HHS Public Access

Author manuscript

J Med Entomol. Author manuscript; available in PMC 2024 May 12.

Published in final edited form as:

J Med Entomol. 2023 May 12; 60(3): 590–603. doi:10.1093/jme/tjad042.

Identifying suitable habitat for *Ixodes scapularis* (Acari: Ixodidae) infected with *Anaplasma phagocytophilum* (Rickettsiales: Anaplasmataceae), *Babesia microti* (Piroplasmida: Babesiidae), and *Borrelia miyamotoi* (Spirochaetales: Spirochaetaceae) to guide surveillance efforts in the eastern United States

James C. Burtis[*],

Erik Foster,

Christina M. Parise,

Rebecca J. Eisen

Division of Vector-Borne Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Pr evention, Fort Collins, CO 80521, USA

Abstract

Understanding the distribution of infected ticks is informative for the estimation of risk for tickborne diseases. The blacklegged tick, *Ixodes scapularis* (Acari: Ixodidae), is the primary vector for 7 medically significant pathogens in United States. However, knowledge of the ranges of these pathogens in host-seeking ticks is incomplete, particularly for those occurring at low prevalence. To aid in prioritizing costly field sampling efforts, we estimated ranges of suitable habitat for *Anaplasma phagocytophilum*, *Babesia microti*, and *Borrelia miyamotoi* in the eastern United States based on existing county-level surveillance records. The resulting suitability maps were compared against those developed previously for *Bo. burgdorferi* s.s., which shares similar ecology but has been detected in a greater number of counties. The overall accuracy of the habitat suitability models was high (AUC ≥ 0.92) for all 4 pathogens. The most important predictors were related to temperature and moisture. The upper midwestern and northeastern states were predicted to be highly suitable for all 4 pathogens. Based on our models, we prioritized sampling in 431, 275, and 539 counties currently lacking pathogen records that our models classified as suitable for *A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi*, respectively. As a second-tier priority, we identified 311 (*A. phagocytophilum*), 590 (*Ba. microti*), and 252 (*Bo. miyamotoi*) counties, based on high suitability scores for *Bo. burgdorferi*. Our models can be used to improve cost-effectiveness of field sampling efforts aimed at improving accuracy and completeness of pathogen distribution maps.

[*]Corresponding author: ptd6@cdc.gov.

Supplementary Material

Supplementary material is available at *Journal of Medical Entomology* online.

Keywords

blacklegged tick; Lyme disease; anaplasmosis; babesiosis; relapsing fever

Introduction

The blacklegged tick, *Ixodes scapularis* (Acari: Ixodidae) (Say), is the primary vector of multiple tick-borne pathogens that account for the majority of vector-borne disease cases reported in the United States (Eisen and Eisen 2018, Rosenberg et al. 2018). *Ixodes scapularis* is a vector for *Borrelia burgdorferi* (Spirochaetales: Spirochaetaceae) sensu stricto, the primary causative agent of Lyme disease, the most commonly reported vector-borne disease in the United States (Kugeler et al. 2021). Additionally, *I. scapularis* vectors several other medically significant pathogens including a more recently discovered and less common agent of Lyme disease (*Bo. mayonii*) (Pritt et al. 2016), *Anaplasma phagocytophilum* (anaplasmosis), *Babesia microti* (babesiosis), *Bo. miyamotoi* (hard tick-borne relapsing fever), *Ehrlichia muris eauclairensis* (ehrlichiosis), and Powassan virus (Powassan virus disease) (Eisen and Eisen 2018). Although less commonly detected in *I. scapularis* than *Bo. burgdorferi* s.s., each of these pathogens has been detected in *I. scapularis* collected from upper midwestern states and most (excepting *Bo. mayonii* and *E. muris eauclairensis*) have been detected in northeastern states (Fleshman et al. 2021, 2022). While reported case numbers are currently low relative to Lyme disease, human cases of anaplasmosis, babesiosis, and Powassan virus disease are increasing and spreading geographically in the northeastern and upper midwestern United States (Dahlgren et al. 2015, Gray and Herwaldt 2019, Campbell and Krause 2020).

In 2018, the US Centers for Disease Control and Prevention (CDC) initiated a national tick and tickborne pathogen surveillance effort intended to provide data driven assessments to the public, clinicians, and policy makers regarding when and where persons are at risk for exposure to medically important ticks and their associated human pathogens (CDC 2018, Eisen and Paddock 2021). This effort includes collaborations between CDC and public health partners to collect, test, and document tick and pathogen presence as well as abundance at county spatial scales. Though records are incomplete, the effort has revealed a substantial increase in the number of US counties where *I. scapularis* and its associated pathogens have been detected (Eisen et al. 2016, Fleshman et al. 2021, 2022).

Many efforts to model risk for *I. scapularis*-borne diseases in the eastern United States have focused on determining the habitat suitability for blacklegged ticks (Estrada-Peña 2002, Brownstein et al. 2003, Hahn et al. 2016), but fewer efforts have focused on modeling acarological risk for Lyme disease based upon the likelihood of human encounters with infected host-seeking ticks (Diuk-Wasser et al. 2012). Counties classified as suitable for sustaining *Bo. burgdorferi* s.s.-infected *I. scapularis* provided greater concordance with counties reporting high Lyme disease incidence in the eastern United States compared with models of the vector's distribution alone (Burtis et al. 2022). However, counties with potentially suitable habitats for other *I. scapularis*-associated pathogens have not been modeled across range of *I. scapularis* in the eastern United States.

A challenge to building habitat suitability models is access to broad geographic records of pathogen presence in host-seeking ticks. Tick-based surveillance efforts have yielded county records for the 7 pathogens vectored by *I. scapularis* in the United States (Fleshman et al. 2021, 2022). However, the resulting maps are likely underestimates of the true distributions of these pathogens, in part because surveillance efforts are limited and detection of rare pathogens requires substantial testing effort (Lehane et al. 2021, Fleshman et al. 2022). We used these records of infected *I. scapularis* to predict the range of suitable habitat for the 3 pathogens for which we had sufficient data to construct models: *A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi*. These infection records represent presence-only data (with pseudo-absence), therefore the resulting suitability maps are highly dependent upon the presence records upon which they are built, and sampling effort is not evenly distributed across the eastern United States; therefore, we assume our models will predict a restricted range of suitable habitat. Recognizing similarities in transmission cycles between these pathogens and *Bo. burgdorferi* s.s. (LoGiudice et al. 2003, Barbour et al. 2009, Hersh et al. 2012, Keesing et al. 2012), we compared the predicted distributions of suitable habitat for these 3 pathogens against a previously published suitability model for *Bo. burgdorferi* s.s. which developed habitat suitability models using an ensemble modeling approach (Burtis et al. 2022). The intent of comparing habitat suitability against *Bo. burgdorferi* s.s. was to determine whether predicted suitable habitat aligned across pathogens, and whether these distributions were defined by similar associations with the environmental predictors used in the models.

Methods

Field Data for *Ixodes scapularis*-Borne Pathogens

We generated county-level datasets for the eastern United States independently for 3 pathogens: *A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi*. County observations of infected *I. scapularis* for all 3 pathogens were derived from historical records and data submitted to the ArboNET Tick Module through 2021 (Fleshman et al. 2022). The ArboNET Tick Module is a database maintained by the CDC through which public health agencies and academic partners submit data documenting the presence and abundance of ticks and the presence and prevalence of their associated pathogens. Testing methods for the pathogen data included in this database conformed with CDC tick surveillance guidelines (CDC 2018). For each of the 3 pathogens, counties wherein at least one *I. scapularis* of any life stage tested positive for infection were coded as “present”, while those counties without records were coded as “no records”.

Climate and Landscape Predictors

The ecology of the 3 pathogens (*A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi*) and *Bo. burgdorferi* s.s. is similar (Slajchert et al. 1997, LoGiudice et al. 2003, Barbour et al. 2009, Hersh et al. 2012, Keesing et al. 2012, Han 2019, Larson et al. 2021). In order to compare the functional relationship between environmental predictors and the predicted suitability for each pathogen, we used the same predictive variables that had been used previously to construct *Bo. burgdorferi* s.s. habitat suitability models (Burtis et al. 2022). A full list of potential predictors can be found in Supplementary Materials

(Supplementary Table S1). Our final models included 8 climate predictors (Table 1) from WorldClim (Hijmans et al. 2005) that were updated with Daymet records (Thornton et al. 1997, 2016) collected between 1980 and 2015 (Johnson et al. 2017). Percent forest cover was also included as an influential predictor of *I. scapularis* presence in previous modeling efforts (Estrada-Peña 2002, Diuk-Wasser et al. 2010, Hahn et al. 2016, Burtis et al. 2022) and was derived from the 2006 national land cover database (Fry et al. 2011). Predictors were generated at a spatial resolution of 1 km × 1 km and the Zonal Statistics tool in QGIS (v. 3.14.1) (QGIS Development Team 2009) was used to generate county averages. We limited candidate predictor variables to those used in the *Bo. burgdorferi* s.s. mode to simplify comparability among the resulting pathogen distribution models. However, to determine if a better fit could be achieved using an expanded variable selection method, we constructed models using an alternative variable selection strategy described in the Supplementary Materials.

Pathogen Risk Modeling

To increase confidence in model predictions, rather than relying on a single statistical modeling framework, we employed a modeling ensemble approach. We generated 5 models based upon each of the 3 datasets (*A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi*) covering the eastern United States, which represents the expected range of *I. scapularis* (Diuk-Wasser et al. 2010, Hahn et al. 2016, Burtis et al. 2022). There were 283 county-level observations for *A. phagocytophilum*, 154 for *Ba. microti*, and 232 for *Bo. miyamotoi* (Figs. 1–3). We used 5 modeling algorithms, (i) boosted regression tree (BRT), (ii) generalized linear model (GLM), (iii) maximum entropy (Maxent), (iv) multivariate adaptive regression splines (MARS), and (v) random forest (RF) (Talbert and Talbert, 2001). The *Bo. burgdorferi* s.s. ensemble model combined output from 3 of these modeling approaches (GLM, MARS, Maxent). Probability values (suitability scores hereafter) were generated by the models for each county, which were converted to binary form (high or low suitability) using 2 different suitability score thresholds. This process is fully described in the next section. The output of a previous habitat suitability modeling effort to identify counties that were suitable for *Bo. burgdorferi* s.s., based on 402 county-level observations was included for comparison with pathogen-specific model outcomes described herein (Burtis et al. 2022).

