

Multi-Scale Analysis of Lyme Disease Ecology

Rebecca Michelle Bingham-Byrne * and Esra Ozdenerol *

Spatial Analysis and Geographic Education Laboratory, Department of Earth Sciences, University of Memphis, Memphis, TN 38152, USA

* Correspondence: rmbingham@memphis.edu (R.M.B.-B.); eozdenrl@memphis.edu (E.O.)

Abstract: Lyme disease is a zoonotic infectious disease. Increased public interest in Lyme disease has caused increased efforts by researchers for its surveillance and control. The main concept for this paper is to determine the mammalian species composition of areas at high risk for Lyme disease utilizing GIS-based (Geographic Information Systems) techniques coupled with k-means clustering, random forest, and multinomial logistic regression. Cluster analysis results were similar to previous work involving maps that display areas where people are at high risk for developing Lyme disease. There were differences in which mammal species presence had associations with Lyme disease risk observed at the two different scales within this analysis, with some overlap observed between the national scale and the smaller regions, as well as some overlap between the Rocky Mountain and Southeast regions that was not found at the national scale. This is an investigative analysis to determine which species are needed for habitat suitability analyses in efforts to prioritize vaccine deployment locations. There has been limited research on vaccine deployment for Lyme disease. Increasing our understanding of not only the vaccine but also the interactions between the components of disease transmission is necessary to control this infectious disease successfully.

Keywords: Lyme disease; risk modeling; species composition; vaccine deployment; wildlife vaccination

Citation: Bingham-Byrne, R.M.; Ozdenerol, E. Multi-Scale Analysis of Lyme Disease Ecology. *Rheumato* **2024**, *4*, 88–130. <https://doi.org/10.3390/rheumato4020008>

Academic Editors: Bruce M. Rothschild and Hajime Kono

Received: 31 December 2023

Revised: 10 April 2024

Accepted: 23 April 2024

Published: 6 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Lyme disease most commonly occurs in the upper Midwest and Northeastern United States, but cases are starting to emerge in other areas, including California, contracted via *Ixodes pacificus* [1]. Humans develop Lyme disease after being bitten by a tick, usually *Ixodes scapularis*, that is infected with one of the many bacterial species that cause Lyme disease, such as those from the genus *Borrelia* [2]. This tick must have been attached for 36–48 hours or more. Neither tick density nor infection rates of this bacterium influence the survival of the rodent host [3]. However, this bacterium does influence human health. Currently, no vaccine is available for humans for this disease [4], but there is one monoclonal antibody and one vaccine in human clinical trials [5,6]. However, there are oral vaccine baits available for wildlife [7], although not yet for Lyme borreliosis. Therefore, it is important to develop methods to prioritize areas for vaccine deployment to improve overall public health.

Controlling Lyme disease by reducing black-legged ticks (*I. scapularis*) and their hosts has proven difficult [8] as well as inadequate and inefficient in reducing disease risk [9]. However, the use of fipronil-based bait boxes has reduced the abundance of questing nymphs and the potential of encountering *Borrelia* spp.-infected nymphs [10]. Recently, it was shown in a very large trial that reducing *I. scapularis* in backyards did not affect the human incidence of Lyme disease [11]. Even if alternative methods such as vaccination, medication, and contraception of wildlife populations prove successful in Lyme disease prevention, there will always be a need to evaluate the risk of these diseases through general or targeted surveillance [9].

To determine the best way of conducting vaccine deployment, researchers need to understand the dynamics of not only the vaccine but also the target species and their natural habitat [9]. There are few efforts related to Lyme disease vaccine deployment. Ozdenrol [12] mentioned that spatial analysis is imperative to understand how to map Lyme disease risk. These risk maps can be used to determine cost-effective areas for efficient vaccine deployment. Researchers have argued the necessity of developing models to predict areas that are suitable for Lyme disease risk by finding environments suitable for the tick species as well as their hosts [13], such as the white-footed mouse (*Peromyscus leucopus*) and white-tailed deer (*Odocoileus virginianus*), which can be done via a combination of remote sensing and GIS tools. However, this method may not work for every area. While there is some knowledge of *P. leucopus*, *O. virginianus*, and *I. scapularis*, less is known or reported about the species that *I. scapularis* feeds on. Other species may also play an important role in the transmission of Lyme disease to humans. It is imperative to determine what species are associated with a high incidence of Lyme disease in humans. Therefore, the purpose of this study is to determine the species composition in areas of high Lyme disease incidence. This is a preliminary analysis in developing Lyme disease risk maps, which help determine where to prioritize Lyme disease prevention efforts.

