The geographic distribution and ecological preferences of the American dog tick, *Dermacentor variabilis* (Say), in the U.S.A.

A. M. J A M E S[1], C. B U R D E T T[2], M. J. M C C O O L[1], A. F O X[1]
and P. R I G G S[1]

[1]U.S. Department of Agriculture, Animal and Plant Health Inspection Service, Veterinary Services, Fort Collins, CO, U.S.A. and
[2]Department of Biology, Colorado State University, Fort Collins, CO, U.S.A.

Abstract. Equine piroplasmosis (EP), caused by two parasitic organisms, *Theileria equi* and *Babesia caballi*, is a tick-borne disease of recent concern in horses in the U.S.A. Outbreaks of EP have been detected in Florida, Missouri, Kansas and Texas. In 2009, EP transmission in Texas occurred through the adults of two tick species, *Amblyomma mixtum* [formerly known as *Amblyomma cajennense* (Fabricius, 1787)] Koch (Ixodida: Ixodidae) and *Dermacentor variabilis* (Say) (Ixodida: Ixodidae), the American dog tick (ADT). In this study, we developed a continent-scale map for the distribution of the EP vector species *D. variabilis*, using a presence-only modelling approach to assess the habitat preferences of this tick. We used identification records from our tick geodatabase of locations in which the presence of the ADT had been noted. The potential distribution of the ADT in the U.S.A. was estimated from environmental factors using the maximum entropy approach based on localities in which there is a high probability of occurrence according to habitat suitability. Elevation and temperature were found to be biologically significant environmental variables influencing the presence of this tick species. Properly designed and constructed probability surfaces using maximum entropy offer one useful approach to the mapping of distribution ranges of tick species based on suitable habitat in the U.S.A.

Key words. *Dermacentor variabilis*, ecological modelling, equine piroplasmosis, habitat suitability, landscape, maximum entropy, Rocky Mountain spotted fever.

Introduction

Our ability to predict species distributions is increasingly important as many distributions shift according to ongoing human-induced changes in land use, resource availability and climate (Estrada-Peña *et al*., 2006; Gray *et al*., 2009). It is a central premise of biogeography that climate has particularly strong effects on the distributions of many species at biogeographic scales. As biogeographic scales preclude experimentation, models for species distribution – constructed using available locational data – have become widely used to develop and test

hypotheses about how ecological factors may affect species distributions. These models can explore the relationships between species occurrence and sets of predictor variables and produce different, but useful outputs for assessing land use and climatic effects (Estrada-Peña *et al*., 2006). Many modelling techniques have been used in this assessment, including generalized linear modelling (GLM), generalized additive modelling (GAM), classification and regression tree (CART) modelling, and principal components analysis (PCA) (Franklin, 2010).

The importance of understanding the factors influencing the broad-scale distribution of a species is even more critical for

Correspondence: Angela M. James, Center for Epidemiology and Animal Health, U.S. Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) Veterinary Service, Building B, 2150 Center Avenue, Fort Collins, CO 80524, U.S.A. Tel.: + 1 970 494 7278; Fax: + 1 970 494 7269; E-mail: Angela.M.James@aphis.usda.gov

species that pose threats to animal and human health. For example, the number of tick-borne diseases has risen dramatically over the last 30 years and more than 10 newly recognized spotted fever rickettsioses have been recorded since 1984 (Paddock & Telford, 2011). There is concern that climate changes may alter tick–host–pathogen interactions, expansion of tick ranges and the potential for the establishment of new tick species in an area (Gage *et al.*, 2008; Porretta *et al.*, 2013). Gray *et al.* (2009) indicated that several tick species have expanded their ranges into new areas with warming temperatures in Europe. Both biotic (i.e. unrestricted animal movement, abundance of tick hosts, animal management practices) and abiotic (i.e. human-induced land cover changes, habitat structure, temperatures) factors are likely to influence changes in tick distributions on a global scale (Estrada-Peña *et al.*, 2012).

The characterization of environmental factors that are suitable for a tick species is the most commonly used strategy to estimate the geographic range of that species, identify areas where tick species may become established, and assess potential disease risk. Efforts have been made at global, regional and local scales to characterize the major environmental factors influencing tick distributions using statistical and/or correlative approaches (Estrada-Peña *et al.*, 2012). Logistic regression modelling methods have been used to create habitat suitability models for an African tick species and the vector of Lyme disease in North America (Cummings, 2000; Brownstein *et al.*, 2005). A bioclimate envelope modelling method has been used to describe the distribution of the southern cattle tick, and the maximum entropy method (Maxent) has been used to estimate habitat suitability for six ixodid tick species in the Mediterranean region (Estrada-Peña *et al.*, 2006; Torina *et al.*, 2008).

In the present study, we used the maximum entropy approach to model the potential distribution of the American dog tick, *Dermacentor variabilis*, an important vector of several pathogens to humans, domestic animals and wildlife. These pathogens include *Rickettsia rickettsii*, the causative agent of Rocky Mountain spotted fever (RMSF), *Anaplasma marginale*, the agent responsible for bovine anaplasmosis, *Francisella tularensis*, the causative agent of tularaemia, and *Theileria equi*, a causative agent of equine piroplasmosis (Scoles *et al.*, 2011). We evaluated the relative performance of this approach in predicting the distribution of this tick species in the U.S.A. and in identifying important environmental factors that may influence its present distribution. The model output from this approach will be helpful in assessing potential changes in the distribution of the American dog tick (ADT) that may arise in response to changes in climate or land use, and in identifying sites where animals or humans may be at increased risk for exposure to this tick species.

Materials and methods

Tick species data sources

The model developed for this study is based on a set of identification records for the ADT, *D. variabilis*, obtained from the Smithsonian Institution National Museum of Natural History. Most datasets from natural history museums include large numbers of records that confirm the presence of the species and data on its absence are generally scarce, especially in regions where little sampling has occurred. When absence data are available, they may be of limited value because sampling intensity in space and time can be highly clustered. In addition to presence-only data, sampling bias is inherent in museum datasets as some areas are more heavily sampled as a result of their ease of access or an increased likelihood of obtaining a sample that depends on pre-existing knowledge about the ecological preferences of a tick species. Most tick species represented in the database are known from only a few localities at a national scale and are unlikely to reflect the full range of the species within the U.S.A.