The BRT is a flexible approach that is useful when there are complex relationships between predictor and response variables. It is a boosting approach that iteratively fits trees. It first generates a sequence of simple trees and then creates new trees based upon their residuals (Elith et al. 2008, Merow et al. 2014). The GLM is a generalized ordinary least squares regression with a link to a binomial function. This approach is suitable when relationships between predictor and response variables are relatively simple (e.g., follow a somewhat linear, unimodal, or s-shaped pattern) (Guisan et al. 2017). Maxent is a maximum entropy approach. It relies on the generation of probability distributions to match the dataset. The distribution which maximizes entropy (or uncertainty) is selected. It is highly flexible and particularly and is well-suited for presence-only datasets (Elith et al. 2011). MARS represents a more flexible regression than GLM. This allows relationships between predictor and response variables to deviate from simple shapes (e.g., linear, unimodal, s-shaped). This method allows for the analysis of complex relationships between predictors and responses

(Guisan et al. 2017). RF is a bagging or bootstrapping aggregation approach. Trees are fit to different bootstrapped subsamples of the data and then the mean probability across all runs is determined (Breiman 2001). The ensemble approach used here can help to account for differences between model assumptions that can affect their output.

We built each model using the entire pathogen specific presence/pseudo-absence dataset containing all counties with known records of each pathogen. These models are referred to as the "training" models hereafter. To evaluate model performance, we used a 10-fold cross-validation method to generate "testing" runs for the data. This method splits the data into 10 equal subsets, then runs the model 10 separate times, removing one subset with each run. This allows for the examination of variation in model outcome when the underlying datasets differ slightly. The Receiver Operating Characteristic (ROC) curve and resulting area under the curve (AUC) were compared between the "testing" and "training" models. The ROC curve plots the true positive rates (sensitivity) and false positive rates (1-specificity) across a range of suitability scores. The AUC is derived from the ROC and is a measure of model accuracy that is independent of sensitivity or specificity thresholds. An AUC value of 1 represents a model with a 'perfect' fit and values <0.5 indicate that the model is poorly fit to the data. If the AUC values differed by >0.05 between the testing and training runs, consistent with previous studies we considered the models to be overfit (Jensen et al. 2005, Edwards et al. 2006, Phillips et al. 2006, 2009, Dormann et al. 2008, Zimmermann et al. 2009, Elith et al. 2011, Hahn et al. 2016, Guisan et al. 2017, Burtis et al. 2022).

We also checked for differences between training and testing runs in specificity, sensitivity, positive predictive value, negative predictive value, and percent correctly classified; large differences can indicate an unstable model (Elith et al. 2008). Sensitivity is measured as the percentage of counties where the pathogen is known to occur that were accurately classified as suitable accounting for false negatives (e.g., number of counties where the pathogen has been recorded and classified by the model as suitable, divided by the total number of counties where the pathogen has been recorded). Specificity represents the percentage of counties where the pathogen has not been reported that were accurately classified as unsuitable by the models accounting for false positives. Positive predictive value represents the percentage of suitable counties that are accurately classified, while negative predictive value represents the percentage of unsuitable counties that are accurately classified. Percent correctly classified is a measure of overall model accuracy which represents the percentage of suitable and unsuitable counties that are correctly classified. Model evaluation followed the same process for the *Bo. burgdorferi* s.s. model as well. All models were run using the VisTrails Software for Assisted Habitat Suitability Modeling (SAHM v. 2.2.3) using default model settings.

Model Thresholding and Construction of Consensus Map

The models generate continuous values for suitability scores. To evaluate the models and visualize suitability, we generated binary outcomes (high- or low-suitability) from the output of each of our 3 models, GLM, MARS, and Maxent. For comparisons between testing and training model runs, we dichotomized the models using a threshold suitability score that

maximized the sum of sensitivity and specificity as this is this is expected to yield the most constrained geographic distribution of highly suitable counties when using a presence-only dataset (Liu et al. 2013). When thresholding models based on balanced sensitivity and specificity, the model sensitivities were high, between 85% and 97% (Table 2). We have greater confidence in presence than the pseudo-absence data in our dataset, and to allow us to visualize these differences we dichotomized the threshold values that yielded 90% or 95% sensitivity to construct the consensus maps. That is, counties with scores above the given threshold were scored as high-suitability and all others were scored as low-suitability.

We overlaid the binary output from each of the 3 models (GLM, MARS, Maxent) to create consensus maps showing the geographic distribution of suitable habitat for each of the 3 pathogens. If a county was classified as highly suitable by 2 or more models, then that county was considered high-suitability for the consensus map. Any counties classified by one or none of the models as highly suitable were classified as low-suitability. Consensus maps were generated using models set at 90% and 95% sensitivity. Consensus maps were compared against the original county-level datasets to calculate their sensitivities, specificities, positive predictive values, negative predictive values, and percentages correctly classified. The binary output was combined to create consensus maps using QGIS (v. 3.14.1).

Comparison Against Previously Generated *Bo. burgdorferi* s.s. Consensus Maps

We compared the distributions of suitable habitat for *A. phagocytophilum*, *Bo. miyamotoi*, and *Ba. microti* with the previously published distribution of counties classified as suitable for *Bo. burgdorferi* s.s. infected *I. scapularis* using the same methods described above (Burtis et al. 2022). The overlap of the 3 pathogen-specific consensus maps was compared against the *Bo. burgdorferi* s.s. map when the sensitivity of the maps was set to 95% in order to highlight the extent of potentially suitable habitat for each pathogen. We classified counties that lack records, but were predicted to be suitable for one of the pathogens (*A. phagocytophilum*, *Bo. miyamotoi*, *Ba. microti*) as high priority for tick surveillance. Those that were not classified as suitable for one of the 3 pathogens, but were predicted to be suitable for only *Bo. burgdorferi* s.s. were classified as moderate priority. Counties that are not classified as suitable for any pathogen were classified as low priority for surveillance efforts.

Results

Variable Assessment and Model Performance

The outputs resulting from the BRT and RF modeling algorithms were found to be overfit and were omitted from inclusion in consensus maps. The BRT models for *Bo. miyamotoi* showed differences in AUC >0.05 between training and testing runs. When the sum of sensitivity and specificity was maximized, all 3 models showed large differences in sensitivity between testing and training runs, with *A. phagocytophilum* showing a 14.5% difference, *Ba. microti* a 28.5% difference, and *Bo. miyamotoi* a 16.7% difference. The RF model showed large differences between the training and testing run in specificity (>9%) and sensitivity (>38%) for all 3 pathogens. Upon visual inspection, the RF and BRT

models also rarely predicted suitable habitat where there was not already a county-level observation. Adjusting model parameters did not improve the results over default settings. Therefore, only the outputs from the GLM, MARS, and Maxent models were included when constructing the consensus maps for the 3 pathogens (Figs. 1–3).

The same 8 predictors were used for the *A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi* models (Table 1). For the GLM, MARS, and Maxent models the differences in AUC between training and testing runs were <0.02 for all 3 pathogen models. Additionally, differences in sensitivity, specificity, NPV, and PPV were ≤7% across all models for the 3 pathogens. The full list of model selection criteria and performance metrics for the testing and training runs are shown in Table 2. The very small differences between the test and training runs indicate the models were not overly sensitive to changes in the presence/pseudo-absence counties included to construct the models. That is, regardless of which randomly selected 10% of the data were excluded from the model build set, the suitability classification for each county was similar among model runs. The Maxent modeling algorithm does not have a post-hoc variable selection process, so it used all predictors, with those that did not explain a large amount of variation having low normalized contribution values (Table 1). The geographic distributions of the predictors selected are shown in Fig. 4.

Predictor Response Curves

Outcomes were similar between models constructed using the *Bo. burgdorferi* s.s. predictors (Burtis et al. 2022) and those selected uniquely for each pathogen based on percent deviance explained (Supplementary Figs. S1–S5, Supplementary Tables S2–S4). We therefore used the same 8 predictors that were used to construct the *Bo. burgdorferi* s.s. models to construct habitat suitability models for the other 3 pathogens. Although the same 8 variables (Table 1) were used to construct suitability maps for each of the 3 pathogens, the percent contributions of each variable and shapes of response curves varied among the pathogens. To a lesser extent, these factors also varied among the modeling algorithms used for each pathogen.

Anaplasma phagocytophilum response curves.—For *A. phagocytophilum* predictions, all predictors were used by the modeling algorithms, except for precipitation of the coldest quarter (BIO19) which was dropped from the MARS and GLM models, and mean temperature of the wettest quarter (BIO8) which was dropped from the MARS model (Table 1). Based on the normalized contribution values, the maximum temperature of the warmest month (BIO5), mean diurnal temperature range (BIO2), mean temperature of the driest quarter (BIO9), and percent forest cover each contributed more than 15% for at least 2 of the 3 algorithms (Table 1).

The relationship between the suitability score and BIO2 was sigmoidal for all 3 modeling algorithms (GLM, MARS, Maxent) with lower diurnal temperature ranges being more suitable than higher diurnal ranges. The slope of the relationships differs between models, with Maxent being the shallowest, followed by MARS, and GLM being the steepest. All 3 models classify mean diurnal range above 12 °C as low-suitability. The relationship for BIO5 was unimodal for all 3 modeling algorithms. Habitat was predicted to be most

suitable when mean temperatures of the warmest month were between 25 °C and 31 °C, with a slight increase in suitability shown for the MARS model above 35 °C. For BIO9, suitability is highest for all 3 modeling algorithms when mean temperatures of the driest quarter (typically winter in the northern United States and summer in the south) are near −10 °C; suitability rapidly declines between −10 °C and 0 °C. Percent forest cover was an important predictor for the MARS and Maxent models, though both response curves were relatively flat with suitability scores changing <0.1 overall. These models also allow for interactions between predictors, which likely explains their relative importance despite their flat curves. The MARS model shows a gradual unimodal relationship with suitability, peaking at 60%, while the Maxent model predicted increasing suitability between 0% and 30%, the relationship is then flat up to 60%, and above this it declines slightly (Fig. 5). Generally, *A. phagocytophilum* suitability was highest in areas with low diurnal temperature ranges throughout the year, cooler maximum temperatures during the warmest months of the year (summer), and very cold average temperatures in the driest quarter of the year (winter in the northern United States); these conditions were found predominantly in counties in the Northeast and Upper Midwest (Figs. 1 and 4).

Babesia microti response curves.—All 8 predictors were used for the *Ba. microti* models, except for the mean temperature of the wettest quarter (BIO8) which was dropped from the MARS model (Table 1). Variable contributions differed across the models, showing less consistency than the *A. phagocytophilum* or *B. miyamotoi* models. The mean temperature of the driest quarter (BIO9) contributed close to 50% for GLM (47.1%) and Maxent (52.2%) but contributed only 14.2% to the MARS model. By contrast, the leading variable in the MARS model was BIO5 (37.8%), which contributed less than 5% in the GLM and Maxent models. Similarly, BIO2 was the second leading variable in the MARS model but contributed less than 7% to the GLM and Maxent models. For GLM and Maxent, all other variables contributed less than 20% (Table 1).

