Landscape features predict the current and forecast the future geographic spread of Lyme disease

Allison M. Gardner[1], Natalie C. Pawlikowski[2], Sarah A. Hamer[3], Graham J. Hickling[4], James R. Miller[5], Anna M. Schotthoefer[6], Jean I. Tsao[7] and Brian F. Allan[2]

[1]School of Biology and Ecology, University of Maine, 5722 Deering Hall, Orono, ME 04469, USA
[2]School of Integrative Biology, University of Illinois, 505 S. Goodwin Avenue, Urbana, IL 61801, USA
[3]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA
[4]The Center for Wildlife Health, University of Tennessee Institute of Agriculture, Knoxville, TN 37966, USA
[5]Department of Natural Resources and Environmental Sciences, University of Illinois, 1102 S. Goodwin Ave, Urbana, IL 61801, USA
[6]Marshfield Clinic Research Institute, 1000 N Oak Ave, Marshfield, WI 54449, USA
[7]Department of Fisheries and Wildlife and Department of Large Animal Clinical Sciences, Michigan State University, 480 Wilson Rd., East Lansing, MI 48824, USA

(iD) AMG, 0000-0001-8438-0634; BFA, 0000-0002-1155-3271

Lyme disease, the most prevalent vector-borne disease in North America, is increasing in incidence and geographic distribution as the tick vector, *Ixodes scapularis*, spreads to new regions. We re-construct the spatial-temporal invasion of the tick and human disease in the Midwestern US, a major focus of Lyme disease transmission, from 1967 to 2018, to analyse the influence of spatial factors on the geographic spread. A regression model indicates that three spatial factors—proximity to a previously invaded county, forest cover and adjacency to a river—collectively predict tick occurrence. Validation of the predictive capability of this model correctly predicts counties invaded or uninvaded with 90.6% and 98.5% accuracy, respectively. Reported incidence increases in counties after the first report of the tick; based on this modelled relationship, we identify 31 counties where we suspect *I. scapularis* already occurs yet remains undetected. Finally, we apply the model to forecast tick establishment by 2021 and predict 42 additional counties where *I. scapularis* will probably be detected based upon historical drivers of geographic spread. Our findings leverage resources dedicated to tick and human disease reporting and provide the opportunity to take proactive steps (e.g. educational efforts) to prevent and limit transmission in areas of future geographic spread.

1. Background

The ongoing emergence of vector-borne diseases presents a major public health challenge, with the number of human cases reported in the United States tripling between 2004 and 2016 [1]. Numerous tick species are expanding beyond their historical range distributions and invading new regions. *Ixodes scapularis* (the blacklegged tick) currently is undergoing significant range expansions in two foci centred in the northeastern [2–4] and upper Midwestern states [5–7], and the range expansion of human cases of tick-borne disease is concomitant with *I. scapularis* range expansions [8–9]. The most important tick-borne pathogen of humans in North America is the spirochetal bacterium *Borrelia burgdorferi sensu stricto*, which causes an estimated 300 000 human cases of Lyme disease per year in the US [10].

Many wildlife species are important hosts for *I. scapularis*, natural reservoirs for *B. burgdorferi*, or both [11]. For example, white-tailed deer (*Odocoileus virginianus*) may contribute as dispersers and blood meal hosts of ticks despite

their incompetence for *B. burgdorferi* [12], while other species, such as white-footed mice (*Peromyscus leucopus*), may be important as both tick hosts and pathogen reservoirs [13]. Dispersal by long-distance movements of wildlife hosts that vary in blood meal host frequency and reservoir competence, such as white-tailed deer and migratory songbirds, may play a key role in ongoing range expansions for both tick and pathogen [14–19]. As a result, several alternative mechanisms of invasion for establishing an enzootic cycle have been proposed, including invasion by tick preceding invasion by pathogen, invasion by pathogen preceding invasion by tick, or by both occurring simultaneously [18].

Despite similarities between the hypothesized ecological drivers of Lyme disease emergence and expansion in the Northeast and Midwest, studies indicate differences between the regions in the dominant blood meal and reservoir host species that may contribute to differences in tick and pathogen dispersal [20–22] and the foci also are composed of closely related but distinct assemblages of circulating sequence types of *B. burgdorferi* [23–25]. At the landscape scale, the Midwest focus for Lyme disease historically has centred on forested regions of Wisconsin and Minnesota, which provide habitats that are highly suitable for *I. scapularis* populations and *B. burgdorferi* hosts in both regions [2,26,27], and where the first Midwestern cases of Lyme disease were reported [28]. Subsequently, expansions occurred into more agriculture-dominated landscapes in patterns that initially were consistent with the distribution of forested habitats [29,30]. For example, Cortinas & Kitron [14] found a pattern of the geographic spread of *I. scapularis* that correlated with the distribution of relatively scarce forest habitats in central Illinois: continuously forested riparian habitats associated with the Illinois River. More recently, however, established *I. scapularis* populations and *B. burgdorferi* transmission have been discovered in other habitats that do not occur in the Northeast, such as prairies [31–33]. Therefore, the geographic expansion of Lyme disease may be best understood by examining factors specific to each region [11,34,35].

Here, we use published literature and a national database reporting detection of *I. scapularis* and human Lyme disease, respectively, to evaluate the role of landscape features in predicting the invasion of the Lyme disease system in the Midwestern US. Specifically, we hypothesize that forest cover, adjacency to a river and adjacency to a county where *I. scapularis* previously has been reported are predictive of *I. scapularis* detection at the county level. Associations between each of these factors and detection of *I. scapularis* or human Lyme disease have been reported previously [14,36,37], so these factors were considered likely candidates for mediating the invasion of tick and pathogen at the spatial-temporal scale considered here. Additionally, we use these data to reconstruct the invasion history for both *I. scapularis* and human Lyme disease in the Midwest, predict counties in the Midwest most likely to become invaded next by *I. scapularis*, analyse whether the initial detection of *I. scapularis* is associated with a change in Lyme disease prevalence and, based on human Lyme disease incidence, infer counties where *I. scapularis* probably occurs already but remains undetected. We propose that a model based on the invasion history of Lyme disease in the Midwest can be used to predict the future expanding distribution of exposure risk and can be incorporated into proactive efforts for managing the continuing emergence of Lyme disease in the region. The unusually extensive tick and human disease occurrence data reported for Lyme disease offer a proof-of-concept for developing models to predict ongoing range expansions for other tick species and associated human illnesses.

2. Methods

(a) Data collection and processing

To reconstruct the invasion history of the blacklegged tick in the Midwestern US, we conducted a literature search (last updated 9 July 2018) for published, peer-reviewed scientific articles (excluding literature reviews) documenting the presence of *I. scapularis* in 10 states (figure 1*a*). In Web of Science, we used the title and abstract search terms 'Ixodes scapularis' OR 'Ixodes dammini' in combination with the names of each of the states. Articles were selected for inclusion in the study if they indicated the year and location of tick collections. For each article, including studies in which no *I. scapularis* were detected, the year of the study, state and counties sampled, and collection method were recorded (electronic supplementary material, table S1). Data were aggregated at the county level and in 5-year time intervals, the finest spatial-temporal resolution shared between *I. scapularis* and Lyme disease case data.