We developed a geographic information system (GIS)-based framework that integrates two national tick identification datasets into a single geodatabase for mapping tick species distributions. The first dataset was derived from a tick museum dataset from the Smithsonian National Museum of Natural History. This dataset represents individual surveys conducted over several decades by multiple researchers using different sampling techniques and intensity, and periods of time. The second and unrelated national tick identification dataset was derived from the National Veterinary Service Laboratories (NVSL) and represents a veterinary passive surveillance system for ticks based on their importance to animal agriculture. The museum dataset was used to develop our distribution model and the agricultural dataset was used to independently validate the distribution model. A total of 1695 ADT records collected between 1942 and 1998 were selected from the museum dataset. All of these records included adequate reference to spatial coordinates (latitude and longitude coordinates were specified) or provided detailed information on locality sufficient to determine referenced latitude and longitude coordinates using the Getty Thesaurus. A total of 841 tick museum records remained after duplicate records had been removed. We also removed all records older than 1950, which gave us a final dataset of 336 points representing 216 counties within 42 states (Table 1 and Fig. 1A). Records with identical year reports were chosen based on host associations and ranges of those hosts, with smaller sized host and smaller host ranges chosen first. These localities, chosen to create the distribution model, all corresponded to the described regional distribution of the ADT within the U.S.A. as reported in publications by other collectors (McEnroe, 1979a; Sonenshine, 1991).

Environmental data sources

For the environmental data, we used 24 different environmental variables that may influence the distribution of North American tick populations. Our variables included vegetation indices, topographic variables, and climatic variables from 1971 to 2000 coinciding with the time period of the tick species collection data points (i.e. both datasets were matched in temporal scale). Environmental layers were processed using ESRI ArcGIS software to resample and reclassify all layers into the same spatial extent (i.e. continental U.S.A.) and 1-km spatial resolution. The spatial resolution of the climate layers was

Table 1. Characteristics of American dog tick, *Dermacentor variabilis*, samples submitted to the Smithsonian National Museum for Natural History during 1950–1998.

Number and stage of submitted samples	Number of records based on host or vegetation	U.S. coverage	Number of records based on locality
1397, males	159, vegetation	42 states	141 mile marker or latitude/longitude coordinates
1203, females	90, wildlife	216 counties	112 national park and wildlife research station centroids*
4320, adults	59, livestock	—	68 city centroids*
147, larvae and nymphs	28, humans	—	15 island centroids*

*Latitude and longitude coordinates based on Getty Thesaurus descriptions.
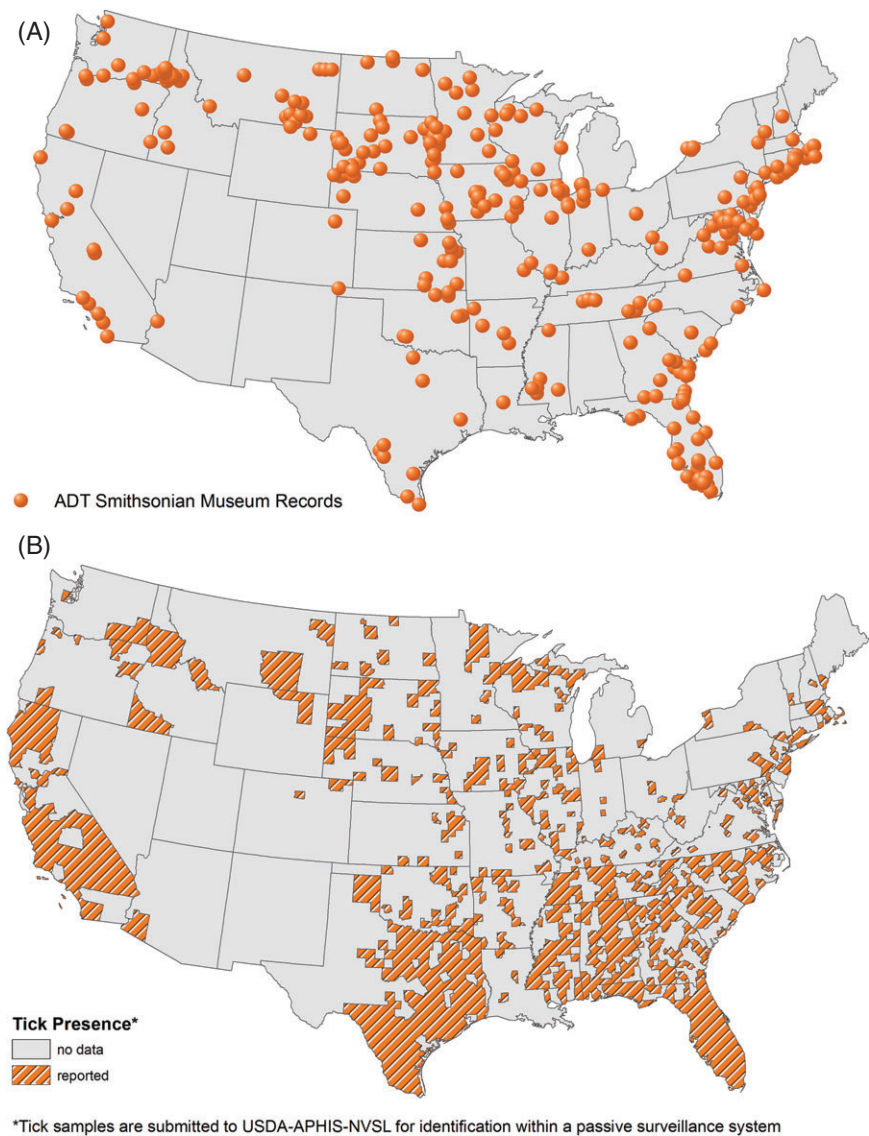

Fig. 1. (A) Map showing Smithsonian National Museum of Natural History records (point localities) for the American dog tick (ADT), *Dermacentor variabilis*, used to assess the potential distribution of this tick species in the U.S.A. (B) Reported distribution of *D. variabilis* within the continental U.S.A., 1998–2011. All tick samples submitted to U.S. Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) National Veterinary Service Laboratories (NVSL) are for identification purposes and are submitted within a passive surveillance system.