Despite substantial differences in the normalized contributions of each variable across the models, the shapes of the response curves were consistent. For BIO 9, the suitability score was highest when the mean temperature of the driest quarter was below 0 °C. For BIO2, suitability scores were highest when the mean diurnal temperature range was low (~8 °C) and dropped substantially when the mean diurnal temperature range was above 10 °C. High temperatures during dry periods (BIO9) explain why much of the Southeast is classified as unsuitable. The low temperatures during the driest quarter and the low diurnal temperature range through the year in the upper Midwest and Northeast explain, in part, the relatively higher suitability scores observed in those regions compared with the southeast (Figs. 2 and 4).

***Borrelia miyamotoi* response curves.—**All predictors were used by the modeling algorithms for *Bo. miyamotoi*, except for the mean temperature of the wettest quarter (BIO8) which was dropped from the MARS model (Table 1). Similar to the *A. phagocytophilum* models, the maximum temperature of the warmest month (BIO5), mean diurnal temperature range (BIO2), and mean temperature of the driest quarter (BIO9) were important predictors for all 3 modeling algorithms.

The relationship between BIO5 and habitat suitability was unimodal for all 3 models, with a peak in suitability when the maximum temperature of the warmest month was between 25 °C and 32 °C. For BIO2, suitability decreased as the mean diurnal temperature range increased from 6 to 12 or 14 °C. Suitability scores were highest when the mean temperature of the driest quarter was low, with suitability declining between −10 °C and 0 °C for the GLM and MARS models, and more slowly between −10 °C and 20 °C for the Maxent model. Similar to the *A. phagocytophilum* predictions, suitable habitat for *Bo. miyamotoi* was highest in areas with low diurnal temperature ranges, cool summer temperatures, and cold temperatures during the driest quarter, which were focused predominantly in the Northeast and Upper Midwest (Figs. 3 and 4).

Consensus Model Performance Against County-Level Observations

Anaplasma phagocytophilum model performance.—The area predicted to be highly suitable by the 3 modeling algorithms based upon the *A. phagocytophilum* dataset largely overlapped with one another. The consensus map yielded an overall accuracy (AUC) of 0.88. When the sum of sensitivity and specificity was maximized, sensitivity of the consensus map was 96%. Sensitivity was relatively high in the balanced output, therefore we concentrated on models set to 90% and 95% sensitivity thresholds. At 90% sensitivity, the consensus map predicted much of the Upper Midwest and Northeast to be suitable, with a small gap in Illinois, southern Indiana, western Ohio, and central Michigan. The specificity of the consensus map was 86%. In other words, 86% of counties lacking records of *I. scapularis* infected with *A. phagocytophilum* were classified as low-suitability. When sensitivity was increased to 95%, the suitable area expanded to cover much of Indiana and Ohio and the specificity decreased to 82%. The negative predictive values were high (>98%) but the positive predictive values were low (<43%) at both sensitivity settings (Table 3). Overall, there was a high false positive rate, with many counties classified as suitable that lacked observations of infected *I. scapularis* in northern New Hampshire, Massachusetts, and Maine. Additionally, some counties in northeastern Maine were not predicted to be suitable at either sensitivity level, despite *A. phagocytophilum* being present in one coastal county. Much of Michigan is classified as suitable, despite largely lacking observations of *A. phagocytophilum* outside of the perimeter of the state (Fig. 1).

Babesia microti model performance.—As with the *A. phagocytophilum* models, and despite differences in the relative contribution of the predictive variable, the area predicted to be highly suitable by the 3 modeling algorithms based upon the *Ba. microti* dataset largely overlapped with one another. However, owing to differences in the relative contributions of variables, the MARS model uniquely predicted highly suitable habitat in southeastern Wisconsin and in northern and eastern counties in the lower peninsula of Michigan where the Maxent and GLM models predicted low suitability. The consensus map yielded an overall accuracy (AUC) of 0.91. When the sum of sensitivity and specificity was maximized, sensitivity of the consensus map was 98%. As with the *A. phagocytophilum* models, sensitivity was relatively high in the balanced output, therefore we concentrated on models set to 90% and 95% sensitivity thresholds. At 90% sensitivity, the consensus map predicted much of the Northeast, south to Pennsylvania, to be suitable along with a large portion of Minnesota and Wisconsin in the Upper Midwest. The specificity of the consensus map at

90% sensitivity was 92%. When sensitivity was increased to 95% the suitable area expanded westward from the Northeast into Ohio and included portions of southern Michigan. The specificity at 95% sensitivity decreased to 89%. The negative predictive values were high (>99%) and positive predictive values were low (<36%) at both sensitivity settings (Table 3). Overall, there was a high false positive rate, with counties predicted to be suitable in much of Maine and Ohio that currently lack data for the presence of *Ba. microti*. Few counties with *Ba. microti* observations were predicted to be unsuitable, with the exception of some in northern Illinois and in western Virginia and coastal Maine. Michigan had little data for the presence of *Ba. microti*, therefore much of the state is classified as unsuitable (Fig. 2).

Borrelia miyamotoi model performance.—As with the models for the other 2 pathogens, the area predicted to be highly suitable by the 3 modeling algorithms based upon the *Bo. miyamotoi* dataset largely overlapped with one another. The consensus map yielded an overall accuracy (AUC) of 0.87. When the sum of sensitivity and specificity was maximized, sensitivity of the consensus map was 90%. As with the previous 2 consensus maps, sensitivity was relatively high in the balanced output, therefore we concentrated on models set to 90% and 95% sensitivity thresholds. At 90% sensitivity, the consensus map predicts much of the Northeast and Upper Midwest to be suitable, except for most of Illinois, parts of central Michigan and central Indiana. The specificity of the consensus map at 90% sensitivity was 85%. When sensitivity was increased to 95% the suitable area expanded northward in the Upper Midwest to cover the majority of Michigan and Minnesota. The specificity at 95% sensitivity was decreased to 78%. The negative predictive values were high (>98%) and positive predictive values were low (<36%) at both sensitivity settings (Table 3). Overall, there was a high false positive rate, with counties predicted to be suitable in much of Maine, New Hampshire, and Ohio that currently lack data for the presence of *Bo. miyamotoi*. Many counties in central Michigan are also classified as suitable, despite *Bo. miyamotoi* observations being concentrated along the perimeter of the state. Few counties with *Bo. miyamotoi* observations were predicted to be unsuitable, with those that were mostly concentrated in southern Indiana and Illinois (Fig. 3).

Comparison against Previously Generated *Bo. burgdorferi* s.s. Consensus Map

To prioritize counties where additional tick sampling and pathogen testing is needed to describe the current distribution of human pathogens in ticks with greater accuracy, we compared the consensus maps for *A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi* against consensus maps for *Bo. burgdorferi* s.s. that had been previously constructed (Burtis et al. 2022). The majority of counties (>98%) that were classified as highly suitable for *A. phagocytophilum*, *Ba. microti*, or *Bo. miyamotoi* were also highly suitable for *Bo. burgdorferi* s.s. A total of 1018 counties were classified as suitable for *Bo. burgdorferi* s.s. at 95% sensitivity. Among counties where the pathogen of interest had not been reported, we identified 431, 275, and 539 counties classified as highly suitable for *A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi*, respectively. These are considered high priority for surveillance efforts in order to confirm presence of these pathogens in counties classified as suitable. An additional 311 (*A. phagocytophilum*), 590 (*Ba. microti*), and 252 (*Bo. miyamotoi*) counties were classified as highly suitable for *Bo. burgdorferi* s.s. but not considered highly suitable for the pathogen of interest (either *A. phagocytophilum*,

Ba. microti, or *Bo. miyamotoi*). These are considered moderate priority for surveillance efforts; we expect with adequate numbers of ticks tested, the pathogens are likely to be detected in these counties, but our certainty of detection is lower compared with the high priority group. An additional 1670 (*A. phagocytophilum*), 1676 (*Ba. microti*), and 1672 (*Bo. miyamotoi*) counties were classified as low suitability for any of the 4 pathogens and were therefore categorized as low priority for field collection and pathogen testing efforts. Most counties classified as highly suitable for *Bo. burgdorferi* s.s., but not the other pathogens were clustered along the southern and western border of the suitable areas. The exception being that many of the counties in Ohio, Indiana, Michigan, Illinois, Virginia, and West Virginia that were predicted to be highly suitable for *Bo. burgdorferi* s.s., were classified as low suitability for *Ba. microti* (Fig. 6).

Discussion

We developed habitat suitability models based upon the county-level datasets for 3 *I. scapularis*-borne pathogens: *A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi*. These 3 pathogens are predicted to share similar suitable habitat to *Bo. burgdorferi* s.s. (Burtis et al. 2022) as their underlying transmission dynamics are similar; each relies predominantly on contact between *I. scapularis* (Eisen and Eisen 2018) and small mammal reservoirs that they share in common (LoGiudice et al. 2003, Barbour et al. 2009, Hersh et al. 2012, Keesing et al. 2012). Suitable habitat for these pathogens exists where vector and host communities interact in ways that facilitate their transmission. However, *Bo. burgdorferi* s.s. is more likely to be detected than the other 3 pathogens modeled here due to higher infection prevalence in *I. scapularis* populations and more extensive surveillance efforts (Fleshman et al. 2021, 2022, Lehane et al. 2021). As a result, we had more county presence points occurring over a larger geographic area to use as our model build set for *Bo. burgdorferi* s.s. compared with other pathogens. The output of habitat suitability models is highly dependent upon the distribution of the underlying datasets (Guisan et al. 2017) and this was observed in our models. For example, *Ba. microti* typically lags *Bo. burgdorferi* s.s. in establishment (Diuk-Wasser et al. 2014, Dunn et al. 2014) and had the most limited distribution of observation points upon which to build our habitat suitability models. As a result, model outputs showed a narrower range of predicted suitable habitat for *Ba. microti*. Given these limitations, we believe the resulting models are likely of limited value in identifying the full range of suitable habitat for each pathogen but are useful for prioritizing surveillance activities aimed at documenting the current distribution of human pathogens in host-seeking ticks. Specifically, we classified 431, 275, and 539 counties as high priority for tick collection and pathogen testing efforts due to their lack of existing records and high suitability scores for *A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi*, respectively. The majority of these counties were located in Indiana, Michigan, and Ohio in the Upper Midwest and Maine, New Hampshire, and Massachusetts in the Northeast.