We recorded both the first report of *I. scapularis* in a county and the earliest date at which *I. scapularis* was considered established according to US Centers for Disease Control and Prevention (CDC) criteria. Discerning the specific number of ticks collected per county, and whether the criteria for establishment had been satisfied by these collections, required making several assumptions due to inconsistencies in data reporting among publications (electronic supplementary material, figure S1). By contrast, determining the first report of *I. scapularis* proved relatively straightforward; for this reason, we primarily used the dataset describing the first report of *I. scapularis* in each county for statistical analyses. We then compared those findings to the results produced by models based upon the CDC criteria for an established population to determine if these analyses yielded qualitatively similar conclusions.

Human case data were obtained from the CDC Lyme disease incidence database (http://www.cdc.gov/lyme/resources/ldpublicuse.csv and https://www.cdc.gov/lyme/resources/LD-Case-Counts-by-County-00-18.csv), which provided the number of cases per county according to CDC case definitions at the time of reporting in 5-year intervals (figure 1*b*). These intervals begin when Lyme disease became a nationally reportable disease in 1992; counties with no reports of Lyme disease were not considered 'apparently negative' until after 1992, when Lyme disease became a reportable disease. The Lyme disease case definition included confirmed cases from 1992 to 2007 and both confirmed and probable cases from 2008 onward; the case definitions were revised multiple times throughout the study period. To collect data for years prior to 1992, we conducted a literature search in Web of Science using the title and abstract search terms 'Lyme disease' in combination with the names of each of the states considered. The year of the study, the number of cases, state and county were recorded (electronic supplementary material, table S2).

To understand the spatial relationships among landscape features and the spread of *I. scapularis*, land cover data were obtained from the 2011 National Land Cover Database (NCLD; [38]). We classified forest cover as the sum of three NCLD land classes (i.e. deciduous, evergreen and mixed forest). The percentage forest cover in each county was calculated using the zonal histogram tool in ArcGIS 10.5 (Esri, Redlands, CA). Stream data were obtained from the National Hydrology Database plus v. 2 (http://www.horizon-systems.com/NHDPlus/NHDPlusV2_
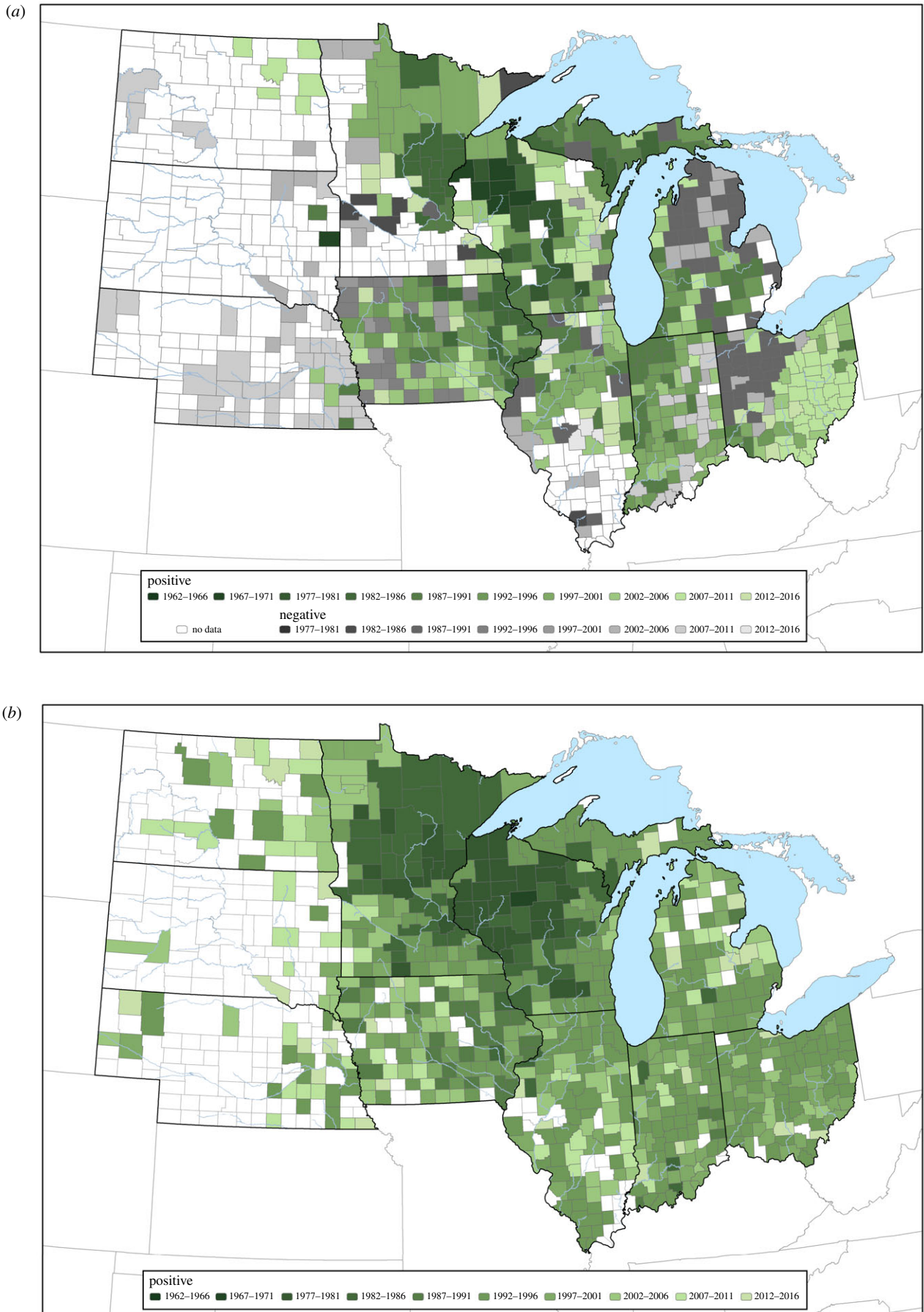
Figure 1. (a) Spread of *I. scapularis* in the Midwestern United States reconstructed using first county-level reports of the tick. Green shading represents the earliest year *I. scapularis* was collected in a county, grey shading represents the most recent year a county was sampled but no ticks were collected and white shading represents counties that never were sampled. (b) Spread of Lyme disease cases in the Midwestern United States. Green shading represents the earliest year Lyme disease was reported in a county.

data. php). Pre-calculated Strahler stream order [39] values were used to classify rivers by size for statistical analyses (electronic supplementary material, figure S2).

Our analyses were conducted in three phases, centred on (i) prediction of the first report of *I. scapularis* in a county, (ii) comparison of the timing and sequence of the first report of *I. scapularis* and the first reports of human cases of Lyme disease in a county and (iii) associations between the initial detection of *I. scapularis* and human Lyme disease incidence in a county. The goals, modelling techniques, inputs and outputs for each model discussed below are summarized in electronic supplementary material, table S3.