Table 2. Description of the bioclimatic variables used in the development of the distribution model for the American dog tick, *Dermacentor variabilis*.

Bioclimatic variable	Definition
1. Annual mean temperature	Mean of all weekly mean temperatures
	Each weekly mean temperature is the mean of that week's maximum and minimum temperatures
2. Mean diurnal range [mean (period max−min)]	Mean of all weekly diurnal temperature ranges
	Each weekly diurnal range is the difference between that week's maximum and minimum temperatures
3. Isothermality 2/7	Mean diurnal range (parameter 2) divided by annual temperature range (parameter 7)
4. Temperature seasonality	ANUCLIM (cov = TRUE) returns the temperature coefficient of variation (CV) as the SD of the weekly mean temperatures expressed as a percentage of the mean of those temperatures (i.e. the annual mean)
	For this calculation, the mean in degrees Kelvin is used
	This avoids the possibility of having to divide by zero, but does mean that the values are usually quite small
	Worldclim (cov = FALSE) returns the t SD of the weekly mean temperatures
5. Maximum temperature of warmest period	Highest of any weekly maximum temperature
6. Minimum temperature of coldest period	Lowest of any weekly minimum temperature
7. Temperature annual range (5−6)	Difference between the max temperature of warmest period (parameter 5) and the min temperature of coldest period (parameter 6)
8. Mean temperature of wettest quarter	The wettest quarter of the year is determined (to the nearest week) and the mean temperature of this period is calculated
9. Mean temperature of driest quarter	The driest quarter of the year is determined (to the nearest week) and the mean temperature of this period is calculated
10. Mean temperature of warmest quarter	The warmest quarter of the year is determined (to the nearest week) and the mean temperature of this period is calculated
11. Mean temperature of coldest quarter	The coldest quarter of the year is determined (to the nearest week) and the mean temperature of this period is calculated
12. Annual precipitation	Sum of all monthly precipitation values
13. Precipitation of wettest period	Precipitation of the wettest week or month, depending on the time step
14. Precipitation of driest period	Precipitation of the driest week or month, depending on the time step
15. Precipitation seasonality (CV)	The coefficient of variation (CV) is the SD of weekly precipitation estimates expressed as a percentage of the mean of those estimates (i.e. annual mean)
16. Precipitation of wettest quarter	The wettest quarter of the year is determined (to the nearest week) and the total precipitation over this period is calculated
17. Precipitation of driest quarter	The driest quarter of the year is determined (to the nearest week) and total precipitation over this period is calculated
18. Precipitation of warmest quarter	The warmest quarter of the year is determined (to the nearest week) and total precipitation over this period is calculated
19. Precipitation of coldest quarter	The coldest quarter of the year is determined (to the nearest week) and total precipitation over this period is calculated

SD, standard deviation.

resampled from 935 m to 1 km and elevation from 800 m to 1 km, resulting in some smoothing of the data, but at a minimal level. The normalized difference vegetation index (NDVI) dataset was imported at 8 km and resampled to 1 km, which resulted in the creation of some redundant information with this layer; however, no data were lost. The PRISM Climate Group's (http://prism.oregonstate.edu) interpolated average monthly precipitation and temperature data values were processed using an aml script to yield the 19 bioclimatic variables (Table 2) (http://www.worldclim.org/bioclim-aml). Bioclimatic variables are derived from climatic datasets (e.g. for minimum and maximum temperatures, and total precipitation) to represent biologically significant information that may influence the distribution of a species (Hijmans *et al.*, 2005; O'Donnell & Ignizio, 2012). Many of the bioclimatic variables are spatially correlated. We calculated Spearman's correlation coefficients to detect correlated bioclimatic variables (Spearman's $r > 0.7$) and to avoid putting these correlated variables into the same model

(Johnson *et al.*, 1995). Bioclimatic variables 1–19, elevation and 12 NDVI values for the months January to December were evaluated individually based on corrected Akaike's information criterion (AICc) values for likely contributions to the models (Table 3). The final model's environmental variables were chosen based on Spearman's correlation coefficients, AICc scores, and biological significance for this tick species (Table 4). A sample of 20% of records of tick occurrences were withheld from each model to use as independent test data. All other Maxent settings relating to model construction were left at their default values (parameters: convergence threshold = 10^5; maximum iterations = 500; regularization value $\beta = 1$).

Model selection

Models of the potential distribution of the ADT were created using the Maxent method (http:/www.cs.princeton.

Table 3. Environmental variables with associated individual corrected Akaike's information criterion (AICc) values.

Environmental variable	Sample size	AIC	AICc	AICc difference
NDVI, May	317	9910.9155	9912.3063	—
NDVI, April	317	9927.6467	9928.848	16.5417
NDVI, June	317	9942.7562	9943.9575	31.6512
NDVI, July	317	9961.2083	9962.4096	50.1033
NDVI, March	317	9974.6203	9976.215	63.9087
NDVI, August	317	9986.7536	9987.9549	75.6486
NDVI, September	317	9990.0403	9991.2416	78.9353
NDVI, November	317	9997.2968	9998.3231	86.0168
NDVI, October	317	10 001.9748	10 004.5337	92.2274
NDVI, February	317	10 006.6377	10 008.2323	95.926
NDVI, December	317	10 018.1077	10 020.6666	108.3603
NDVI, January	317	10 022.5022	10 024.7975	112.4912
Elevation	317	9743.0654	9745.14	—
Bioclimatic 2: mean diurnal range	317	9951.2291	9952.8238	—
Bioclimatic 3: isothermality	317	9965.1249	9966.3262	13.5024
Bioclimatic 16: precipitation of wettest quarter	317	9967.5174	9969.1121	16.2883
Bioclimatic 6: min temperature of coldest month	317	9967.6394	9969.6863	16.8625
Bioclimatic 9: mean temperature of driest quarter	317	9977.6444	9979.6912	26.8674
Bioclimatic 1: annual mean temperature	317	9979.4066	9980.7973	27.9735
Bioclimatic 11: mean temperature of coldest quarter	317	9977.5858	9981.028	28.2042
Bioclimatic 7: annual temperature range	317	9979.2283	9981.0417	28.2179
Bioclimatic 13: precipitation of wettest month	317	9983.9501	9984.9764	32.1526
Bioclimatic 18: Precipitation of warmest quarter	317	9985.0184	9986.2197	33.3959
Bioclimatic 12: annual precipitation	317	9985.6619	9986.5274	33.7036
Bioclimatic 8: mean temperature of wettest quarter	317	9990.6557	9992.469	39.6452
Bioclimatic 4: temperature seasonality	317	9989.7426	9993.8522	41.0284
Bioclimatic 5: max temperature of warmest month	317	9995.3014	9995.8878	43.064
Bioclimatic 10: mean temperature of warmest quarter	317	10 000.6749	10 002.9702	50.1464
Bioclimatic 15: precipitation seasonality	317	10 015.7877	10 016.814	63.9902
Bioclimatic 19: precipitation of coldest quarter	317	10 023.4065	10 024.2721	71.4483
Bioclimatic 17: precipitation of driest quarter	317	10 030.1829	10 030.7692	77.9454
Bioclimatic 14: precipitation of driest month	317	10 040.1391	10 041.5298	88.706