Three predictors used in the *Bo. burgdorferi* s.s. model related to temperature and moisture were consistently highly influential for all 3 pathogens: mean diurnal temperature range (BIO2), maximum temperature of the warmest month (BIO5), and mean temperature of the driest month (BIO9). Diurnal temperature range (BIO2) is higher when there is less moisture in the air to regulate diurnal and nocturnal temperatures. Low moisture coupled with high

maximum temperatures during the warmest month (BIO5) will increase desiccation rates for host-seeking ticks, potentially reducing host-vector contact rates resulting in lower rates of pathogen transmission. In the interior southeastern United States, the driest quarter is often the summer, whereas it is generally the winter or fall in the Upper Midwest or Northeast (Fig. 4). Therefore, high mean temperatures during the driest quarter (BIO9) also likely contribute to increased potential for desiccation. It is possible that dry air temperatures in the winter do not contribute to mortality as strongly since *I. scapularis* tend to be in diapause in sheltered soil refugia during the cold winter months (Gray et al. 2016). Generally, hot and dry conditions were associated with lower suitability scores. These conditions decrease *I. scapularis* survival by increasing their rate of desiccation (Stafford 1994, Ginsberg et al. 2017, Burtis et al. 2019). Dry conditions can also decrease the amount of time that *I. scapularis* is actively questing (Schulze and Jordan 2003, Berger et al. 2014, Burtis et al. 2016). This would directly affect the survival of *I. scapularis* directly, and a reduction in the questing behavior of *I. scapularis* during dry conditions could result in decreased contact rates with reservoir hosts in southern or western states, potentially interrupting transmission cycles (Randolph and Storey 1999). Unfortunately, it is difficult to determine specifically how these predictors affect pathogen transmission cycles as they could be predictive of host distributions or other factors affecting tick-host contact rates.

The counties predicted to be suitable for *Bo. burgdorferi* s.s. are generally more extensive than predicted suitable counties for the other 3 pathogens (Fig. 6). This is likely due to the relatively limited presence data for *A. phagocytophilum*, *Ba. microti*, or *Bo. miyamotoi*-infected *I. scapularis* available to build these models, rather than a reflection on the true distribution of suitable habitat for these pathogens. These limited distributions are explained in part by differences in sampling strategies. Our build set was based on records in ArboNET, some of which were derived from literature review (Fleshman et al. 2021, 2022) and reflect greater effort to detect *Bo. burgdorferi* s.s. compared with other pathogens. However, even within counties where the sampling and testing efforts were similar, *Bo. burgdorferi* s.s. was still more likely to be detected owing to its higher prevalence of infection in *I. scapularis* (Lehane et al. 2021). This is reflected when examining a subset of 408 counties from our dataset wherein all ticks were tested for all 4 pathogens. *Bo. burgdorferi* s.s. was detected in more counties (258) than *A. phagocytophilum* (139 counties), *Ba. microti* (38 counties), or *Bo. miyamotoi* (101 counties).

We believe the true distributions of suitable habitats are more extensive than reflected by our models, highlighting the need for expanded field surveillance efforts to accurately define the current distribution of human pathogens in host-seeking ticks. For example, *Bo. miyamotoi* may also be more widespread than our models predict due to its capacity to be transmitted transovarially (Rollend et al. 2013), which may be the primary route of transmission (Lynn et al. 2022). Recently, *I. scapularis* collected from deer were found infected with *Bo. miyamotoi* in central Oklahoma (Smalley et al. 2022) far outside of the range predicted here. It is unknown if this detection represents adventitious infected ticks, or if the pathogen is established in this region, but this finding highlights the need for expanded field sampling across the range of *I. scapularis*. Despite *I. scapularis* being implicated as a vector relatively recently (Western et al. 1970, Telford et al. 1996, Gugliotta et al. 2013), resulting in a shorter period of time in which efforts were made to detect *Bo. miyamotoi*, observations cover a

similarly broad geographic area to that of *Bo. burgdorferi* s.s. and *A. phagocytophilum*., providing further evidence of its broad distribution in the United States. It is also worth noting that the range of suitable habitat for *A. phagocytophilum* we have reported here may be more extensive than the range of reported human anaplasmosis cases because our records do not differentiate between the human active-variant and deer-associated (v1) variants (Fleshman et al. 2022). However, these variants often co-occur geographically in the upper midwestern and northeastern United States (Massung et al. 2002, Courtney et al. 2003). Further surveillance efforts that differentiate among human disease-causing strains and wildlife-associated strains are needed to refine estimates for human risk of tickborne infections.

To provide current and accurate pathogen distribution maps, ongoing tick surveillance efforts for infected *I. scapularis* should focus on counties that lack observations for the pathogens of interest. Our models can be used to prioritize costly field collection efforts based on a county's likelihood of identifying tick-borne pathogens. Increased tick collection and pathogen testing efforts should be prioritized in western Wisconsin, Michigan, and Ohio in the Upper Midwest and New Hampshire, Maine, in the Northeast and Western Virginia in the Mid-Atlantic region. Broader surveillance for *Bo. miyamotoi* is likely needed given its capacity for transovarial transmission and its recent detection in Oklahoma (Smalley et al. 2022). As new data are generated, suitability models will likely need to be updated to more accurately predict the range of suitable habitat for each pathogen.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank our public health partners across the United States who worked to collect ticks and submit surveillance data to ArboNET or to CDC for pathogen testing. We would also like to thank those working as part of CDC DVBD BDB EET for their pathogen testing and data management efforts which allowed for the generation of the data used for our analyses. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of CDC or the Department of Health and Human Services.

References

Barbour AG, Bunikis J, Travinsky B, Hoen AG, Diuk-Wasser MA, Fish D, Tsao JI. Niche partitioning of Borrelia burgdorferi and Borrelia miyamotoi in the same tick vector and mammalian reservoir species. Am J Trop Med. 2009:81:1120.

Berger KA, Ginsberg HS, Gonzalez L, Mather TN. Relative humidity and activity patterns of Ixodes scapularis (Acari: Ixodidae). J Med Entomol. 2014:51(4):769–776. 10.1603/me13186 [PubMed: 25118408]

Breiman L Random forests. Mach Learn. 2001:45:5–32.

Brownstein JS, Holford TR, Fish D. A climate-based model predicts the spatial distribution of the Lyme disease vector Ixodes scapularis in the United States. Environ Health Perspect. 2003:111(9):1152–1157. 10.1289/ehp.6052 [PubMed: 12842766]

Burtis JC, Fahey TJ, Yavitt JB. Survival and energy use of Ixodes scapularis nymphs throughout their overwintering period. Parasitol. 2019:146:781–790.

Burtis JC, Foster E, Schwartz AM, Kugeler KJ, Maes SE, Fleshman AC, Eisen RJ. Predicting distributions of blacklegged ticks (Ixodes scapularis), Lyme disease spirochetes (Borrelia

burgdorferi sensu stricto) and human Lyme disease cases in the eastern United States. Ticks Tick Borne Dis. 2022:13(5):102000. 10.1016/j.ttbdis.2022.102000 [PubMed: 35785605]

Burtis JC, Sullivan P, Levi T, Oggenfuss K, Fahey TJ, Ostfeld RS. The impact of temperature and precipitation on blacklegged tick activity and Lyme disease incidence in endemic and emerging regions. Parasites Vectors. 2016:9:1–10. [PubMed: 26728523]

Campbell O, Krause PJ. The emergence of human Powassan virus infection in North America. Ticks Tick Borne Dis. 2020:11(6):101540. 10.1016/j.ttbdis.2020.101540 [PubMed: 32993949]

CDC. Centers for disease control and prevention guidelines: "Surveillance for Ixodes scapularis and pathogens found in this tick species in the United States". Fort Collins, CO: Centers for Disease Control and Prevention; 2018. [accessed 2022 Apr 10]. https://www.cdc.gov/ticks/resources/TickSurveillance_Iscapularis-P.pdf.

Courtney JW, Dryden RL, Montgomery J, Schneider BS, Smith G, Massung RF. Molecular characterization of Anaplasma phagocytophilum and Borrelia burgdorferi in Ixodes scapularis ticks from Pennsylvania. J Clin Microbiol. 2003:41(4):1569–1573. 10.1128/JCM.41.4.1569-1573.2003 [PubMed: 12682147]

Dahlgren FS, Heitman KN, Drexler NA, Massung RF, Behravesh CB. Human granulocytic anaplasmosis in the United States from 2008 to 2012: a summary of national surveillance data. Am J Trop Med. 2015:93:66.

Diuk-Wasser MA, Hoen AG, Cislo P, Brinkerhoff R, Hamer SA, Rowland M, Cortinas R, Vourc'h G, Melton F, Hickling GJ, et al. Human risk of infection with Borrelia burgdorferi, the Lyme disease agent, in eastern United States. Am J Trop Med. 2012:86:320.

Diuk-Wasser MA, Liu Y, Steeves TK, Folsom-O'Keefe C, Dardick KR, Lepore T, Bent SJ, Usmani-Brown S, Telford SR, Fish D, et al. Monitoring human babesiosis emergence through vector surveillance New England, USA. Emerg Infect Dis. 2014:20(2):225–231. 10.3201/eid2002.130644 [PubMed: 24447577]

Diuk-Wasser MA, Vourc'h G, Cislo P, Hoen AG, Melton F, Hamer SA, Rowland M, Cortinas R, Hickling GJ, Tsao JI, et al. Field and climate-based model for predicting the density of host-seeking nymphal Ixodes scapularis, an important vector of tick-borne disease agents in the eastern United States. Glob Ecol Biogeogr. 2010:19:504–514.

Dormann CF, Purschke O, Marquez JRG, Lautenbach S, Schroeder B. Components of uncertainty in species distribution analysis: a case study of the great grey shrike. Ecology. 2008:89:3371–3386. [PubMed: 19137944]

Dunn JM, Krause PJ, Davis S, Vannier EG, Fitzpatrick MC, Rollend L, Belperron AA, States SL, Stacey A, Bockenstedt LK, et al. Borrelia burgdorferi promotes the establishment of Babesia microti in the northeastern United States. PLoS One. 2014:9(12):e115494. 10.1371/journal.pone.0115494

Edwards TC Jr, Cutler DR, Zimmermann NE, Geiser L, Moisen GG. Effects of sample survey design on the accuracy of classification tree models in species distribution models. Ecol Model. 2006:199:132–141.