2. Materials and Methods

2.1. Data Collection

2.1.1. Lyme Disease Incidence Rates

The number of new cases of Lyme disease was downloaded for each county for each year between 2000 and 2017 from the Centers for Disease Prevention website (<https://www.cdc.gov/lyme/stats/survfaq.html>, accessed on 10 March 2019). Population estimates for each county were downloaded from the Census Bureau website in two different datasets: the "Intercensal Estimates of Resident Population for Counties and States: 1 April 2000 to 1 July 2010" and the "County Population Totals and Components of Change: 2010–2017" (<https://www.census.gov/programs-surveys/popest/data/datasets.All.html>, accessed on 10 March 2019). The incidence rates per 100,000 people were then calculated for each county in Microsoft Excel (Microsoft Office) using Equation (1) below:

$$\frac{\text{new case counts per county}}{\frac{\text{population per county}}{100,000}}. \quad (1)$$

These data were uploaded into ArcMap 10.6 (Esri, INC) along with a United States county boundary layer, where the data were joined to the county layer to provide a Lyme disease incidence by county GIS shapefile (Figure 1). Because this project covers multiple scales (national and regional), a new variable was created to include the region of the United States in which each county is located.

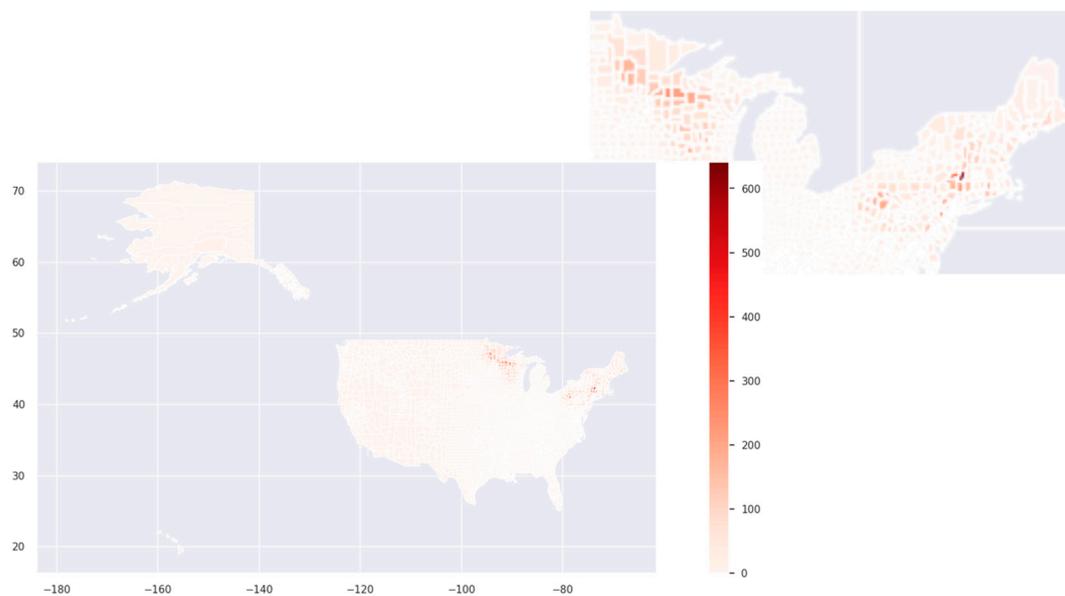


Figure 1. Lyme disease incidence rates on a national scale averaged over an 18-year period. Lighter colors show no incidence and darker colors show high incidence.