NDVI, normalized difference vegetation index.

Table 4. Biological significance of bioclimatic variables in the species distribution model for the American dog tick (ADT), *Dermacentor variabilis*.

Bioclimatic variable∗	Biological significance
Elevation	The ADT is commonly found in habitats (soils, elevation, climate) associated with lower elevations (Bishopp & Trembley, 1945)
	Elevation may take into account not only terrain-specific associations such as soil composition, but also indications of interactions of habitat, ticks and hosts
NDVI July	A surrogate for moisture availability
	Moisture needed for survival during active periods (e.g. host-seeking behaviour) (Randolph, 2000)
Mean diurnal temperature range	Temperature stability important for when immature stages are seasonally active
Isothermality	Related to the quantificationof how large the day–night temperature oscillation is in comparison with the summer–winter oscillation
	Supports immature development
Minimum temperature of coldest month	Affects mortality of eggs
Mean temperature of coldest quarter	Affects mortality of eggs
Precipitation of wettest quarter	Egg development and larval survivability
Precipitation of driest quarter	Lack of winter precipitation will inhibit and delay development, causing high mortality to tick populations

∗The selection of the variables was based on Spearman's statistical test, biological significance, and individual AICc scores.
AICc, corrected AIC; NDVI, normalized difference vegetation index.

edu/~schapire/maxent/). Maximum entropy has been applied widely to model distributions of species, particularly using museum or herbarium datasets, and has performed well when compared with other presence-only methods (Elith *et al.*, 2006). Maxent works by finding the largest spread (maximum entropy) in a geographic dataset of species presences in relation to a set of background environmental variables (Phillips *et al.*, 2009). Maxent Version 3.3.3 (with the default settings) was used to create the ADT model. The default background points were used with each model combination in addition to the regularization parameter.

Model evaluation

Model performance was evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve (Elith *et al.*, 2006) and AICc metric. The AUC is a non-parametric measure used to compare the relative ranking ability of the models produced with different settings for this tick species within the same study region. The AUC from an ROC analysis varies from 0 to 1 and represents a measure of a model's overall ability to discriminate suitable and unsuitable areas, independently of a threshold. We used both the test AUC and the difference between the test and training AUCs to determine the best model.

The ENMTools package was used to create AIC values for all models to select the best performing model. ENMTools is a Perl script with a graphical user interface and is available as an executable file for Windows and Macintosh (Warren *et al.*, 2010). The AICc is a metric that assesses the relative goodness of fit of each model (Burnham & Anderson, 2002). Warren and Seifert (2011) found the AIC metric exhibited the best average performance for selecting models that estimate true model complexity (when using small sample sizes) and that models preferred by AIC more accurately estimate the relative importance of variables.

The performance of our best approximating model was validated using the independent dataset from the NVSL in Ames, Iowa. The ADT identification records obtained from the NVSL dataset represented tick collections from 864 U.S. counties within 43 states from 1989 to 2006 (Fig. 1B).

Results

Environmental variable selection

The NDVI value for the month of May showed the lowest difference among AICc values, followed by NDVI values for April, June and July. Bioclimatic variable 2 had the lowest difference among AICc values, followed by bioclimatic variables 3, 16, 6 and 9 (Table 3). The distribution of the ADT, based on county-level NVSL tick surveillance data, indicates that the tick is found at lower elevations with a high humidity index and warm temperatures (Fig. 2). Therefore, we used temperature and precipitation variables that represented moisture in a microenvironment, temperature variability and extreme temperatures in our candidate models. Lastly, the final eight environmental

datasets selected (bioclimatic variables 2, 3, 5, 6, 13, 15, the NDVI value for July, and elevation), based on biological significance and AICc differences, were not correlated with one another according to Spearman's correlation tests. Eight different candidate models were selected using these variables to determine appropriate combinations for a model of habitat suitability for the ADT.

Model evaluations

The Maxent modelling approach showed good predictive value in describing areas suitable for the ADT species and areas of habitat of low or absent suitability across the U.S.A. The largest per-state percentage areas of highly suitable habitat for the ADT lay in the eastern region of the U.S.A. in Delaware, Maryland, New Jersey, Florida, Massachusetts and Rhode Island. Other states with high percentages of suitable habitat were South Dakota, Iowa and Wisconsin. Suitable environmental conditions were either absent or present only in localized areas in the Rocky Mountain regions of New Mexico, Arizona, Utah, Nevada and Colorado.

The ADT distribution model is considered to usefully represent the overall distribution of this tick species given that the final model test AUC (0.82) lay between 0.7 and 0.9 and was > 0.75 (Swets, 1988; Elith *et al.*, 2006, 2010). The remaining seven combinations had similar test AUC scores in the range of 0.77–0.81 with standard deviations of 0.025–0.028. The minimum difference between training and test AUCs ranged from 3 to 5 for all models created using the default value in Maxent for the background. Regularization parameters (1–10) were evaluated with all model combinations; increasing the regularization parameter did not improve modelling performance. Moreover, the lowest AICc difference using the eight-variable model indicated a good fit (Burnham & Anderson, 2002; Warren *et al.*, 2010). The AICc differences ranged from 0 to 232 for the eight different model combinations using the default settings in Maxent. The eight-variable model had the highest test AUC and the lowest AICc score (Table 5).