Eisen RJ, Eisen L. The blacklegged tick, Ixodes scapularis: an increasing public health concern. Trends Parasitol. 2018:34(4):295–309. 10.1016/j.pt.2017.12.006 [PubMed: 29336985]

Eisen RJ, Paddock CD. Tick and tickborne pathogen surveillance as a public health tool in the United States. J Med Entomol. 2021:58(4):1490–1502. 10.1093/jme/tjaa087 [PubMed: 32440679]

Eisen RJ, Eisen L, Beard CB. County-scale distribution of Ixodes scapularis and Ixodes pacificus (Acari: Ixodidae) in the continental United States. J Med Entomol. 2016:53(2):349–386. [PubMed: 26783367]

Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008:77(4):802–813. 10.1111/j.1365-2656.2008.01390.x [PubMed: 18397250]

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. A statistical explanation of MaxEnt for ecologists. Divers Distrib. 2011:17:43–57.

Estrada-Peña A Increasing habitat suitability in the United States for the tick that transmits Lyme disease: a remote sensing approach. Environ Health Perspect. 2002:110(635):635–640. 10.1289/ehp.02110635 [PubMed: 12117639]

Fleshman AC, Foster E, Maes SE, Eisen RJ. Reported county-level distribution of seven human pathogens detected in host-seeking Ixodes scapularis and Ixodes pacificus (Acari: Ixodidae) in the contiguous United States. J Med Entomol. 2022:59(4):1328–1335. [PubMed: 35583265]

Fleshman AC, Graham CB, Maes SE, Foster E, Eisen RJ. Reported county-level distribution of Lyme disease spirochetes, Borrelia burgdorferi sensu stricto and Borrelia mayonii (Spirochaetales: Spirochaetaceae), in host-seeking Ixodes scapularis and Ixodes pacificus ticks (Acari: Ixodidae) in the contiguous United States. J Med Entomol. 2021:58(3):1219–1233. 10.1093/jme/tjaa283 [PubMed: 33600574]

Fry J, Xian GZ, Jin S, Dewitz J, Homer CG, Yang L, Barnes CA, Herold ND, Wickham JD. Completion of the 2006 national land cover database for the conterminous United States. Photogramm Eng Remote Sensing. 2011:77:858–864.

Ginsberg HS, Albert M, Acevedo L, Dyer MC, Arsnoe IM, Tsao JI, Mather TN, LeBrun RA. Environmental factors affecting survival of immature Ixodes scapularis and implications for geographical distribution of Lyme disease: the climate/behavior hypothesis. PLoS One. 2017:12(1):e0168723. 10.1371/journal.pone.0168723 [PubMed: 28076359]

Gray EB, Herwaldt BL. Babesiosis surveillance—United States, 2011–2015. 2019. MMWR Surveil Summ. 2019:68:1–11.

Gray JS, Kahl O, Lane RS, Levin ML, Tsao JI. Diapause in ticks of the medically important Ixodes ricinus species complex. Ticks Tick Borne Dis. 2016:7(5):992–1003. 10.1016/j.ttbdis.2016.05.006 [PubMed: 27263092]

Gugliotta JL, Goethert HK, Berardi VP, Telford SR. Meningoencephalitis from Borrelia miyamotoi in an immunocompromised patient. N Engl J Med. 2013:368(3):240–245. 10.1056/NEJMoa1209039 [PubMed: 23323900]

Guisan A, Thuiller W, Zimmermann NE. Habitat suitability and distribution models: with applications in R. Cambridge, UK: Cambridge University Press; 2017.

Hahn MB, Jarnevich CS, Monaghan AJ, Eisen RJ. Modeling the geographic distribution of Ixodes scapularis and Ixodes pacificus (Acari: Ixodidae) in the contiguous United States. J Med Entomol. 2016:53(5):1176–1191. 10.1093/jme/tjw076 [PubMed: 27282813]

Han S How is Borrelia miyamotoi maintained among its vector, Ixodes scapularis, and vertebrate host population? Lansing, MI: Michigan State University; 2019.

Hersh MH, Tibbetts M, Strauss M, Ostfeld RS, Keesing F. Reservoir competence of wildlife host species for Babesia microti. Emerg Infect Dis. 2012:18(12):1951–1957. 10.3201/eid1812.111392 [PubMed: 23171673]

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 2005:25(15):1965–1978. 10.1002/joc.1276

Jensen OP, Seppelt R, Miller TJ, Bauer LJ. Winter distribution of blue crab Callinectes sapidus in Chesapeake Bay: application and cross-validation of a two-stage generalized additive model. Mar Ecol Prog Ser. 2005:299:239–255.

Johnson TL, Haque U, Monaghan AJ, Eisen L, Hahn MB, Hayden MH, Savage HM, McAllister J, Mutebi JP, Eisen RJ. Modeling the environmental suitability for Aedes (Stegomyia) aegypti and Aedes (Stegomyia) albopictus (Diptera: Culicidae) in the contiguous United States. J Med Entomol. 2017:54(6):1605–1614. 10.1093/jme/tjx163 [PubMed: 29029153]

Keesing F, Hersh MH, Tibbetts M, McHenry DJ, Duerr S, Brunner J, Killilea M, LoGiudice K, Schmidt KA, Ostfeld RS. Reservoir competence of vertebrate hosts for Anaplasma phagocytophilum. Emerg Infect Dis. 2012:18(12):2013–2016. 10.3201/eid1812.120919 [PubMed: 23171835]

Kugeler KJ, Schwartz AM, Delorey MJ, Mead PS, Hinckley AF. Estimating the frequency of Lyme disease diagnoses, United States, 2010–2018. Emerg Infect Dis. 2021:27(2):616–619. 10.3201/eid2702.202731 [PubMed: 33496229]

Larson RT, Bron GM, Lee X, Zembsch TE, Siy PN, Paskewitz SM. Peromyscus maniculatus (Rodentia: Cricetidae): an overlooked reservoir of tick-borne pathogens in the Midwest, USA? Ecosphere. 2021:12:e03831.

Lehane A, Maes SE, Graham CB, Jones E, Delorey M, Eisen RJ. Prevalence of single and coinfections of human pathogens in Ixodes ticks from five geographical regions in the United States,

2013–2019. Ticks Tick Borne Dis. 2021:12(2):101637. 10.1016/j.ttbdis.2020.101637 [PubMed: 33360805]

Liu C, White M, Newell G. Selecting thresholds for the prediction of species occurrence with presence-only data. J Biogeogr. 2013:40(4):778–789. 10.1111/jbi.12058

LoGiudice K, Ostfeld RS, Schmidt KA, Keesing F. The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. PNAS. 2003:100:567–571. [PubMed: 12525705]

Lynn GE, Breuner NE, Hojgaard A, Oliver J, Eisen L, Eisen RJ. A comparison of horizontal and transovarial transmission efficiency of Borrelia miyamotoi by Ixodes scapularis. Ticks Tick Borne Dis. 2022:13(5):102003. [PubMed: 35858517]

Massung RF, Mauel MJ, Owens JH, Allan N, Courtney JW, Stafford KC III, Mather TN. Genetic variants of Ehrlichia phagocytophila, Rhode Island and Connecticut. Emerg Infect Dis. 2002:8:467. [PubMed: 11996680]

Merow C, Smith MJ, Edwards TC Jr, Guisan A, McMahon SM, Normand S, Thuiller W, Wüest RO, Zimmermann NE, Elith J. What do we gain from simplicity versus complexity in species distribution models? Ecography. 2014:37(12):1267–1281. 10.1111/ecog.00845

Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. Ecol Model. 2006:190:231–259.

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl. 2009:19(1):181–197. 10.1890/07-2153.1 [PubMed: 19323182]

Pritt BS, Respicio-Kingry LB, Sloan LM, Schriefer ME, Replogle AJ, Bjork J, Liu G, Kingry LC, Mead PS, Neitzel DF, et al. Borrelia mayonii sp. nov., a member of the Borrelia burgdorferi sensu lato complex, detected in patients and ticks in the upper midwestern United States. Int J Syst Evol Microbiol. 2016:66:4878. [PubMed: 27558626]

QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation; 2009. http://qgis.org.

Randolph SE, Storey K. Impact of microclimate on immature tick-rodent host interactions (Acari: Ixodidae): implications for parasite transmission. J Med Entomol. 1999:36(6):741–748. 10.1093/jmedent/36.6.741 [PubMed: 10593075]

Rollend L, Fish D, Childs JE. Transovarial transmission of Borrelia spirochetes by Ixodes scapularis: a summary of the literature and recent observations. Ticks Tick Borne Dis. 2013:4(1–2):46–51. 10.1016/j.ttbdis.2012.06.008 [PubMed: 23238242]

Rosenberg R, Lindsey NP, Fischer M, Gregory CJ, Hinckley AF, Mead PS, Paz-Bailey G, Waterman SH, Drexler NA, Kersh GJ, et al. Vital signs: trends in reported vectorborne disease cases—United States and Territories, 2004–2016. MMRW. 2018:67:496.

Schulze TL, Jordan RA. Meteorologically mediated diurnal questing of Ixodes scapularis and Amblyomma americanum (Acari: Ixodidae) nymphs. J Med Entomol. 2003:40(4):395–402. 10.1603/0022-2585-40.4.395 [PubMed: 14680102]

Slajchert T, Kitron UD, Jones CJ, Mannelli A. Role of the eastern chipmunk (Tamias striatus) in the epizootiology of Lyme borreliosis in northwestern Illinois, USA. J Wildl Dis. 1997:33(1):40–46. 10.7589/0090-3558-33.1.40 [PubMed: 9027689]

Smalley R, Zafar H, Land J, Samour A, Hance D, Brennan RE. Detection of Borrelia miyamotoi and Powassan Virus lineage II (Deer Tick Virus) from Odocoileus virginianus harvested Ixodes scapularis in Oklahoma. Vector Borne Zoonotic Dis. 2022:22(4):209–216. 10.1089/vbz.2021.0057 [PubMed: 35446170]

Stafford KC III. Survival of immature Ixodes scapularis (Acari: Ixodidae) at different relative humidities. J Med Entomol. 1994:31:310–314. [PubMed: 8189424]

Talbert C, Talbert M. User documentation for the Software for Assisted Habitat Modeling (SAHM) package in VisTrails. Fort Collins (CO): US Geoglogical Survey; 2001.

Telford SR, Dawson JE, Katavolos P, Warner CK, Kolbert CP, Persing DH. Perpetuation of the agent of human granulocytic ehrlichiosis in a deer tick-rodent cycle. PNAS. 1996:93:6209–6214. [PubMed: 8650245]

Thornton PE, Running SW, White MA. Generating surfaces of daily meteorological variables over large regions of complex terrain. J Hydrol. 1997:190(3–4):214–251. 10.1016/s0022-1694(96)03128-9

Thornton PE, Thornton MM, Mayer BW, Wei Y, Devarakonda R, Vose RS, Cook RB. Daymet: daily surface weather data on a 1-km grid for North America, version 3. Oak Ridge (TN): ORNL DAAC; 2016.