2.1.2. Species Presence

The available species ranges for each mammal included in the National Gap Analysis Project (GAP) website (<https://gapanalysis.usgs.gov/species/data/download/>, accessed on 10 March 2019) was downloaded. This included 459 total mammal species distributions for the analysis. While this does not contain all mammals present in the United States, all wild mammal species currently known to be associated with the presence of Lyme disease reported in previous literature were obtainable from this database. These files were extracted from their folders using a Python script for improved efficiency. The range for each species was uploaded into ArcMap 10.6 (Esri, INC), and then the spatial join tool was used to join the original county boundary layer to each species range. This gave a summary of the number of polygons in each species range that intersects each county. These data were exported to a separate text file for each species. Another Python script was used to open these text files in Microsoft Excel and recode intersection counts greater than 0 to "presence" to represent that species was present in that county, and everything that was 0 was set to "absence" to represent that species was absent in that county. Species with either no presence or no absence in a county throughout the area the analysis was interested in were omitted from those analyses.

2.1.3. Descriptive Statistics and Data Management

The Microsoft Excel file, along with GIS shapefiles containing Lyme disease incidence rates, were uploaded to Google Drive, where they could be used for further manipulations, analyses, and visual representations in Google Colab. Several packages were used, including GeoPandas, mapclassify, pymssql, and Yellowbrick, along with other packages that did not need to be separately installed onto the platform. Data were loaded onto the platform, and descriptive statistics were used on Lyme disease incidence data to better understand the structure of the data and the geographical spread of Lyme disease. GeoPandas was used to import the shapefile. Descriptive information such as count, mean, standard deviation, and quantiles were calculated and reviewed.

2.2. Analyses

2.2.1. K-Means Clustering

A k-means clustering analysis was used to cluster counties based on the consistency of yearly incidence rates for the years 2000–2017. The data were subset into three

categories: no_incidence included counties with no occurrences of Lyme disease throughout the timeframe of this dataset; data3 (counties with very low incidence); and high_outliers (counties with high rates of Lyme disease 3 standard deviations or larger above the mean). The k-means cluster analysis was conducted on the high_outliers dataset as it includes the areas of most interest. Before clustering, pairwise regressions were conducted between each year to assess the correlation among years. An elbow plot was created with the KElbowVisualizer from the Yellowbrick.cluster Python package and used to determine the number of clusters of consistent risk levels ranging from 2 (lowest risk with lowest mean average incidence) to k (highest risk with highest mean average incidence). Once the analysis was finished, no_risk was assigned a cluster rank of 0 ("no risk"), and data3 was assigned the rank 1 (very low-risk cluster). The three subsets of data were merged back together into a single dataset. Descriptive details such as count, mean, standard deviation, and quantiles were calculated, and a frequency histogram was created to visualize the spread of counties into each cluster.

2.2.2. Random Forest Model

To determine which species may be important classifiers for Lyme disease risk, all species presence/absence data that had varying levels of presence throughout the area of analysis were included in the random forest model to determine variable importance. The data were separated into training and test data. The training data included 80 percent of the observations in each cluster. The test data were the remaining observations in the dataset. The training and test datasets remained the same throughout the three different predicting models. Clusters were used as the label to be predicted, and species presence was used as the exploratory variable. A 5-fold cross-validation score was calculated for both the training and test data and was used to determine if the model was successful in predicting Lyme disease risk. A variable importance bar chart was used to visualize species' presence/absence contribution to the model, and the elbow rule was used to choose which species would be included in the multinomial logistic regression analysis.

2.2.3. Multinomial Logistic Regression Analysis

A multinomial logistic regression analysis was used because the response (risk clusters) was non-binary. The risk clusters obtained from the k-means cluster analysis were used as separate classes for the outcome variables to determine the impact of species presence or species absence in areas of various Lyme disease risk levels. The model was set to be a multi_class multinomial with the "lbfgs" (limited-memory Broyden–Fletcher Goldfarb–Shanno) optimization technique so that the model could handle the multinomial loss. The penalty was also set to "l2" because the optimization technique chosen can only be done with this type of penalty. After fitting the model with the training data, the test data were used to assess the model's balanced training and balanced test accuracy. All estimates whose absolute value is greater than 0.100 were recorded to be used in future analyses. Finally, counties that were misclassified and classified correctly from the test data were extracted to visualize inaccuracies in the model to obtain a better idea of how to proceed in the future.