We also carefully considered the background region and sampling size for this tick species across the continental U.S.A. and the influence on our model. The background points used in our model were a random subsample of our entire study area – continental U.S.A. – and represented both the regions in which tick specimens had been collected or observed over a 50-year period and those in which we wanted to project our model. The randomly selected background points created by the maximum entropy methodology at a national scale provided a relatively useful model with the higher test AUCs and minimal difference between training and test AUCs.

The Maxent final eight-variable tick distribution model used in the present study provides habitat suitability probability values ranging from 0 to 1 across all grid cells (Fig. 2A). In this study, we used the lowest presence threshold (LPT), which sets the threshold at the lowest value of the prediction for any of the presence localities to determine cut-offs for suitable vs. unsuitable localities for the purposes of graphical presentation (Peterson *et al.*, 2008). The categories for suitable habitat were
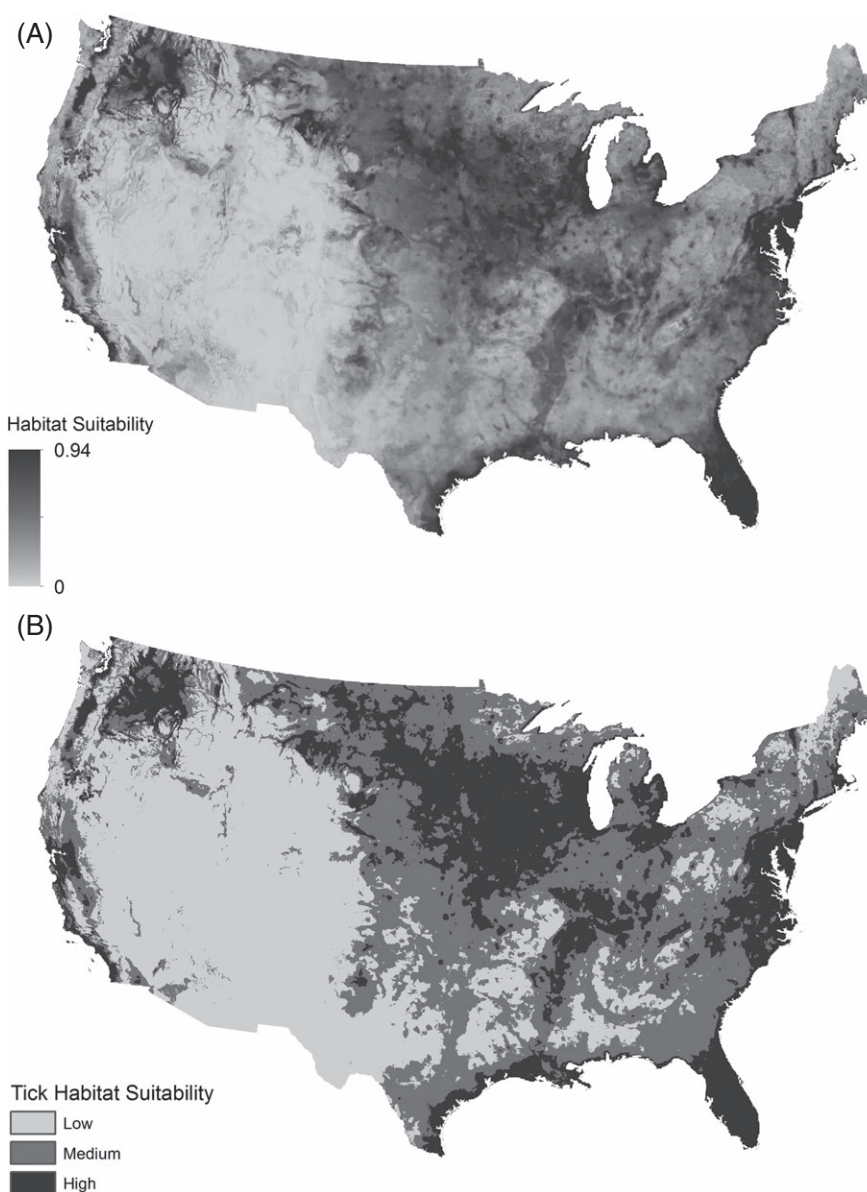
Fig. 2. (A) Potential distribution of the American dog tick, *Dermacentor variabilis*, in the U.S.A. based on habitat suitability produced by Maxent modelling. (B) Potential distribution of *D. variabilis* based on three classes of habitat suitability.

further characterized using Maxent's ROC curve and associated histogram to divide suitability into medium and high classes (Hosmer & Lemeshow, 2000). The objective of using this curve was to determine the optimal cut point that maximized both specificity and sensitivity. Specificity and sensitivity curves crossed at 0.1 and thus this was chosen as our optimal cut point for separating unsuitable (low) from suitable (medium and high) habitat (Fig. 2B).

An independent tick identification dataset from the NVSL in Ames, Iowa was used to validate the ADT habitat suitability eight-variable model. All 847 counties reporting this tick species intersected with medium to high habitat suitability as associated with our model. A total of 283 counties intersected with the

medium habitat suitability, and 565 counties intersected with high habitat suitability (Fig. 3).

Environmental variable contributions

The percentage contribution is an estimate of the relative contribution of an environmental variable to the Maxent model or the order that Maxent used to constructed the model. However, the permutation importance depends only on the final Maxent model, not the path used to obtain it. The contribution of each variable is determined by randomly permuting the values of that variable among the training points (both presence and

Table 5. Evaluation of Maxent climatic models for the distribution of the American dog tick, *Dermacentor variabilis*.