Western KA, Benson GD, Gleason NN, Healy GR, Schultz MG. Babesiosis in a Massachusetts resident. N Engl J Med. 1970:283(16):854–856. 10.1056/NEJM197010152831607 [PubMed: 4989787]

Zimmermann NE, Yoccoz NG, Edwards TC Jr, Meier ES, Thuiller W, Guisan A, Schmatz DR, Pearman PB. Climatic extremes improve predictions of spatial patterns of tree species. PNAS. 2009:106:19723–19728. [PubMed: 19897732]

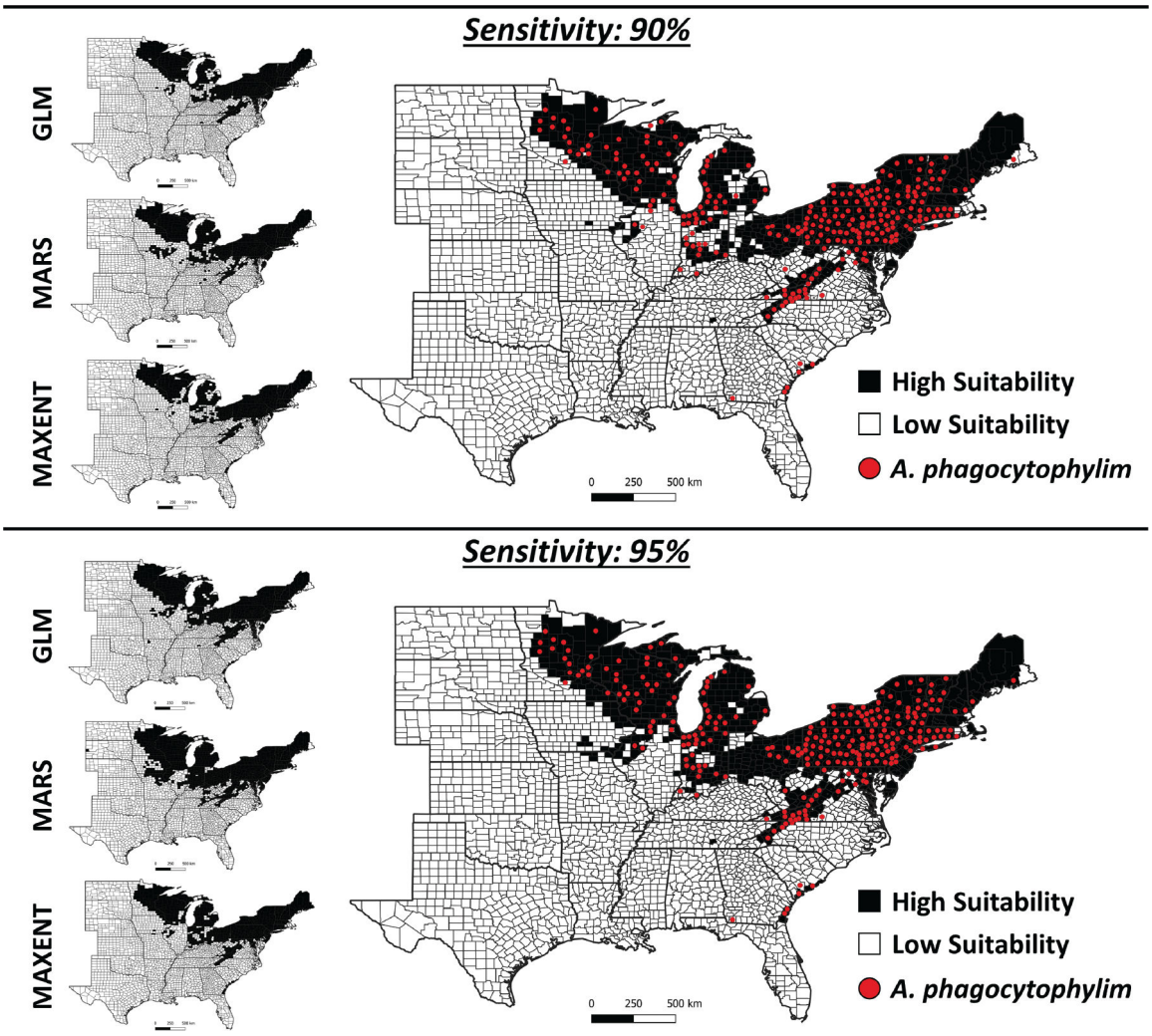

Fig. 1.

Counties predicted to be highly suitable for detecting *A. phagocytophilum*-infected *I. scapularis* based on 3 individual models (GLM, MARS, Maxent), shown in maps on the left. The large consensus maps to the right show counties predicted to be highly suitable when ≥2 of the individual models predict high-suitability. Points on the large maps represent counties where *A. phagocytophilum* was detected in field collected host-seeking *I. scapularis*. The maps are shown at 2 levels of sensitivity (90%, 95%).

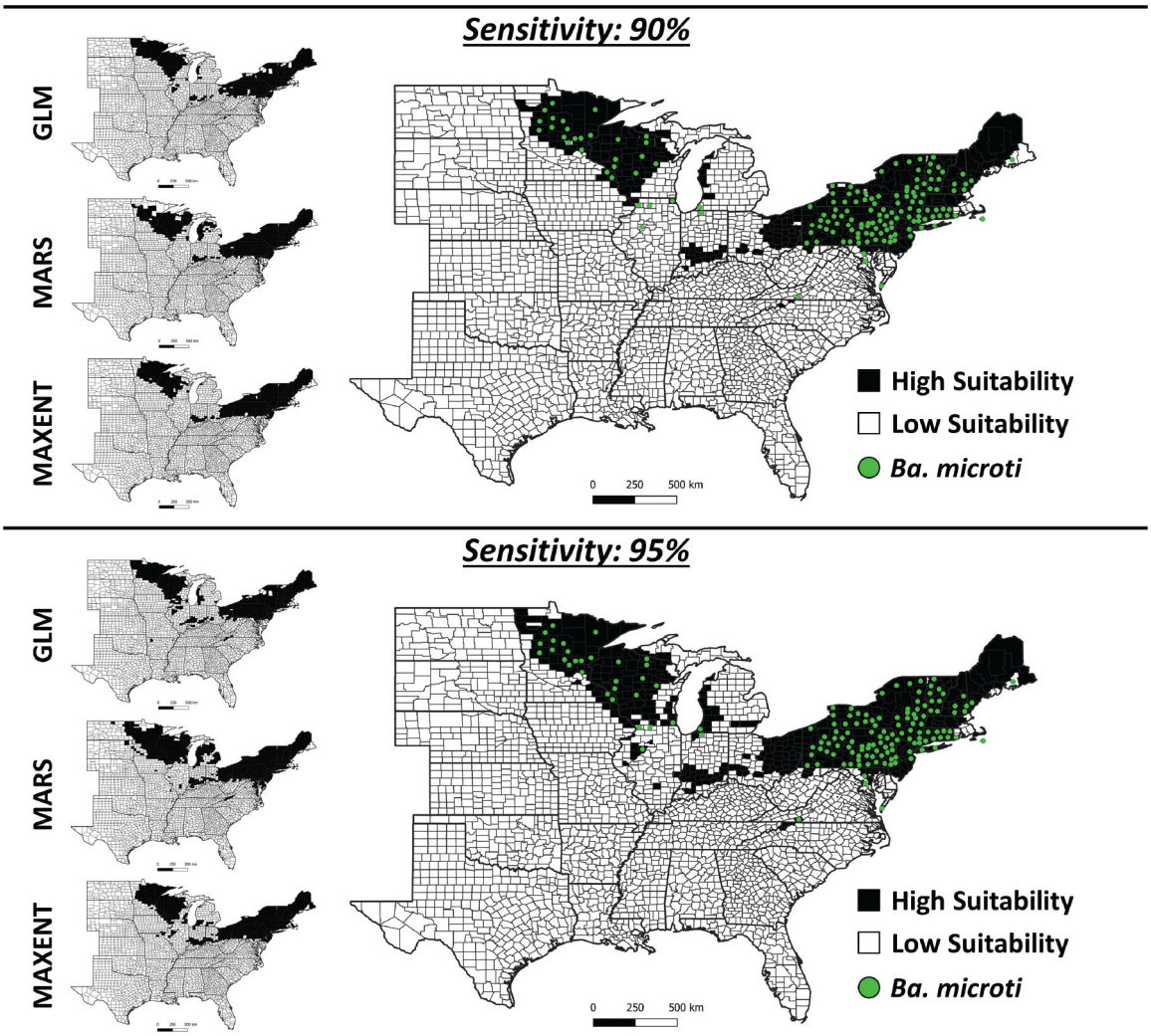

Fig. 2.

Counties predicted to be highly suitable for detecting *Ba. microti*-infected *I. scapularis* based on 3 individual models (GLM, MARS, Maxent), shown in maps on the left. The large consensus maps to the right show counties predicted to be highly suitable when ≥2 of the individual models predict high-suitability. Points on the large maps represent counties where *Ba. microti* was detected in field collected host-seeking *I. scapularis*. The maps are shown at 2 levels of sensitivity (90%, 95%).