3. Results

3.1. National Analysis

3.1.1. Descriptive Statistics

Appendix A includes all the figures and tables associated with the descriptive statistics for this study. Incidence rates seemed consistent across the 18-year period of the project, with areas that had higher rates of Lyme disease in one year also having higher rates in other years. Furthermore, the data are less correlated the further apart the years are from one another. For the k-means clustering analysis, clusters were determined based on

consistency, i.e., consistently high incidence rates or consistently low incidence rates. Therefore, clustering was conducted based on the incidence rates of all years (2000–2017).

Most counties in the United States had no or low incidence rates, providing a very skewed dataset (see Appendix A). Therefore, the data must be subset before the cluster analysis to account for the abundance of areas with no incidence rates throughout the timeframe. After removing counties with absolutely no incidence (no_incidence) or very low incidence (data3) throughout the timeframe, the counties in areas more prone to Lyme disease (high_outliers) were still highly skewed toward a lower incidence rate.

3.1.2. K-Means Clustering

The Distortion Score Elbow plot (Figure 2) shows that the number of clusters best suited for the data gathered is 4. The k-means analysis was conducted with $k = 4$, resulting in varying frequencies in each group (0: 293, mean average incidence = 31.76; 1: 86, mean average incidence = 103.67; 2: 18, mean average incidence = 230.69; 3: 1, average incidence = 641.17). It took less than a second to finish. These cluster groups were related back to average incidence to determine Lyme disease risk levels for each cluster (0: low or recoded to risk cluster 2, 1: medium or recoded to risk cluster 3, 2: high or recoded to risk cluster 4, 3: very high or recoded to risk cluster 5).

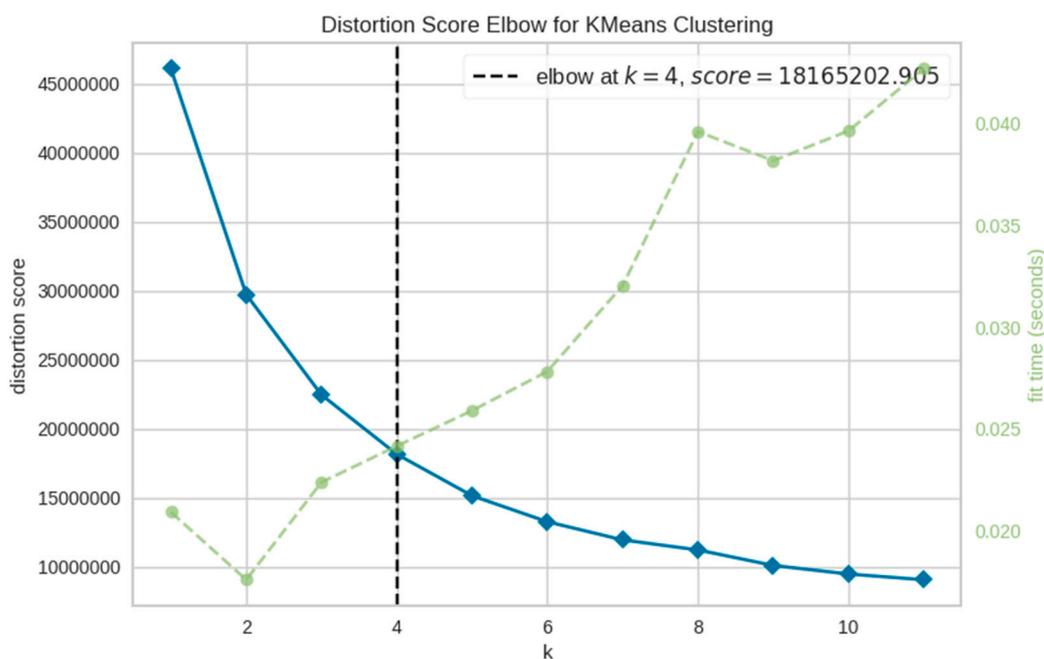


Figure 2. Distortion Score Elbow plot for Lyme disease incidence rates (years 2000–2017). The blue line refers to the distortion scores per k groups. The green dashed line refers to the time it took to train the clustering model per k groups.