(b) Prediction of first report of *Ixodes scapularis*

First, we tested the hypotheses that the first report of *I. scapularis* in a county is predicted by forest cover, the county's adjacency to or inclusion of a river and adjacency to a county where *I. scapularis* had previously been reported. Tick report status was the binary response variable, coded as '0' if the tick had not been reported (hereafter a 'negative county') or '1' if the tick had been reported in a county (hereafter a 'positive county') by a given time interval (electronic supplementary material, figure S3). Logistic regression has been used by others to analyse vector presence/absence in multiple disease systems [9,37,40]. Here, we used survival analysis as an alternative [41] due to variation in the start and end time points of sampling efforts among counties (i.e. censoring) that violates the assumptions of logistic regression. Of 835 counties in the dataset, 266 counties were Type I right-censored (grey in figure 1a), meaning that the counties for which sampling has been documented remained negative for the presence of *I. scapularis* by the end of the final time interval, and 243 counties were randomly right-censored (white in figure 1a), meaning that the counties presumably have not been sampled for *I. scapularis* to date. The latter counties were excluded from further analyses due to the bias introduced by censoring driven by the experimenter, in this case by the choice not to search for ticks in certain geographical areas; thus, 592 counties that previously had been sampled for ticks were considered. These counties also were left-truncated, meaning that not all counties began surveillance for the tick simultaneously, and the year of initial sampling effort was specified throughout all phases of the analysis to account for this unavoidable limitation of the dataset [42].

A Cox regression model was fitted to tick report status using PROC PHREG in SAS v. 9.4 (SAS Institute Inc., Cary NC). We used the exact calculation of the partial likelihood to handle ties where the first report of the tick occurred during the same time interval in multiple counties and specified the time step of each county's entry into the model based on the year of initial sampling effort. We evaluated forest cover as a continuous predictor, adjacency to a river as a binary predictor and adjacency to a previously positive county as a time-dependent binary predictor in our models. Because the effect of adjacency to a river may be related to river size, we fitted multiple candidate models using adjacency to intermediate order streams (Strahler orders 6–8) and high-order streams (orders 9–10) as predictors; adjacency to low-order streams (orders 1–5) was excluded because each county contained a low-order stream. We also examined the possibility that the effect of adjacency to a previously positive county in predicting tick report status was associated with a time lag and included a positive adjacent county during the current time interval or in the one or two 5-year time intervals prior. The model among these permutations of predictors that minimized the Akaike information criterion (AIC) was considered the most parsimonious (electronic supplementary material, table S4). To determine whether *I. scapularis* establishment data would reach the same qualitative conclusions, we repeated the analysis for the 592 counties considered using the CDC establishment criteria [43] (electronic supplementary material, figure S4 and table S4).

To test the predictive capability of the regression model, we validated the selected model using the data from the first 10 time intervals (i.e. 1962–2011) and used this model to predict the new counties, among 168 counties that had been determined to be negative by the most recent sampling effort, that would become positive during the eleventh time interval (i.e. 2012–2016). The classification threshold used to determine whether a county was predicted to become positive (1) or remain negative (0) was 0.6. This threshold was selected using receiver operating characteristic (ROC) curve analysis. This analysis evaluates the ability of the model to classify between two groups (i.e. positive and negative counties); the closer the area under the curve (AUC) to 1, the better the model's ability to classify [44]. Here, we calculated ROC curves for classification thresholds between 0.2 and 0.8 in increments of 0.1 and identified the value that maximized the AUC (electronic supplementary material, table S6).

After model validation, we assimilated the data from the eleventh time interval (i.e. 2012–2016) and used the updated model to predict the new counties that will become positive during the as-yet incomplete reporting of the twelfth time interval (i.e. 2017–2021). We forecast 137 counties that had been determined to be negative by a most recent sampling effort in 1992 or later. Consistent with our model validation, the classification threshold used to determine whether a county was predicted to become positive (1) or remain negative (0) was 0.6.

(c) Comparison of first report of *Ixodes scapularis* and first report of human cases

Second, we tested the hypothesis that the first report of *I. scapularis* typically precedes the first human case of Lyme disease at the county level by using Kaplan–Meier analysis to compare the survival and hazard functions of reports of the blacklegged tick and human cases of Lyme disease. The survival function estimates the probability that a report of either a tick or a Lyme disease case has not occurred in a county by a given time interval, while the hazard function estimates the probability that a county which has not previously reported the occurrence of a tick or a Lyme disease case will do so during a given time interval. The analysis was conducted using PROC LIFETEST. Again, 592 counties were considered. Because Lyme disease has been a reportable illness since 1992, we assumed a lack of informative right-censoring for the human case data; however, non-positive counties prior to 1992 were considered left-truncated. Again, we specified the time step of each county's entry into the model based on the year of initial sampling effort.

(d) Presence of *Ixodes scapularis* and human Lyme disease incidence

Third, we tested the hypothesis that the first report of *I. scapularis* in a county is associated with a concomitant increase in Lyme disease incidence in humans. We fitted a general linear model with repeated measures to Lyme disease annual incidence per 100 000 people for 592 counties that previously had been searched for ticks. The discrete fixed predictors were report status of *I. scapularis*, time interval and their interaction. The repeated variable was county and an independent variance components covariance structure was modelled to reflect the autocorrelation of incidence over time within a county. Orthogonal contrasts generated using PROC IML were used to test the functional form of the time effect. The analysis was conducted using PROC MIXED.

Finally, recognizing that a relationship exists between Lyme disease incidence and the timing of the first report of *I. scapularis* within a county, we conducted an analysis to identify counties where we suspect *I. scapularis* is present yet remains undetected. For the 592 counties that previously had been searched for ticks,

Table 1. Parameter estimates, tests of significance, and risk ratios (95% confidence interval) for four predictors of county-level first reports of *I. scapularis* for the minimum AIC regression model from 1962 to 2016: forest cover, adjacency to medium-sized or large rivers (Strahler stream order) and adjacency to a county where ticks had been reported two time intervals previously.

parameter	d.f.	parameter estimate	standard error	χ^2	p	hazard ratio 95% CI lower	upper
forest cover	1	0.008	0.002	11.27	<0.001	1.003	1.013
medium-sized river	1	0.139	0.079	3.06	0.080	0.983	1.343
large river	1	−1.053	0.415	6.45	0.011	0.155	0.786
positive county ($t-2$)	1	0.762	0.128	35.69	<0.001	1.668	2.750

we fitted a binary logistic regression model with the presence (i.e. first report status) of *I. scapularis* as the response and Lyme disease incidence per 100 000 people during the most recent time interval (i.e. 2012–2016) as the predictor. We then used the model to identify counties where *I. scapularis* likely already has first occurred among 203 counties in which *I. scapularis* most recently has been reported as absent and 243 counties that have not been searched for *I. scapularis* to date. We considered the presence of *I. scapularis* to be 'suspected' if the probability of occurrence for a given county was greater than 0.7. The analysis was conducted using PROC LOGISTIC.