	AIC scores				AUC scores				
Model combinations	Sample no.	AIC	AICc	AICc diff.	Variable no.	AUC_Tr	SD	AUC_Tt	SD	AUC diff.
Model 1: BioClim 2, 3, 5, 6, 13, 15 (temp/precip) + elevation + NDVI July	317	9479.53	9540.42	0	8	0.8644	0.01	0.8262	0.0249	0.0382
Model 2: BioClim 2, 3, 5, 6 (temp) + elevation + NDVI July	317	9525.71	9575.3	34.88	6	0.8549	0.01	0.8091	0.0246	0.0458
Model 3: BioClim 13, 15 (precip) + elevation + NDVI July	317	9570.24	9604.98	64.56	4	0.836	0.011	0.792	0.0259	0.044
Model 4: BioClim 2, 3, 5, 6, 13, 15 (temp/precip) + elevation	317	9550.7	9606.57	66.14	7	0.8508	0.01	0.806	0.025	0.0448
Model 5: BioClim 2, 3, 5, 6 (temp) + elevation	317	9604.3	9632.28	91.86	5	0.8242	0.011	0.7797	0.0255	0.0445
Model 6: BioClim 13, 15 (precip) + elevation	317	9616.97	9636.43	96.01	3	0.8105	0.011	0.7718	0.0273	0.0387
Model 7: BioClim 2, 3, 5, 6, 13, 15 (temp/precip) + NDVI July	317	9697.47	9754	213.58	7	0.8494	0.012	0.795	0.0264	0.0544
Model 8: BioClim 2, 3, 5, 6, 13, 15 (temp/precip)	317	9746.01	9772.54	232.12	6	0.8201	0.012	0.772	0.0276	0.0481

AIC, Akaike's information criterion; AICc, corrected AIC; AUC, area under the curve; NDVI, normalized difference vegetation index; SD, standard deviation.

background) and measuring the resulting decrease in training AUC. Based on percentage contributions, four environmental variables contributed almost 78% of influence in the final model. Variable importance in the Maxent model indicated that elevation was the best predictor of potential suitable habitat, contributing 36.6% to model development. The NDVI value for the month of July was the second best predictor, contributing 17.2%, followed by diurnal temperature ranges, or temperature fluctuations, at 12.2%, and maximum temperature of the warmest month at 11.2%. All eight model combinations showed a similar trend, with elevation representing the largest contributing variable, followed by NDVI value in July, diurnal temperature ranges (bioclimatic variable 2), maximum temperature of the warmest month (bioclimatic variable 5), and minimum temperature of the coldest month (bioclimatic variable 6). However, the order of permutation importance for each variable changed slightly, with 77% of permutation importance determined by elevation, NDVI value for July, temperature of the warmest month (bioclimatic variable 5), and precipitation seasonality (bioclimatic variable 15).

Elevation was the most significant environmental variable in all models. This derives from the fact that many of the localities from well-sampled portions of the species' range lie at lower elevations than other areas. For the variable precipitation, evaluated localities corresponded to very few precipitation variables other than precipitation seasonality. By contrast, the same three or four environmental variables – elevation, NDVI and maximum temperature of the warmest month, or mean diurnal temperature range – were significant in model performance in the well-sampled portions of the species' distribution

for all sample sizes, complexity or background selection areas.

Discussion

Tick species distributions can be predicted using a variety of statistical and spatial methods that look at associations between environmental factors and the presence of a tick species of economic importance. However, in most cases, we do not have good tick collection information from well-designed tick surveys that can provide starting points to generate predictive maps on an ongoing basis (Estrada-Peña, 2006). Very few current descriptive maps of distributions of tick species in the U.S.A. are available because the maintenance of such species distribution maps requires intensive sampling and resources. Therefore, predictive tick species distribution maps provide us with the best methods to understand how environmental patterns may relate to the distribution of an individual tick species so that we can fill in the gaps in areas for which we have no presence data for an economically important tick species. Moreover, predictive tick distribution maps provide opportunities to understand possible changes in tick species distributions as a result of environmental changes caused by human interventions or mitigations, land use and climate change (Randolph, 1994).

The environmental variables that proved to have the greatest influence on our final model were elevation, vegetation indices (NDVI) for the month of July (humidity on the ground), and bioclimatic variables 2 (mean diurnal temperature range) and
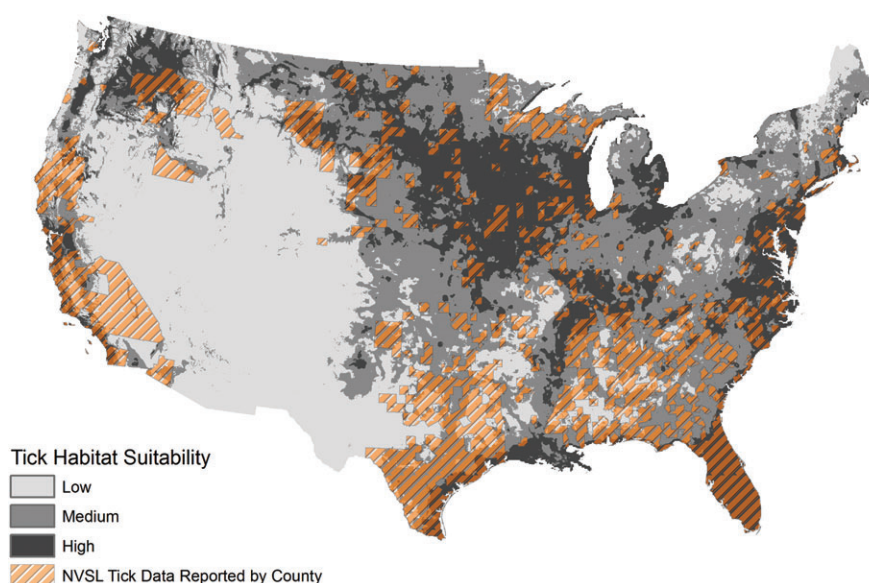
Fig. 3. Validation of the potential distribution of the American dog tick, *Dermacentor variabilis*, based on habitat suitability using National Veterinary Service Laboratories (NVSL) county-based tick data for 1998–2011.
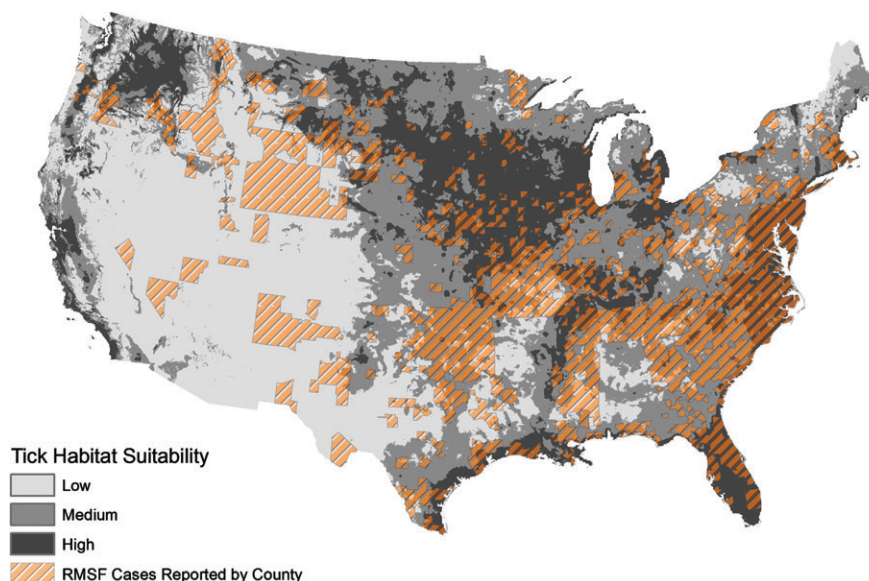