The previously removed counties of no incidence (mean average incidence = 0, no risk coded to rank cluster 0) and very low risk (mean average incidence = 1.21, high risk coded to rank cluster 2) were appended back into the dataset after their clusters were manually assigned. The counts for the clustering analysis show that the counties are highly skewed toward no and lower risk categories (Figure 3), with 907 (28.88%) counties in cluster 0 (no risk), 1836 (58.45%) counties in cluster 1 (very low risk), 293 (9.33%) counties in cluster 2 (low risk), and 86 (2.74%) counties in cluster 3 (medium risk), 18 (0.57%) counties in cluster 4 (high risk), and 1 (<0.1%) county was in cluster 5 (very high risk).

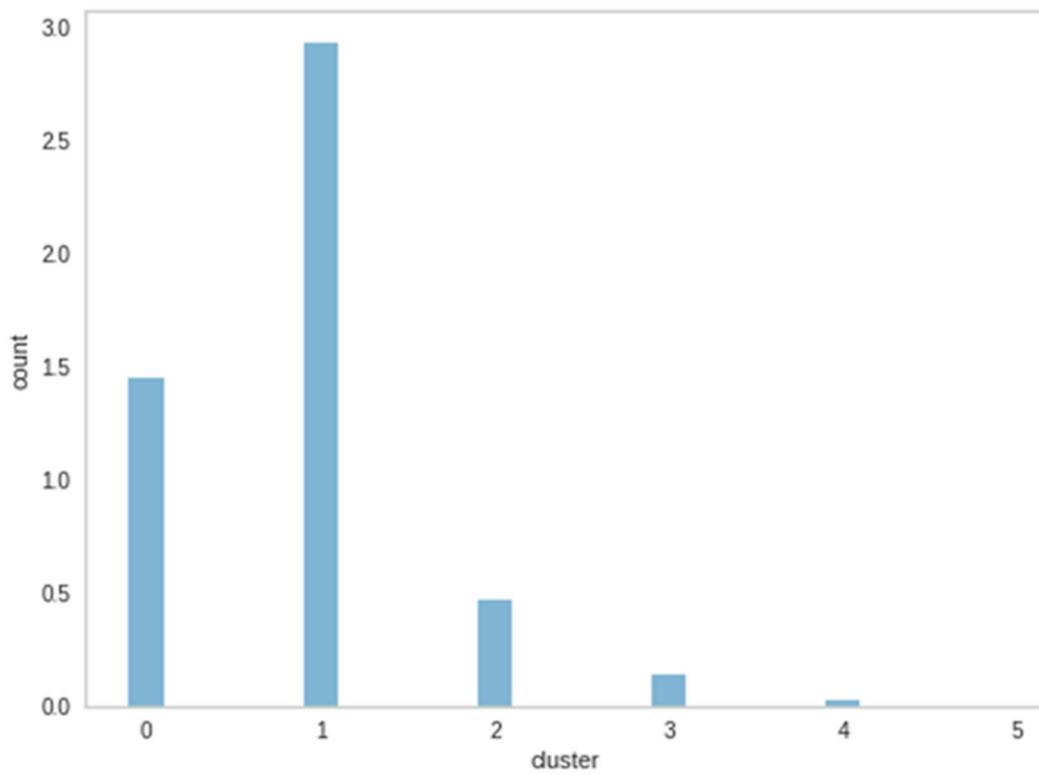


Figure 3. Lyme Disease risk cluster histogram.

These clusters were grouped spatially throughout the nation in similar ways as observed with known Lyme disease cases. Most areas of high risk were found in the Upper Midwest and Northeast regions of the United States (Figure 4).