3. Results

(a) Invasion history for *Ixodes scapularis* and Lyme disease

We identified 65 peer-reviewed journal articles to reconstruct the invasion history of *I. scapularis* in the Midwest (electronic supplementary material, table S1), aggregated at the county level and in 5-year time intervals. The earliest reports of *I. scapularis* occurred from 1967 to 1970 in west-central Wisconsin. The most recent reports of *I. scapularis* identified in our literature search (last updated 9 July 2018) were from eastern South Dakota and east-central Illinois [19,45]. Overall, *I. scapularis* appears to be spreading in all cardinal directions within the Midwest, both based on the first report of the tick (figure 1*a*; electronic supplementary material, figure S3) and using CDC criteria [43] to infer establishment of a tick population (electronic supplementary material, figure S4). We used human case data from 1992 to 2016 from the CDC supplemented by 13 journal articles published prior to 1992 to determine the first occurrence of Lyme disease in each county (electronic supplementary material, table S2). While the initial spread of human cases mirrored the range expansion of *I. scapularis*, cases later were reported in most counties in the study region, including those in which *I. scapularis* had not yet been detected (figure 1*b*; electronic supplementary material, figure S5).

(b) Prediction of first report of *Ixodes scapularis*

After consideration of multiple candidate models using our *I. scapularis* invasion history database (electronic supplementary material, table S4), we found that per cent forest cover, adjacency to a medium-sized river [39], and adjacency to a county where the tick had been detected between 5 and 10 years previously, were positively correlated to first reports of *I. scapularis*, and adjacency to a large river was negatively correlated (table 1). The negative correlation with adjacency

to a large river may be due to large river size in this analysis being associated with lower forest cover than both medium-sized ($F = 11.37$; d.f. = 1, 581; $p < 0.001$) and small rivers ($F = 14.44$; d.f. = 1, 581; $p < 0.001$). An analysis using the presence of an established tick population according to CDC criteria as the response variable reached qualitatively similar conclusions (electronic supplementary material, table S5).

To validate the predictive capability of the regression model, we calibrated the selected model using data from 1962 to 2011 and used this model to predict counties in which the tick first would be detected between 2012 and 2016. ROC curve analysis demonstrated that, using a classification threshold of 0.6, the model effectively predicts invasion status (AUC = 0.95). Of 32 counties in which *I. scapularis* first was reported between 2012 and 2016, 29 (90.6%) were correctly predicted to become positive whereas three (9.4%) were incorrectly predicted to remain negative. Of 136 counties that remained negative during this time interval, 134 (98.5%) were correctly predicted to remain negative, while two (1.3%) were incorrectly predicted to become positive. Next, we applied the regression model to forecast tick report statuses in 2017–2021 for 136 counties; 42 counties are predicted to switch to positive invasion status while 95 counties are predicted to remain negative by the end of 2021 (figure 2).

(c) Comparison of first report of *Ixodes scapularis* and human cases of Lyme disease

Using the *I. scapularis* invasion history and Lyme disease case databases, we tested the hypothesis that the first report of *I. scapularis* typically precedes the first case of Lyme disease in a county. Initially, prior to Lyme disease becoming a nationally notifiable disease, first reports of *I. scapularis* and cases of Lyme disease occurred roughly simultaneously (electronic supplementary material, figure S6A and S6B). After 1991, however, Lyme disease cases began to precede the first reports of *I. scapularis* ($\chi^2 = 83.669$; d.f. = 1; $p < 0.001$). In particular, 65 counties first reported Lyme disease in 1992–1996, coinciding with Lyme disease becoming a reportable illness; from 2002 onward, the rate of new counties reporting their first Lyme disease cases remained steady, yet higher than the rate of first reports of *I. scapularis*.

(d) Presence of *Ixodes scapularis* and incidence of Lyme disease in humans

We tested the hypothesis that the first report of *I. scapularis* in a county is associated with a concomitant increase in Lyme
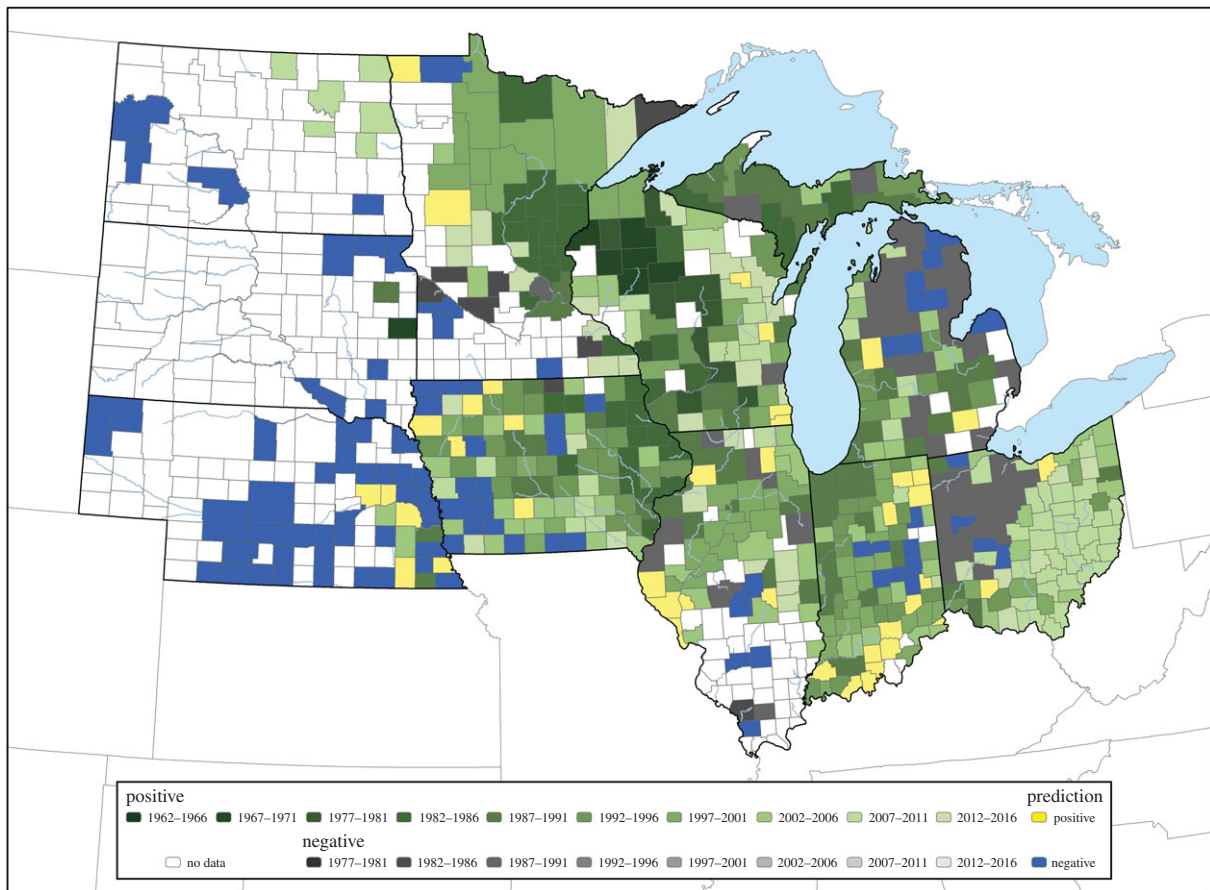
Figure 2. Predictions of counties that will report the first observation of *I. scapularis* and counties predicted to remain negative by 2021. Predictions are overlaid on the spread of *I. scapularis* in the Midwestern United States reconstructed using first county-level reports of the tick. Yellow shading represents counties that are predicted to report the first observation of *I. scapularis* and blue shading represents counties that are predicted to remain negative.