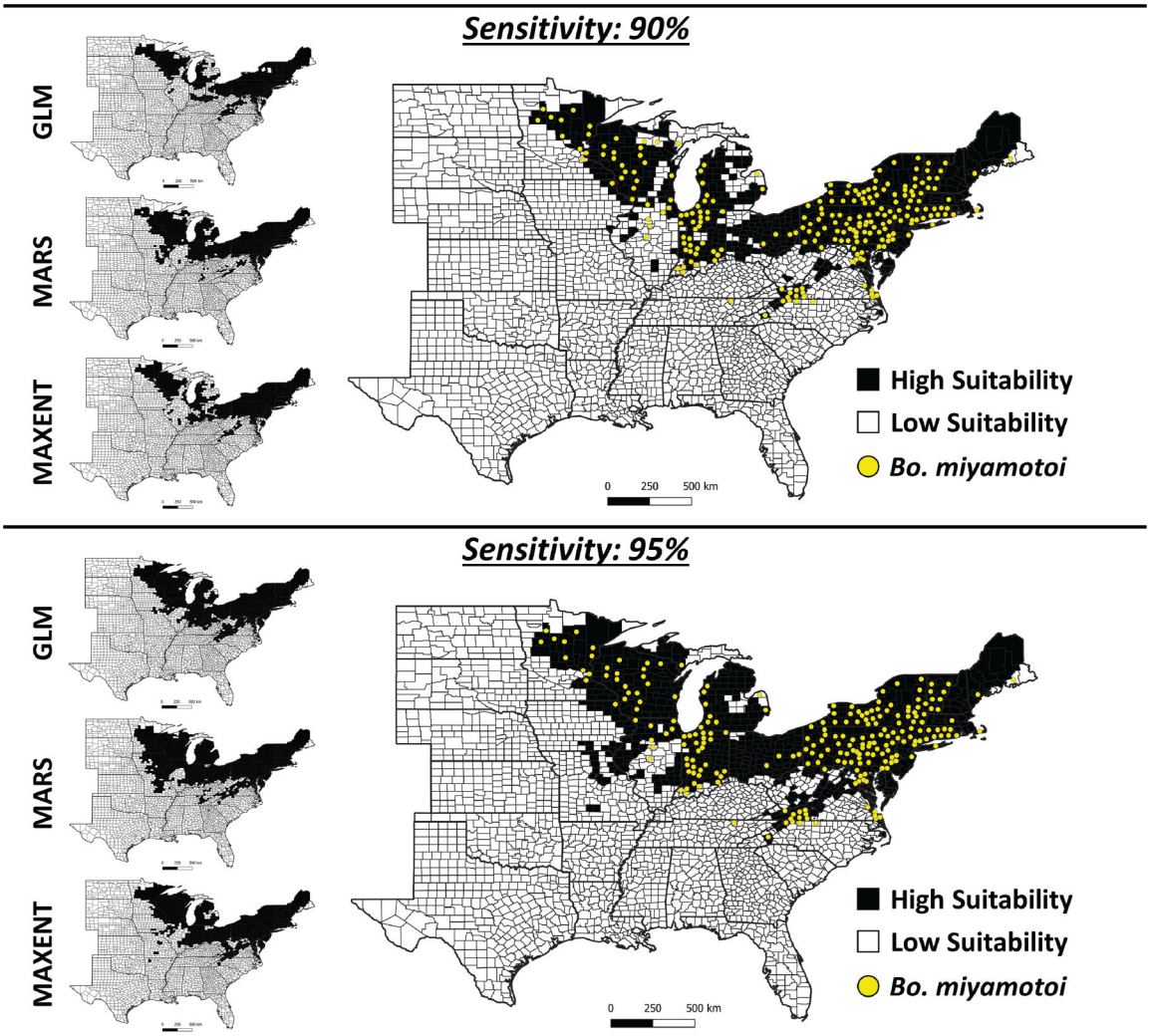

Fig. 3.

Counties predicted to be highly suitable for detecting *Bo. miyamotoi*-infected *I. scapularis* based on 3 individual models (GLM, MARS, Maxent), shown in maps on the left. The large consensus maps to the right show counties predicted to be highly suitable when ≥2 of the individual models predict high-suitability. Points on the large maps represent counties where *Bo. miyamotoi* was detected in field collected host-seeking *I. scapularis*. The maps are shown at 2 levels of sensitivity (90%, 95%).

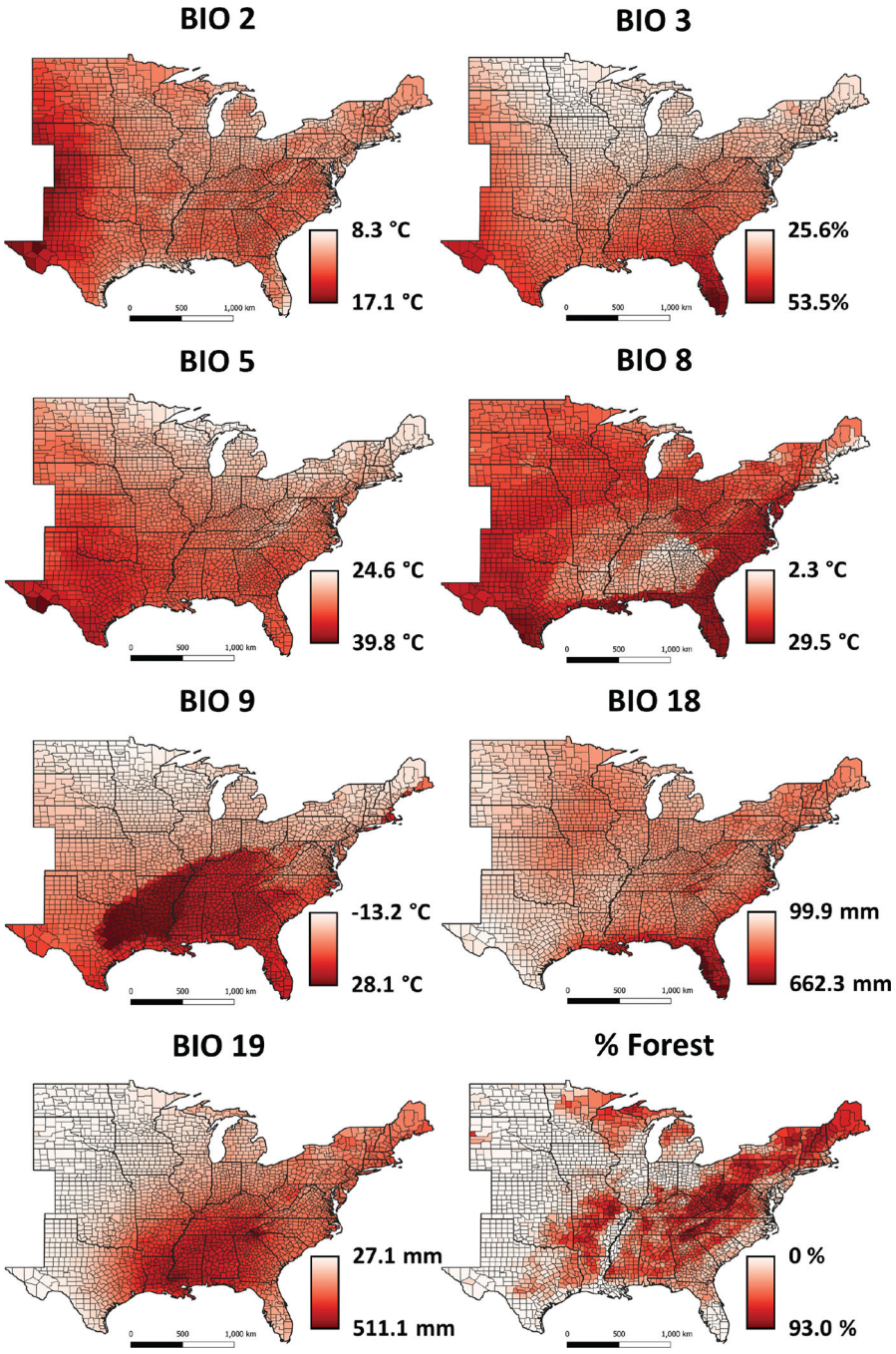

Fig. 4.

Maps showing the geographic distribution of the 8 predictors used by the 3 algorithms (GLM, MARS, Maxent) to model suitable habitat for the 3 pathogens (*A. phagocytophilum*, *Ba. microti*, *Bo. miyamotoi*).

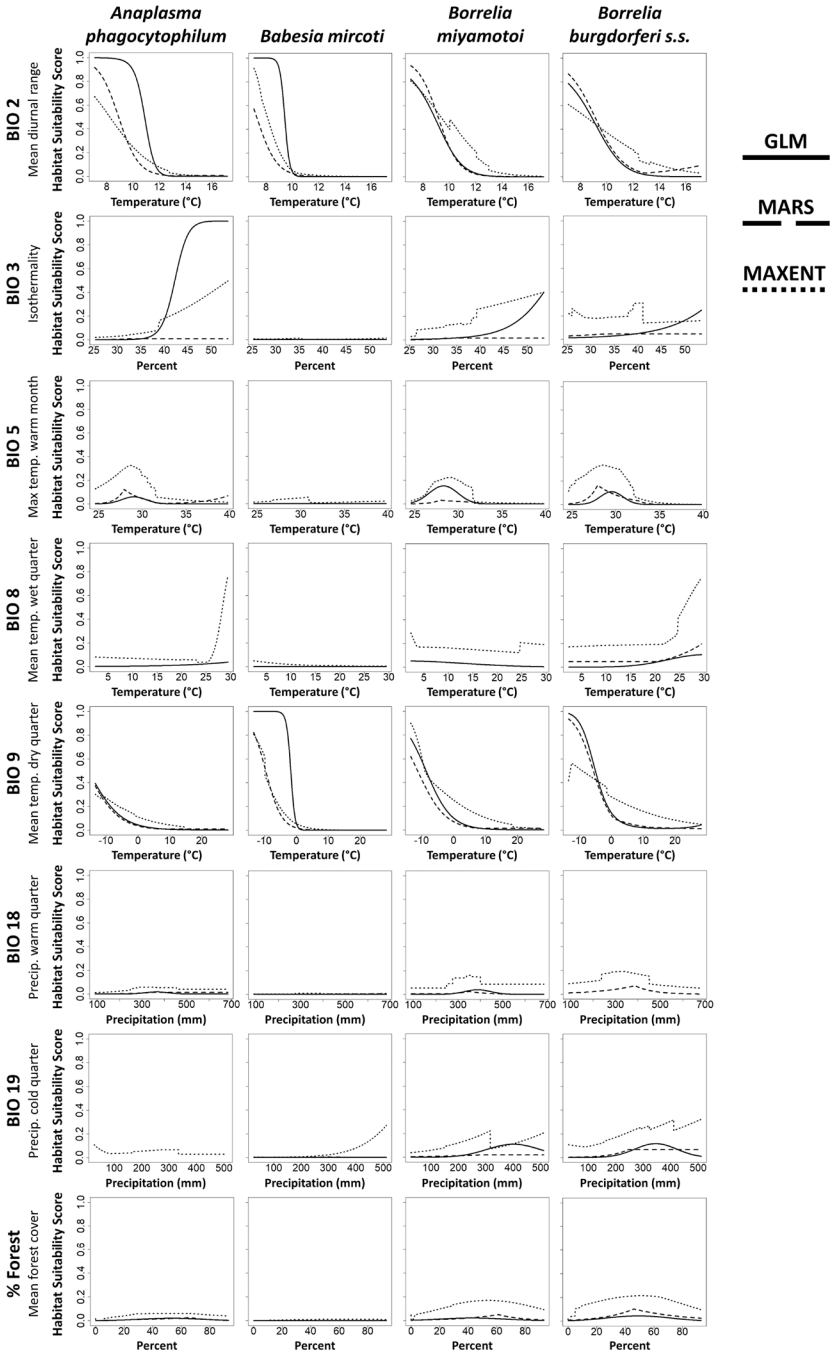

Fig. 5.

Response curves for the predictive variables included in the climate suitability models using the 3 pathogen datasets; *A. phagocytophilum*, *Ba. microti*, and *Bo. miyamotoi*. *Bo. burgdorferi* s.s. response curves are included for comparison. The different line types represent the modeling algorithms, solid lines are GLM, dashed lines are MARS, and dotted lines are Maxent. Not all parameters were used in all models.