Fig. 4. Potential distribution of the American dog tick, *Dermacentor variabilis*, based on habitat suitability overlaid with Centers for Disease Control and Prevention (CDC) data on cases of Rocky Mountain spotted fever (RMSF) by county for 1998–2011.

5 (maximum temperature of the warmest month). All three of these environmental variables are consistent with some of the factors found by McEnroe (1979b), and Sonenshine (1991) to influence the distribution, seasonal activity and abundance of the ADT. The ADT is commonly found in habitats (soils, elevation, climate) associated with lower elevations (Bishopp & Trembley, 1945). Diurnal temperature ranges, or temperature fluctuations, will affect tick development and activity on a regional basis, with day length accentuating the effects of temperature. Temperature stability and moisture are important to tick immature stage development and seasonal activity. Temperature extremes will affect both egg and immature mortality, and moisture will influence both egg development and larval survival. In the southeast of the U.S.A., this tick species can go through multiple generations in a single year, whereas in colder regions of the country it may present only one generation in a single year. The ADT will enter into a diapause or suspended physiological activity in response to adverse conditions such as colder temperatures (Sonenshine, 1991). Winter temperatures across the U.S.A. are likely to represent one of the factors that determine the geographic range of the tick, with the duration of the mean temperature and water availability in an area directly influential on

immature development, survival and seasonal activity (Harlan & Foster, 1990). Randolph (2000) also confirmed that temperature and humidity were major factors in describing the presence of *Ixodes ricinus* L. (Ixodida: Ixodidae) and tick-borne disease encephalitis in Europe.

In the model, NDVI values for July had relatively high influence on predicting the likelihood of suitable habitat for this tick species. The NDVI is a spectral vegetative index that quantifies the level of photosynthetic activity of the vegetation in an area and is an indicator of vegetation biomass. According to Randolph (1994), the NDVI is a significant variable for predicting tick distributions because it has a biological relationship to tick biology, which refers to the availability of moisture for free-living ticks. The moisture index during the month of July is likely to reflect the overlap in seasonal activity of all three stages (larvae, nymphs and adults) of this tick species. All stages of the ADT are actively seeking hosts (free-living stages) during this time and moisture availability is critical to their survival and presence in certain areas of the U.S.A. The ADT is one of the most widely distributed tick species in the U.S.A. and one of the limitations of its broad range is likely to relate to variations in humidity across the continent.

The monitoring of changes in the distribution of the ADT will be important because the extreme high and low temperatures that are anticipated to occur with climate change are likely to influence the host-finding rates and development of this tick species (Ogden *et al.*, 2008). For example, an early simulation model for the ADT showed that ambient temperature and moisture changes affect the lifecycle of this tick species (Mount & Haile, 1989). Moreover, climate change will affect not only the seasonal activity of this tick species, but also its transmission of pathogens such as RMSF, a rickettsial disease in humans, 90% of cases of which occur in the eastern half of the U.S.A., where the ADT is the major vector. We obtained a dataset for RMSF from the National Notifiable Disease Surveillance System (NNDSS) database maintained by the Centers for Disease Control and Prevention and found that the majority of counties reporting RMSF cases from 2000 to 2007 intersected with our model's designated areas of suitable habitat for the ADT, which supports previous work showing a close association between the presence of the ADT and RMSF cases (Schriefer & Azad, 1994). The Rocky Mountain wood tick, *Dermacentor andersoni*, the other major tick vector of RMSF, resides in regions of low suitability for the ADT and thus accounts for some cases of RMSF that overlap with these areas (Fig. 4). Changes in the distribution of cases of RMSF that have occurred over the last decade have been shown to be highly correlated with man-made ecological changes, climate fluctuations and increased levels of interaction among ticks, humans and mammalian hosts (Schriefer & Azad, 1994). The same ecological changes also affect both domestic and wild animal populations, resulting in an increased risk for disease transmission with changes in the distributions of tick vectors (Gray *et al.*, 2009). We may observe similar changes in the distribution of tick-transmitted equine piroplasmosis with changes in the distribution of the ADT across the U.S.A. that occur in response to changes in host density, landscape and climate. Our current ADT habitat suitability model may be a useful baseline tool for monitoring the changes in the distribution of this tick species and the risk for contact and disease to human and equine populations.

The Maxent modelling approach has proven to be a helpful tool in determining an approximation of the distribution of the ADT at a national scale. It appears to be a good method for use with presence-only museum data because it is insensitive to spatial errors associated with historically collected locational data (Elith *et al.*, 2010). Further advances are required to develop modelling methods to better qualify the distributions of tick species and the related climate variables, and to establish statistical methodologies for evaluating presence-only tick location data. In addition, it would be beneficial to develop protocols for assessing habitat requirements at a local scale for a variety of public and agriculturally important tick species in order to refine tick mitigations at a state level based on the biological requirements of a tick species rather than on a generalized taxonomic approach to ticks. The creation of models applied on partial ranges (i.e. regional) might represent a better approach to avoiding the generalization of tick distribution at a local level from a national-scale model. A major challenge in creating predictive tick habitat models refers to the collection of systematic tick presence/absence data, tick population density and host density indices that can be used to develop suitable statistical models for monitoring changes in tick-borne disease risk that occur with climate and land use changes in the U.S.A. An accuracy assessment (i.e. 'ground-truthing') of the ADT distribution model using systematic field collections of tick specimens from areas of high probability would add valuable input into future developments of tick species distribution modelling and provide further validation of our distribution model.