disease incidence. We found that Lyme disease incidence increased quadratically over time across the Midwest ($F = 19.85$; d.f. = 9, 5309; $p < 0.001$; figure 3a; electronic supplementary material, figure S7) and was significantly higher within any given county after the first report of the tick than before the first report of the tick ($F = 28.99$; d.f. = 1, 380; $p < 0.001$; figure 3b). Based on the modelled relationship between the presence of *I. scapularis* and human Lyme disease incidence ($\chi^2 = 21.432$; d.f. = 1; $p < 0.001$), we identified 31 counties in which *I. scapularis* is predicted to occur, but from which the tick had not yet been reported as of the time of our literature search. Among these 31 counties, mean Lyme disease incidence from 2012 to 2016 was 238.3 ± 7.8 s.e. cases per 100 000 people. These include nine counties for which the most recent published surveys failed to detect *I. scapularis* and 22 counties with no published surveys for *I. scapularis* to date (electronic supplementary material, table S7). Conversely, there were no counties in which *I. scapularis* has been detected, but from which no human Lyme disease cases have been reported.

4. Discussion

Globally, many arthropod disease vectors and the pathogens they transmit are expanding in distribution. Here, we demonstrate that a regression model that incorporates per cent forest cover, adjacency to rivers and adjacency to a county where the tick had been detected previously is predictive of detection of invasion by *I. scapularis*, arguably the

most important tick vector in North America [46]. In our model, which has both high sensitivity (90.6%) and high specificity (98.5%), the strongest predictor of invasion was adjacency to a county where the tick had been detected about 10 years previously. The importance of this proximity by *I. scapularis* to the detection of newly invaded counties, similar to the findings of [9] and [47], may be driven both by natural and social factors. While the relative contribution to tick dispersal of both short- and long-distance movements by wildlife remains an area of ongoing investigation [15], the close proximity of a reproducing population of *I. scapularis* may increase propagule pressure along invasion fronts, increasing the probability that dispersed ticks will establish populations in new areas [41]. Similarly, the presence of an established *I. scapularis* population nearby may enhance awareness, public health messaging and surveillance activities, increasing the probability of detection of newly established populations.

Per cent forest cover also was an important predictor of invasion, consistent with previous research which demonstrates that forest habitats are an important component of the ecological niche for *I. scapularis* and many of its vertebrate hosts [48]. Adjacency to a medium-sized river was positively correlated, and adjacency to a large river was negatively correlated, with invasion by *I. scapularis*. This latter correlation may be explained, in part, by the Mississippi River at its widest being associated with low forest cover. Overall, adjacency to a river was a less important predictor variable compared to forest cover. Therefore, previous reporting that rivers may be associated with invasion by *I. scapularis* [14]
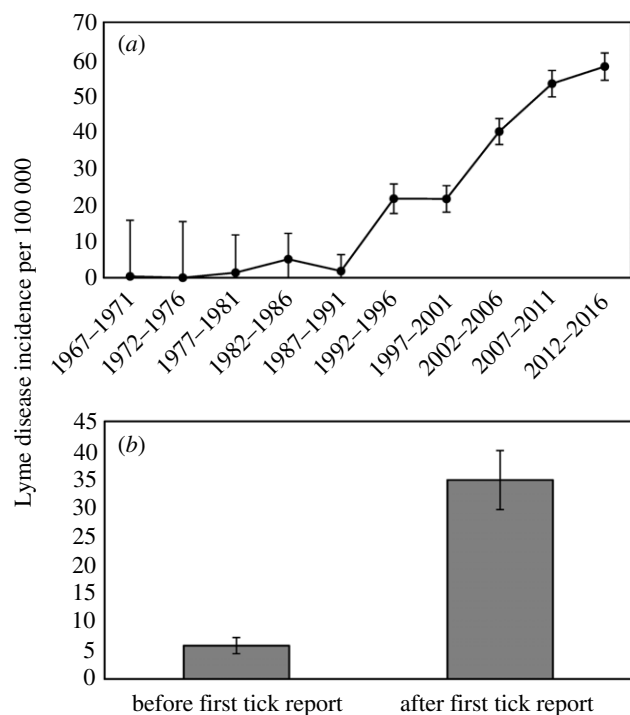
Figure 3. Mean (±1 s.e.) county-level incidence of Lyme disease (reported cases per 100 000 population) in humans (*a*) over time, 1967–2016, and (*b*) before and after the first report of *I. scapularis* in all counties.

and other tick species [49] may reflect close association of forest habitats, and movement of wildlife, with rivers in the Midwest.

Numerous species distribution models have identified associations between occurrence and density of *I. scapularis* and environmental variables that we did not consider in our model, such as soil order, forest type and elevation [47,50,51]. Climate conditions can impact tick survival and phenology, tick fecundity and tick questing behaviour, with consequences for host encounter rates and recruitment [52–56]. In particular, it is widely hypothesized that low winter temperatures constrain the distribution of ticks and climate warming has contributed to the geographic range expansion of *I. scapularis* within recent decades [57,58]. Yet there is evidence that effects of cold ambient air temperature on overwinter tick survival may be mitigated by microhabitat conditions such as the presence of insulating leaf litter or snowpack [59]. We chose to exclude climate variables from this analysis because aggregating these data over 5-year intervals and scaling to the county resolution to which our study was confined could produce unreliable estimates and relationships to tick detection [60]. Similarly, although *I. scapularis* densities generally appear higher in deciduous compared to coniferous forest [50], in the county spatial units considered in our model, forest type is less likely to determine presence/absence of ticks than simply whether forest habitat is available. Our more simplified regression model based on landscape features and predicted movement patterns of hosts accurately predicted the pattern and probability of invasion and could be incorporated into a proactive framework for managing the ongoing emergence of Lyme disease, including surveillance efforts that typically are operationalized at the county scale [61].

Prior to Lyme disease becoming a nationally reportable illness, first reports of *I. scapularis* and human cases of Lyme disease occurred roughly simultaneously within counties. Subsequently, however, first reports of Lyme disease cases began to precede the first reports of *I. scapularis*, and at present, more counties report Lyme cases than detection of its tick vector. This may be due to a strong effect of the designation of Lyme disease as a nationally reportable illness, higher reporting of travel-associated cases of Lyme illness and undetected populations of *I. scapularis*. We tested the hypothesis that the first report of *I. scapularis* in a county is associated with an increase in Lyme disease incidence and indeed we found incidence increased in counties after the first report of the tick. An important caveat, however, is that the detection of the tick may enhance awareness of Lyme disease risk among residents and local medical professionals. Using this relationship, we identified 31 counties where we suspect *I. scapularis* already occurs yet remains undetected as of our literature search ending in 2018, providing a warning that undetected tick populations could be causing local clusters of Lyme disease cases.

