7	<i>mean</i>
Anthropogenic	<i>Human population density (km⁻²)</i>	<i>numeric</i>	<i>Human population diversity in the range of species</i>	6	<i>mean</i>

variables	²⁾				
	<i>Crops</i>	<i>categorical</i>		6	<i>mode</i>
	<i>Livestock density</i>	<i>categorical</i>		6	<i>mode</i>
	<i>Land use</i>	<i>categorical</i>		6	<i>mode</i>
Vectors	<i>Mosquitoes</i>	<i>numeric</i>	<i>additive distribution of mosquito vectors (probability × 10)</i>	8	<i>mean</i>
	<i>Ticks</i>	<i>numeric</i>	<i>additive distribution of tick vectors (probability × 10)</i>		<i>mean</i>

86

87

88

89

90

91

92

93

Supplementary Table 5: Known Flavivirus mosquitoes and tick vectors and viruses carried by these vectors

Vector	Number of viruses	Viruses						
<i>Aedes aegypti</i>	7	DENV	ILHV	SLEV	WSLV	WNV	YFV	ZIKV
<i>Culex pipiens</i>	6	JEV	SLEV	TMUV	USUV	WNV	ZIKV	
<i>Culex tritaeniorhynchus</i>	6	BAGV	DENV	JEV	TMUV	WSLV	WNV	
<i>Aedes albopictus</i>	5	DENV	MVEV	USUV	WNV	ZIKV		
<i>Aedes vexans</i>	5	JEV	TBEV	USUV	WSLV	WNV		
<i>Culex quinquefasciatus</i>	5	JEV	SLEV	TMUV	WNV	ZIKV		
<i>Aedes africanus</i>	5	BOUV	WSLV	WNV	YFV	ZIKV		
<i>Culex annulirostris</i>	4	JEV	KUNV	MVEV	WNV			
<i>Psorophora ferox</i>	4	ILHV	ROCV	SLEV	WNV			
<i>Mansonia uniformis</i>	4	JEV	WSLV	WNV	ZIKV			
<i>Ixodes persulcatus</i> *	4	LIV	OMSKV	TBEV				
<i>Culex univittatus</i>	3	USUV	WSLV	WNV				
<i>Culex vishnui</i>	3	JEV	TMUV	WNV				
<i>Culex pseudovishnui</i>	3	JEV	KUNV	TMUV				
<i>Dermacentor marginatus</i> *	3	OMSKV	TBEV	WNV				
<i>Culex bitaeniorhynchus</i>	3	JEV	MVEV	WNV				
<i>Aedes scapularis</i>	3	ILHV	ROCV	SLEV				
<i>Sabethes chloropterus</i>	3	ILHV	SLEV	YFV				
<i>Hyalomma marginatum</i> *	3	LIV	TBEV	WNV				
<i>Aedes luteocephalus</i>	3	WSLV	YFV	ZIKV				
<i>Anopheles hyrcanus</i>	3	JEV	WNV					
<i>Amblyomma variegatum</i> *	2	WNV	YFV					
<i>Rhipicephalus muhsamae</i> *	2	WSLV	WNV					
<i>Ixodes ricinus</i> *	2	LIV	TBEV					
<i>Dermacentor reticulatus</i> *	2	OMSKV	TBEV					
<i>Amblyomma cajennense</i> *	1	CPCV						

95 *Tick species.

96 **Note:** APOIV: Apoi virus, BAGV: Bagaza virus, BANV: Banzi virus, BOUV: Bouboui virus, CPCV: Cacipacore virus, DBV: Dakar bat virus, DENV:

97 Dengue virus, ENTV: Entebbe bat virus, IGUV: Iguape virus, ILHV: Ilheus virus, JEV: Japanese encephalitis virus, KFDV: Kyasanur forest virus, KUNV:

98 Kunjin virus, LIV: Louping-ill virus, MODV: Modoc virus, MVEV: Murray Valley encephalitis virus, OMSKV: Omsk hemorrhagic fever virus, RBV: Rio

99 *Bravo virus, ROCV: Rocio virus, SLEV: St. Louis Encephalitis virus, TBEV: Tickborne encephalitis virus, TMUV: Tembusu virus, UGSV: Uganda S virus,*
100 *USUV: Usutu virus, WNV: West Nile virus, WSLV: Wesselsbron virus, YFV: Yellow fever virus, ZIKV: Zika virus.*

101

102 *Supplementary References*

103 1. IUCN. The IUCN Red List of Threatened Species. Version 2017-1 edn (2017).

104

105 2. BirdLife International and Nature Serve. Bird species distribution maps of the world. BirdLife International, Cambridge, UK
106 and NatureServe, Arlington, USA. (2015).

107

108 3. Vilela B, Villalobos F. letsR: a new R package for data handling and analysis in macroecology. *Methods Ecol Evol* **6**, 1229-
109 1234 (2015).

110

111 4. Wilman H, Belmaker J, Simpson J, de la Rosa C, Rivadeneira MM, Jetz W. EltonTraits 1.0: Species-level foraging attributes
112 of the world's birds and mammals. *Ecology* **95**, 2027-2027 (2014).

113

114 5. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land
115 areas. *Int J Climatol* **25**, 1965-1978 (2005).

116

117 6. FAO. FAO Geonetwork. Food and Agriculture Organization of the United Nations (2017).

118

119 7. Jenkins CN, Pimm SL, Joppa LN. Global patterns of terrestrial vertebrate diversity and conservation. *Proc Natl Acad Sci U S A*
120 **110**, E2602-2610 (2013).

121

122 8. The Walter Reed Biosystematics Unit. VectorMap. <http://vectormap.si.edu/>. Smithsonian Institution (2016).

123

124