References

Bishopp, F.C. & Trembley, H.L. (1945) Distribution and hosts of certain North American ticks. *Journal of Parasitology*, **31**, 1–54.

Brownstein, J.S., Holford, T.R. & Fish, D. (2005) Effect of climate change on Lyme disease risk in North America. *EcoHealth*, **2**, 38–46.

Burnham, K. & Anderson, D. (2002) *Model Selection and Multi-Model Inference: A Practical Information–Theoretic Approach*, 2nd edn. Springer, New York, NY.

Cummings, G.S. (2000) Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography*, **27**, 441–455.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

Estrada-Peña, A. (2006) Prediction of habitat suitability for ticks. *Annals of the New York Academy of Sciences*, **1078**, 275–284.

Estrada-Peña, A., Corson, M., Venzal, J.M., Mangold, A.J. & Guglielmone, A. (2006) Changes in climate and habitat suitability for the cattle tick *Boophilus microplus* in its southern Neotropical distribution range. *Journal of Vector Ecology*, **31**, 158–167.

Estrada-Peña, A., Ayllon, N. & de la Fuente, J. (2012) Impact of climate trends on tick-borne pathogen transmission. *Frontiers in Physiology*, **3**, 64.

Franklin, J. (2010) *Mapping Species Distributions: Spatial Inference and Prediction: Ecology, Biodiversity and Conservation*. Cambridge University Press, Cambridge.

Gage, K., Burkot, T., Eisen, R. & Hayes, E. (2008) Climate and vectorborne diseases. *American Journal of Preventive Medicine*, **35**, 436–450.

Gray, J.S., Dautel, H., Estrada-Peña, A., Kahl, O. & Lindgren, E. (2009) Effects of climate change on ticks and tick-borne diseases in Europe. *Interdisciplinary Perspectives on Infectious Disease*, **2009**, 1–12.

Harlan, H.J. & Foster, W.A. (1990) Micrometeorologic factors affecting field host-seeking activity of adult *Dermacentor variabilis* (Acari: Ixodidae). *Journal of Medical Entomology*, **27**, 471–479.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hosmer, D.W. & Lemeshow, S. (2000) *Applied Logistic Regression*, 2nd edn. John Wiley & Sons, New York, NY.

Johnson, N.L., Kotz, S. & Balakrishnan, N. (1995) *Distributions in Statistics: Continuous Univariate Distributions*, 2nd edn, Vol. 2. John Wiley & Sons, New York, NY.

McEnroe, W.D. (1979a) *Dermacentor variabilis* (Say) in eastern Massachusetts. *Recent Advances in Acarology*, Vol. 2 (ed. by J.G. Rodriguez), pp. 145–153. Elsevier, Amsterdam.

McEnroe, W.D. (1979b) The effect of the temperature regime on *Dermacentor variabilis* (Say) populations in eastern North America. *Acarologia*, **20**, 58–67.

Mount, G.A. & Haile, D.G. (1989) Computer simulation of the population dynamics of the American dog tick, *Dermacentor variabilis* (Acari: Ixodidae). *Journal of Medical Entomology*, **26**, 60–76.

O'Donnell, M.S. & Ignizio, D.A. (2012) *Bioclimatic Predictors for Supporting Ecological Applications in the Conterminous United States* U.S. Geological Survey Data Series 691. U.S. Department of the Interior, Washington, DC.

Ogden, N.C., Onge, L., Barker, I.K. *et al.* (2008) Risk maps for range expansion of the Lyme disease vector, *Ixodes scapularis*, in Canada now and with climate change. *International Journal of Health Geographics*, **7**, 24.

Paddock, C.D. & Telford, S.R. (2011) Through a glass, darkly: the global incidence of tick-borne diseases. Institute of Medicine Committee on Lyme disease and other tick-borne diseases. *State of the Science Workshop: Critical Needs and Gaps in Understanding Prevention, Amelioration, and Resolution of Lyme and Other Tick-Borne Diseases: The Short-Term and Long-Term Outcomes: Workshop Report*. National Academies Press, Washington, DC. http://www.ncbi.nlm.nih.gov/books/NBK57015 [accessed on 3 July 2014].

Peterson, A.T., Papes, M. & Soberón, J. (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modelling. *Ecological Modelling*, **213**, 63–72.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

Porretta, D., Mastrantonio, V., Amendolia, S. *et al.* (2013) Effects of global changes on the climatic niche of the tick *Ixodes ricinus* inferred by species distribution modelling. *Parasites and Vectors*, **6**, 271.

Randolph, S.E. (1994) Population dynamics and density dependent seasonal mortality indices of the tick *Rhipicephalus appendiculatus* in eastern and southern Africa. *Medical and Veterinary Entomology*, **8**, 351–368.

Randolph, S.E. (2000) Ticks and tick-borne disease systems in space and from space. *Advances in Parasitology*, **47**, 217–243.

Schriefer, M.E. & Azad, A.F. (1994) Changing ecology of Rocky Mountain spotted fever. *Ecological Dynamics of Tick-borne Zoonoses* (ed. by D.E. Sonenshine & T.N. Mather), pp. 314–324. Oxford University Press, New York, NY.

Scoles, G.A, Hutcheson, H.J., Schlater, J.L., Hennager, S.G., Pelzel, A.M. & Knowles, D.P. (2011) Equine piroplasmosis associated with *Amblyomma cajennense* ticks, Texas, U.S.A. *Emerging Infectious Diseases*, **17**, 1903–1905.

Sonenshine, D.E. (1991) *Biology of Ticks* Vol. **1**. Oxford University Press, New York, NY.

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.

Torina, A., Alongi, A., Naranjo, V. *et al.* (2008) Prevalence and genotypes of *Anaplasma* species and habitat suitability for ticks in a Mediterranean ecosystem. *Applied and Environmental Microbiology*, **74**, 7578–7584.

Warren, D.L. & Seifert, S.N. (2011) Environmental niche modelling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, **21**, 335–342.

Warren, D.L., Glor, R.E. & Michael, T. (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography*, **33**, 607–611.