We elected to focus on the first published report of *I. scapularis* as opposed to reporting based on the CDC criteria of collection within 1 year of either two *I. scapularis* specimens belonging to different life stages or six specimens belonging to a single life stage [43]. This distinction differentiates our effort from a 2016 update to nationwide reporting data of *I. scapularis* and *Ixodes pacificus*, which categorized tick reports based on the CDC criteria [3]. We made this decision first because reporting in the literature of life stage and the number of *I. scapularis* collected was inconsistent and required us to make numerous assumptions to determine if these criteria had been met for a given location (electronic supplementary material, figure S1). Second, the CDC criteria do not require any standardization by search effort, which often also was not reported. For example, a recent investigation was the first to report according to the CDC criteria the establishment of *I. scapularis* in South Dakota [45]. Because search effort was carefully reported in this study, we know it required 23 person-hours of drag/flag sampling to collect seven adult *I. scapularis*. Thus, the criteria for establishment were met, but because of intensive sampling. In the absence of information on search effort from many studies, we believe first published report to be potentially a more reliable indicator of the arrival of *I. scapularis* in a county. We also repeated our analyses based upon a reconstructed invasion history following the CDC criteria for established populations and reached qualitatively comparable conclusions, although the timeline of invasion occurred later for some counties compared to first report (electronic supplementary material, table S4 and figure S4). Our findings were qualitatively similar to those of [3] in terms of which counties in the Midwest have become invaded since [43] and which counties are predicted to become invaded in the future.

We took advantage of data made possible by the designation of Lyme disease as a nationally reportable illness after 1991 and extensive surveillance efforts by many researchers to characterize the distribution of *I. scapularis*. However, there could be important sources of bias in using these data, including variation in surveillance efforts, imperfect reporting of tick occurrence and human cases [62], assumptions we made to glean data from literature, limitations in using the first discovery of either *I. scapularis* or Lyme disease, and the spatial-temporal grain used in the analyses. Lyme disease cases are reported in the county of residence rather than the location of

presumed exposure, and citizen-submitted and veterinary tick specimens may be subject to similar inaccuracies. Furthermore, our predictions concerning future invasions partly are driven by historical rates of detection of *I. scapularis* or Lyme disease. A significant increase in surveillance efforts due to additional funding or research interest could, for example, increase the probability of discovery of undetected populations of *I. scapularis*. The absence of information from most studies concerning search effort makes it difficult to address how detection rates may change if search effort increases. This is another argument to standardize sampling and report search effort across studies [63]. Guidelines proposed in 2019 to standardize sampling methods and locations will aid in this effort (https://www.cdc.gov/ticks/resources/TickSurveillance_Iscapularis-P.pdf).

Our analysis suggests that within the Midwest, the distribution of the Lyme disease system has expanded considerably over the previous 50 years, greatly increasing the prevalence and geographic distribution of human Lyme disease in this region. Moreover, our predictions suggest this expansion is ongoing, and that the vector tick probably already is established in some areas from where it has not yet been detected or published (as is the case, for example, in Michigan; J.I.T. 2020, unpublished data). The regression model used here appears to predict accurately in the short term (i.e. 5-year windows) the pattern and probability of invasion based on landscape features and could be incorporated into a more proactive framework for managing the ongoing emergence of Lyme disease. This could include an 'early warning' system used to notify medical and public health professionals that there is an imminent risk for the local establishment and even environmental interventions to attempt to prevent it [64]. Future efforts should build upon and improve the utility of our model by using techniques (e.g. Bayesian models) that lend themselves well to long-term forecasting and assimilation of new data, such as the detection of *I. scapularis* in a new county, to aid public health decision making in real time. To improve access to tick occurrence data, we encourage researchers and public health agencies to publish tick surveillance efforts more frequently, and we also encourage the development of a publicly available curated database, with associated metadata of surveillance methods and search effort, to which researchers could report surveillance data.

5. Conclusion

In addition to its public health significance, this study demonstrates the potentially fruitful yet underexplored intersection between disease ecology and invasion biology to understand the emergence and spread of vector-borne disease systems. Spatial configuration and type of landscape features are well known to influence the outcomes of biological invasions, including altering propagule pressure and the receptivity of the colonized habitat. Our analytical approach rooted in these principles of invasion biology can help generate new understanding of the spatial dynamics of, and potential to mitigate, the geographic spread of the ecologically complex Lyme disease system. Moreover, regionally widespread blacklegged tick surveillance efforts and the notifiable illness status of Lyme disease yield extensive occurrence datasets for both vector and pathogen, overcoming the low detection probabilities that typically constrain analysis of the early spread of invasive species. Further investment by public health agencies in intensified active tick surveillance may serve to improve power and performance of models to predict the future spread of disease vector ticks, as demonstrated in this study, as the width of prediction intervals decreases, model sensitivity and specificity increase. Thus, through surveillance efforts that are both systematic and extensive, vector-borne disease systems may provide more general insights into the landscape processes that drive biological invasions.

References

1. Rosenberg R *et al.* 2018 Vital signs: trends in reported vectorborne disease cases—United States and Territories, 2004–2016. *MMWR Morb. Mortal. Wkly. Rep.* **67**, 496. (doi:10.15585/mmwr.mm6717e1)

2. Diuk-Wasser MA *et al.* 2012 Human risk of infection with *Borrelia burgdorferi*, the Lyme disease agent, in eastern United States. *Am. J. Trop. Med. Hyg.* **86**, 320–327. (doi:10.4269/ajtmh.2012.11-0395)

3. Eisen RJ, Eisen L, Beard CB. 2016 County-scale distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the continental United States. *J. Med. Entomol.* **53**, 349–386. (doi:10.1093/jme/tjv237)

4. Hickling GJ, Kelly JR, Auckland LD, Hamer, SA. 2018 Increasing prevalence of *Borrelia burgdorferi sensu stricto*–infected blacklegged ticks in Tennessee Valley, Tennessee, USA. *Emerg. Infect. Dis.* **24**, 1713. (doi:10.3201/eid2409.180343)

5. Hamer SA, Hickling GJ, Walker ED, Tsao, JI. 2014 Increased diversity of zoonotic pathogens and *Borrelia burgdorferi* strains in established versus incipient *Ixodes scapularis* populations across the Midwestern United States. *Infect. Genet. Evol.* **27**, 531–542. (doi:10.1016/j.meegid.2014.06.003)

6. Lantos PM *et al.* 2017 Geographic expansion of Lyme disease in Michigan, 2000–2014. *Open Forum Infect. Dis.* **4**, 1. (doi:10.1093/ofid/ofw269)