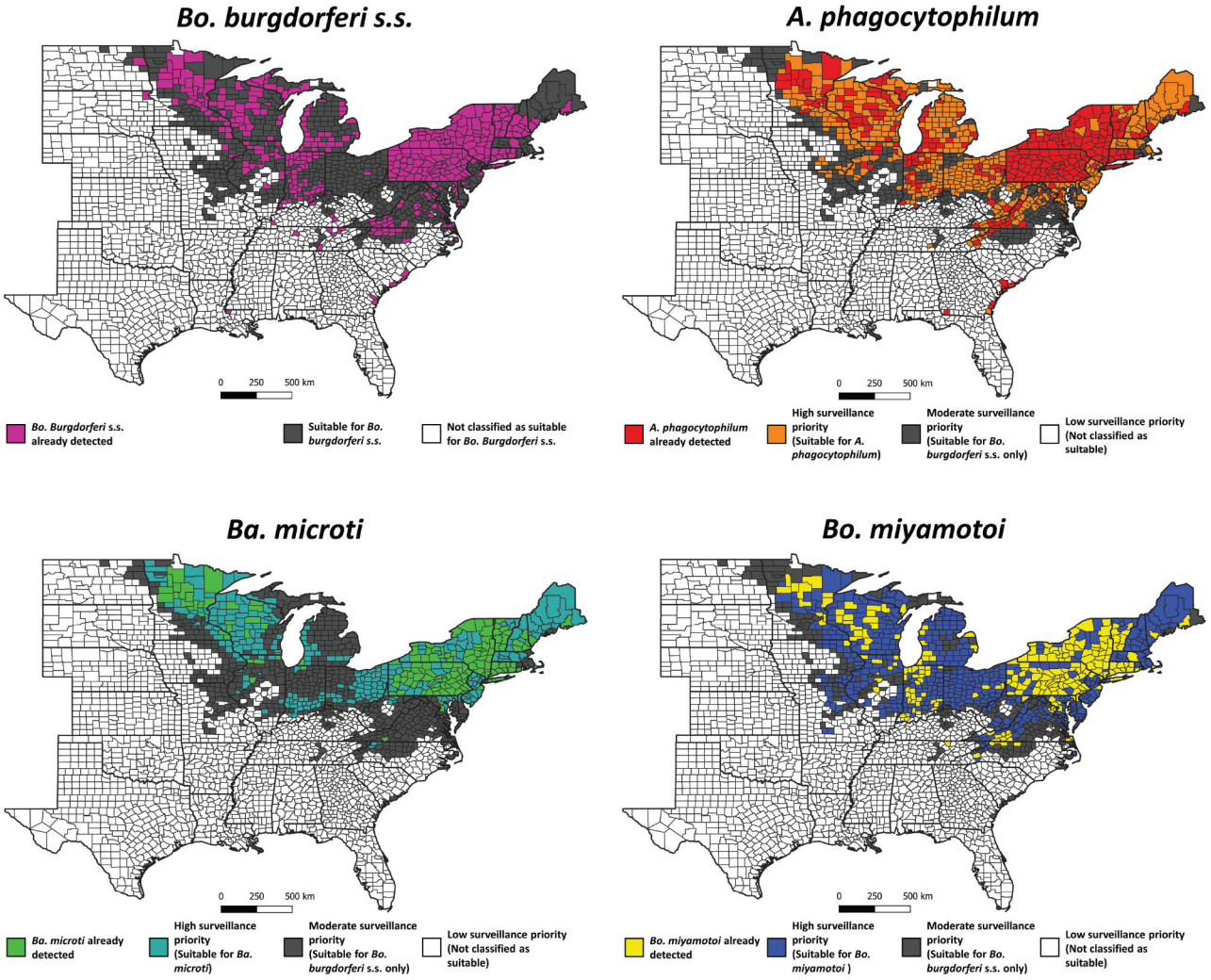

Fig. 6.

Maps showing the counties where pathogens have already been detected, those predicted to be suitable for *A. phagocytophilum*, *Ba. microti*, or *Bo. miyamotoi* at 95% sensitivity, and those predicted to be suitable for *Bo. burgdorferi* s.s. at 95% sensitivity. Maps show counties where pathogens have already been detected, those predicted to be suitable without records (highest priority for surveillance), and those without records that are suitable for *Bo. burgdorferi* s.s., but not other pathogens (moderate priority for surveillance). Counties predicted not to be suitable for any pathogens are the lowest priority for surveillance. Additional information regarding the *Bo. burgdorferi* s.s. models can be found in Burtis et al. (2022).

Table 1.

Relative contributions of the climate predictors selected by the modeling algorithms for the pathogen datasets; *A. phagocytophilum, Ba. microti,* and *Bo. miyamotoi*

	Normalized contribution values (%)		
Predictors	**GLM**	**MARS**	**Maxent**
A. phagocytophilum models			
Mean diurnal temp. range (BIO2)	24.9	19.3	11.9
Isothermality (BIO3)	12.0	7.8	7.4
Max temp. warmest month (BIO5)	21.9	27.4	15.3
Mean temp. of wettest quarter (BIO8)	1.3	—	2.4
Mean temp. of driest quarter (BIO9)	24.7	18.9	34.0
Precip. of warmest quarter (BIO18)	8.4	4.2	1.1
Precip. of coldest quarter (BIO19)	-	—	2.7
Percent forest cover	6.8	22.4	25.1
Ba. microti models			
Mean diurnal temp. range (BIO2)	6.8	25.5	5.0
Isothermality (BIO3)	4.6	1.9	4.7
Max temp. warmest month (BIO5)	4.9	37.8	4.3
Mean temp. of wettest quarter (BIO8)	3.3	—	1.3
Mean temp. of driest quarter (BIO9)	47.1	14.2	52.2
Precip. of warmest quarter (BIO18)	12.0	10.8	3.2
Precip. of coldest quarter (BIO19)	15.3	4.4	9.8
Percent forest cover	5.9	5.3	19.3
Bo. miyamotoi models			
Mean diurnal temp. range (BIO2)	11.6	15.0	14.9
Isothermality (BIO3)	3.4	4.3	5.0
Max temp. warmest month (BIO5)	20.8	54.5	24.3
Mean temp. of wettest quarter (BIO8)	1.8	—	1.5
Mean temp. of driest quarter (BIO9)	34.4	13.1	27.2
Precip. of warmest quarter (BIO18)	10.4	3.3	3.5
Precip. of coldest quarter (BIO19)	8.9	1.9	6.4
Percent forest cover	8.5	7.9	17.1

Normalized Contribution Values: Describes the contribution of each predictor to the predictive power of each of the 3 models (GLM, MARS, Maxent).

Table 2.

Model selection criteria and performance metrics for the testing and training runs of each modeling algorithm used to construct the consensus models based upon the 3 pathogen datasets; *A. phagocytophilum, Ba. microti,* and *Bo. miyamotoi*

	GLM		MARS		Maxent	
Performance metric	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**
A. phagocytophilum models						
AUC	0.94	0.95	0.93	0.93	0.94	0.96
Percent correctly classified	82.5	82.4	82.3	81.2	86.3	86.2
Mean threshold	0.08	0.08	0.09	0.07	0.28	0.26
Sensitivity	0.93	0.96	0.88	0.93	0.89	0.96
Specificity	0.81	0.81	0.82	0.80	0.86	0.85
PPV	0.37	0.37	0.36	0.35	0.43	0.43
NPV	0.99	0.99	0.98	0.99	0.98	0.99
Correlation coefficient	0.64	0.66	0.63	0.64	0.62	0.67
Ba. microti models						
AUC	0.95	0.97	0.95	0.96	0.96	0.98
Percent correctly classified	88.3	90.5	87.3	86.8	89.5	89.3
Mean threshold	0.05	0.07	0.06	0.06	0.20	0.17
Sensitivity	0.92	0.95	0.90	0.95	0.92	0.97
Specificity	0.88	0.90	0.87	0.86	0.89	0.89
PPV	0.32	0.37	0.30	0.30	0.34	0.35
NPV	0.99	0.99	0.99	0.99	0.99	0.99
Correlation coefficient	0.60	0.68	0.56	0.59	0.60	0.66
Bo. miyamotoi models						
AUC	0.93	0.94	0.92	0.93	0.93	0.95
Percent correctly classified	83.6	86.2	82.4	79.9	85.3	87.0
Mean threshold	0.11	0.12	0.09	0.08	0.31	0.31
Sensitivity	0.85	0.90	0.88	0.93	0.84	0.91
Specificity	0.84	0.86	0.82	0.79	0.85	0.87
PPV	0.33	0.37	0.31	0.29	0.35	0.39
NPV	0.98	0.99	0.99	0.99	0.98	0.99

	GLM		MARS		Maxent	
Performance metric	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**
Correlation coefficient	0.57	0.60	0.55	0.57	0.54	0.60

AUC: The AUC is a measure of model accuracy. A value of 1 indicates a “perfect” model and values ≥0.5 indicate a poor distinction between counties classified as high- or low-suitability.

Percent Correctly Classified: (True Positive + True Negative)/(True Positive + True Negative + False Positive + False Negative).

Mean threshold: Probability threshold at which presence is with the sum of sensitivity and specificity maximized.

Sensitivity: True Positive/(True Positive + False Negative).

Specificity: True Negative/(True Negative + False Positive).

Positive Predictive Value (PPV) = True Positive/(True Positive/False Positive).

Negative Predictive Value (NPV) = True Negative/(False Negative + True Negative).

Correlation Coefficient: Linear relationship between the field data and model output.

Table 3.

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and percent correctly classified (PCC) for the 3 consensus maps against the *A. phagocytophilum, Ba. microti,* and *Bo. miyamotoi* binary (reported/not reported) datasets, with the sensitivities of their 3 component models set at 2 different levels (90%, 95%)

Consensus layers	Model settings	Sensitivity	Specificity	PPV	NPV	PCC
A. phagocytophilum consensus map	Sens: 90%	90.1%	86.2%	43.3%	98.7%	86.6%
	Sens: 95%	96.1%	82.1%	38.7%	99.5%	83.6%
Ba. microti consensus map	Sens: 90%	92.2%	91.6%	39.9%	99.5%	91.6%
	Sens: 95%	96.8%	89.2%	35.1%	99.8%	89.6%
Bo. miyamotoi consensus map	Sens: 90%	89.7%	84.9%	35.8%	98.9%	85.3%
	Sens: 95%	94.8%	78.1%	29.0%	99.4%	79.6%

The model settings refer to the sensitivities set for the GLM, MARS, and Maxent models.