7. Oliver JD, Bennett SW, Beati L, Bartholomay LC. 2017 Range expansion and increasing *Borrelia burgdorferi* infection of the tick *Ixodes scapularis* (Acari: Ixodidae) in Iowa, 1990–2013. *J. Med. Entomol.* **54**, 1727–1734. (doi:10.1093/jme/tjx121)

8. Kugeler KJ, Farley GM, Forrester JD, Mead PS. 2015 Geographic distribution and expansion of human Lyme disease, United States. *Emerg. Infect. Dis.* **21**, 1455–1457. (doi:10.3201/eid2108.141878)

9. Bisanzio D, Fernández MP, Martello E, Reithinger R, Diuk-Wasser MA. 2020 Current and future spatiotemporal patterns of Lyme disease reporting in the Northeastern United States. *JAMA Netw. Open* **3**, e200319. (doi:10.1001/jamanetworkopen.2020.0319)

10. Hinckley AF *et al.* 2014 Lyme disease testing by large commercial laboratories in the United States.

Clin. Infect. Dis. **59**, 676–681. (doi:10.1093/cid/ciu397)

11. Halsey SJ, Allan BF, Miller JR. 2018 The role of *Ixodes scapularis*, *Borrelia burgdorferi* and wildlife hosts in Lyme disease prevalence: a quantitative review. *Ticks Tick Borne Dis.* **9**, 1103–1114. (doi:10.1016/j.ttbdis.2018.04.006)

12. Telford III SR, Mather TN, Moore SI, Wilson ML, Spielman A. 1988 Incompetence of deer as reservoirs of the Lyme disease spirochete. *Am. J. Trop. Med. Hyg.* **39**, 105–109. (doi:10.4269/ajtmh.1988.39.105)

13. Hanincová K, Kurtenbach K, Diuk-Wasser M, Brei B, Fish D. 2006. Epidemic spread of Lyme borreliosis, northeastern United States. *Emerg. Infect. Dis.* **12**, 604. (doi:10.3201/eid1204.051016)

14. Cortinas MR, Kitron U. 2006 County-level surveillance of white-tailed deer infestation by *Ixodes scapularis* and *Dermacentor albipictus* (Acari: Ixodidae) along the Illinois River. *J. Med. Entomol.* **43**, 810–819. (doi:10.1093/jmedent/43.5.810)

15. Ogden NH *et al.* 2008 Role of migratory birds in introduction and range expansion of *Ixodes scapularis* ticks and of *Borrelia burgdorferi* and *Anaplasma phagocytophilum* in Canada. *Appl. Environ. Microbiol.* **74**, 1780–1790. (doi:10.1128/AEM.01982-07)

16. Brinkerhoff RJ *et al.* 2011 Regional variation in immature *Ixodes scapularis* parasitism on North American songbirds: implications for transmission of the Lyme pathogen, *Borrelia burgdorferi*. *J. Med. Entomol.* **48**, 422–428. (doi:10.1603/ME10060)

17. Brinkerhoff RJ, Folsom-O'Keefe CM, Tsao K, Diuk-Wasser MA. 2011. Do birds affect Lyme disease risk? Range expansion of the vector-borne pathogen *Borrelia burgdorferi*. *Front. Ecol. Environ.* **9**, 103–110. (doi:10.1890/090062)

18. Hamer SA, Tsao JI, Walker ED, Hickling GJ. 2010 Invasion of the Lyme disease vector *Ixodes scapularis*: implications for *Borrelia burgdorferi* endemicity. *Ecohealth* **7**, 47–63. (doi:10.1007/s10393-010-0287-0)

19. Parker CM, Miller JR, Allan BF. 2017 Avian and habitat characteristics influence tick infestation among birds in Illinois. *J. Med. Entomol.* **54**, 550–558. (doi:10.1093/jme/tjw235)

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

21. Johnson RC, Kodner C, Jarnefeld J, Eck DK, Xu Y. 2011 Agents of human anaplasmosis and Lyme disease at Camp Ripley, Minnesota. *Vector Borne Zoonotic. Dis.* **11**, 1529–1534. (doi:10.1089/vbz.2011.0633)

22. Hamer SA, Lehrer E, Magle SB. 2012 Wild birds as sentinels for multiple zoonotic pathogens along an urban to rural gradient in greater Chicago, Illinois. *Zoonoses Public Health* **59**, 355–364. (doi:10.1111/j.1863-2378.2012.01462.x)

23. Pepin KM *et al.* 2012 Geographic variation in the relationship between human Lyme disease incidence and density of infected host-seeking *Ixodes scapularis* nymphs in the Eastern United States. *Am. J. Trop. Med. Hyg.* **86**, 1062–1071. (doi:10.4269/ajtmh.2012.11-0630)

24. Hanincova K *et al.* 2013 Multilocus sequence typing of *Borrelia burgdorferi* suggests existence of lineages with differential pathogenic properties in humans. *PLoS ONE* **8**, e73066. (doi:10.1371/journal.pone.0073066)

25. Margos G *et al.* 2013 *Borrelia bavariensis* sp. nov. is widely distributed in Europe and Asia. *Int. J. Syst. Evol. Microbiol.* **63**, 4284–4288. (doi:10.1099/ijs.0.052001-0)

26. Guerra M *et al.* 2002 Predicting the risk of Lyme disease: habitat suitability for *Ixodes scapularis* in the north central United States. *Emerg. Infect. Dis.* **8**, 289. (doi:10.3201/eid0803.010166)

27. Johnson TL *et al.* 2018 Prevalence and distribution of seven human pathogens in host-seeking *Ixodes scapularis* (Acari: Ixodidae) nymphs in Minnesota, USA. *Ticks Tick Borne Dis.* **9**, 1499–1507. (doi:10.1016/j.ttbdis.2018.07.009)

28. Scrimenti RJ. 1970 Erythema chronicum migrans. *Arch. Dermatol.* **102**, 104–105. (doi:10.1001/archderm.1970.04000070106017)

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

30. Peterson AT, Raghavan R. 2017 The leading edge of the geographic distribution of *Ixodes scapularis* (Acari: Ixodidae). *J. Med. Entomol.* **54**, 1103. (doi:10.1093/jme/tjx097)

31. Rydzewski J, Mateus-Pinilla N, Warner RE, Hamer S, Weng H-Y. 2011 *Ixodes scapularis* and *Borrelia burgdorferi* among diverse habitats within a natural area in east-central Illinois. *Vector Borne Zoonotic Dis.* **11**, 1351–1358. (doi:10.1089/vbz.2010.0160)

32. Russart NM, Dougherty MW, Vaughan JA. 2014 Survey of ticks (Acari: Ixodidae) and tick-borne pathogens in North Dakota. *J. Med. Entomol.* **51**, 1087–1090. (doi:10.1603/ME14053)

33. Beckmann S, Freund R, Pehl H, Rodgers A, Venegas T. 2019 Rodent species as possible reservoirs of *Borrelia burgdorferi* in a prairie ecosystem. *Ticks Tick Borne Dis.* **10**, 1162–1167. (doi:10.1016/j.ttbdis.2019.06.011)

34. Stone BL, Tourand Y, Brissette CA. 2017. Brave new worlds: the expanding universe of Lyme Disease. *Vector Borne Zoonotic Dis.* **17**, 619–629. (doi:10.1089/vbz.2017.2127)

35. Kilpatrick AM *et al.* 2017 Lyme disease ecology in a changing world: consensus, uncertainty and critical gaps for improving control. *Phil. Trans. R. Soc. B* **372**, 20160117. (doi:10.1098/rstb.2016.0117)

36. Barbour AG, Fish D. 1993 The biological and social phenomenon of Lyme disease. *Science* **260**, 1610–1616. (doi:10.1126/science.8503006)

37. Diuk-Wasser MA *et al.* 2006 Spatiotemporal patterns of host-seeking *Ixodes scapularis* nymphs (Acari: Ixodidae) in the United States. *J. Med. Entomol.* **43**, 11. (doi:10.1603/0022-2585(2006)043[0166:SPOHIS]2.0.CO;2)

38. Homer C *et al.* 2015 Completion of the 2011 National land cover database for the conterminous United States—representing a decade of land cover change information. *Photogramm. Eng. Remote Sensing* **81**, 345–354.

39. Strahler AN. 1957 Quantitative analysis of watershed geomorphology. *Eos* **38**, 913–920. (doi:10.1029/TR038i006p00913)

40. Hales S, de Wet N, Maindonald J, Woodward A. 2002. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *Lancet* **360**, 830–834. (doi:10.1016/S0140-6736(02)09964-6)

41. Leighton PA, Koffi JK, Pelcat Y, Lindsay LR, Ogden NH. 2012. Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector *Ixodes scapularis* in Canada. *J. Appl. Ecol.* **49**, 457–464. (doi:10.1111/j.1365-2664.2012.02112.x)

42. Hyde J. 1977 Testing survival under right censoring and left truncation. *Biometrika* **64**, 225–230. (doi:10.1093/biomet/64.2.225)

43. Dennis DT, Nekomoto TS, Victor JC, Paul WS, Piesman J. 1998. Forum: reported distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the United States. *J. Med. Entomol.* **35**, 629–638. (doi:10.1093/jmedent/35.5.629)

44. Fawcett T. 2006. An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874. (doi:10.1016/j.patrec.2005.10.010)

45. Maestas LP, Adams SL, Britten HB. 2016 First evidence of an established population of *Ixodes scapularis* (Acari: Ixodidae) in South Dakota. *J. Med. Entomol.* **53**, 965–966. (doi:10.1093/jme/tjw038)

46. Eisen RJ, Eisen L. 2018 The blacklegged tick, *Ixodes scapularis*: an increasing public health concern. *Trends Parasitol.* **34**, 295–309. (doi:10.1016/j.pt.2017.12.006)

47. Diuk-Wasser MA *et al.* 2010 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. *Global Ecol. Biogeogr.* **19**, 504–514. (doi:10.1111/j.1466-8238.2010.00526.x)

48. Ostfeld RS, Cepeda OM, Hazler KR, Miller MC. 1995 Ecology of Lyme disease: habitat associations of ticks (*Ixodes scapularis*) in a rural landscape. *Ecol. Appl.* **5**, 353–361. (doi:10.2307/1942027)

49. Cortinas R, Spomer S. 2013 Lone star tick (Acari: Ixodidae) occurrence in Nebraska: historical and current perspectives. *J. Med. Entomol.* **50**, 244–251. (doi:10.1603/ME12207)

50. Ginsberg HS *et al.* 2004 Woodland type and spatial distribution of nymphal *Ixodes scapularis* (Acari: Ixodidae). *Environ. Entomol.* **33**, 1266–1273. (doi:10.1603/0046-225X-33.5.1266)

51. Brownstein JS, Holford TR, Fish D. 2003 A climate-based model predicts the spatial distribution of the Lyme disease vector *Ixodes scapularis* in the United States. *Environ. Health Perspect.* **111**, 1157. (doi:10.1289/ehp.6052)

52. Schulze TL, Jordan RA, Hung RW. 2001 Effects of selected meteorological factors on diurnal questing

of *Ixodes scapularis* and *Amblyomma americanum* (Acari: Ixodidae). *J. Med. Entomol.* **38**, 318–324. (doi:10.1603/0022-2585-38.2.318)

53. Ogden NH *et al.* 2004 Investigation of relationships between temperature and developmental rates of tick *Ixodes scapularis* (Acari: Ixodidae) in the laboratory and field. *J. Med. Entomol.* **41**, 622–633. (doi:10.1603/0022-2585-41.4.622)

54. Gray JS, Dautel H, Estrada-Peña A, Kahl O, Lindgren E. 2009 Effects of climate change on ticks and tick-borne diseases in Europe. *Interdiscip. Perspect. Infect. Dis.* **2009**, e593232. (doi:10.1155/2009/593232)

55. Eisen RJ, Eisen L, Ogden NH, Beard CB. 2016 Linkages of weather and climate with *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae), enzootic transmission of *Borrelia burgdorferi*, and Lyme Disease in North America. *J. Med. Entomol.* **53**, 250–261. (doi:10.1093/jme/tjv199)

56. Ogden NH, Lindsay LR. 2016 Effects of climate and climate change on vectors and vector-borne diseases: ticks are different. *Trends Parasitol.* **32**, 646–656. (doi:10.1016/j.pt.2016.04.015)

57. Ogden NH *et al.* 2008 Risk maps for range expansion of the Lyme disease vector, *Ixodes scapularis*, in Canada now and with climate change. *Int. J. Health Geogr.* **7**, 24. (doi:10.1186/1476-072X-7-24)

58. Sonenshine DE. 2018 Range expansion of tick disease vectors in North America: implications for spread of tick-borne disease. *Int. J. Environ. Res. Public Health* **15**, 478. (doi:10.3390/ijerph15030478)

59. Linske MA *et al.* 2019 Impacts of deciduous leaf litter and snow presence on nymphal *Ixodes scapularis* (Acari: Ixodidae) overwintering survival in coastal New England, USA. *Insects* **10**, 227. (doi:10.3390/insects10080227)

60. Robinson WS. 1950 Ecological correlations and the behavior of individuals. *Am. Sociol. Rev.* **15**, 351–357. (doi:10.2307/2087176)

61. Mader EM *et al*. In press. A survey of tick surveillance and control practices in the United States. *J. Med. Entomol.*

62. Gilliam B *et al.* 2020 Impact of unexplored data sources on the historical distribution of three vector tick species in illinois. *J. Med. Entomol.* **57**, 872–883. (doi:10.1093/jme/tjz235)

63. Eisen RJ, Paddock CD. In press. Tick and tickborne pathogen surveillance as a public health tool in the United States. *J. Med. Entomol.*

64. Fischhoff IR *et al.* 2019 Assessing effectiveness of recommended residential yard management measures against ticks. *J. Med. Entomol.* **56**, 1420–1427. (doi:10.1093/jme/tjz077)