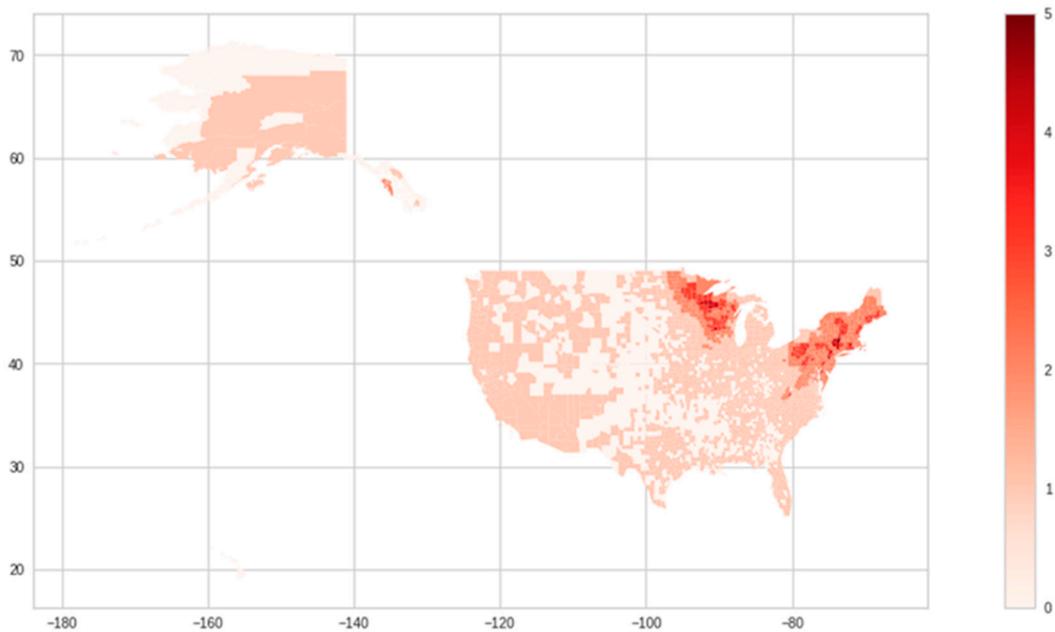


Figure 4. Spatial distribution of k-means clustering. Lighter colors show no risk, and darker colors show high risk.

3.1.3. Random Forest

A total of 453 species were used to train the random forest model. The mean cross-validation score was 70.19% in the training dataset. For the test dataset, the mean cross-validation score was 63.99%. Appendix B includes the variable importance graphs from all random forest models produced in this paper. Based on the variable importance graph, four species were kept. Most of these species were small terrestrial mammals, *Myodes gapperi* (red-backed vole), *Condylura cristata* (star-nosed mole), and *Sorex palustris* (American water shrew). One medium-sized mammal also showed importance to the prediction of Lyme disease risk clusters from this analysis, *Vulpes velox* (swift fox). The other mammal species did not seem to have as much importance in the overall predictive ability of the model. However, it is known from previous research that *Peromyscus leucopus* (white-footed mouse) and *Odocoileus virginianus* (white-tailed deer) are very important in the spread of Lyme disease. Therefore, these two species were also retained for the final model.

3.1.4. Multinomial Regression Analysis

Appendix C contains the heatmaps to determine multicollinearities for all the pairwise comparisons conducted prior to choosing variables for the multinomial regression models. For the national analysis, no further species were excluded because there were no violations in multicollinearity assumptions ($r > 0.8$).

Vulpes velox was positively associated with Lyme disease clusters within the multinomial logistic regression model, while the four small mammal species were negatively associated with Lyme disease risk clusters. Table 1 gives a detailed list of each mammal species with their correlation estimates. Appendix D gives a visual representation of all species correlation coefficients obtained from multinomial logistic regression models.

Table 1. Descriptive statistics for Lyme disease incidence by county.

Mammal Species	Estimate
<i>Myodes gapperi</i> (southern red-backed vole)	-1.462
<i>Condylura cristata</i> (start-nosed mole)	-2.116
<i>Vulpes velox</i> (swift fox)	1.766
<i>Sorex palustris</i> (American water shrew)	-1.235
<i>Peromyscus leucopus</i> (white-footed mouse)	-1.352
<i>Odocoileus virginianus</i> (white-tailed deer)	-0.305

Red filled cell means there is a negative relationship. Blue filled cell means there is a positive relationship. No fill means there is a neutral relationship.

This model had a mean cross-validation score of 65.93% for the training dataset and 62.89% for the test dataset. There were a total of 223 counties that were misclassified from the test data (Figure 5). Of the 405 counties that were correctly classified into the correct Lyme disease risk category (Figure 6), none were counties originally categorized into medium, high, or very high Lyme disease risk.

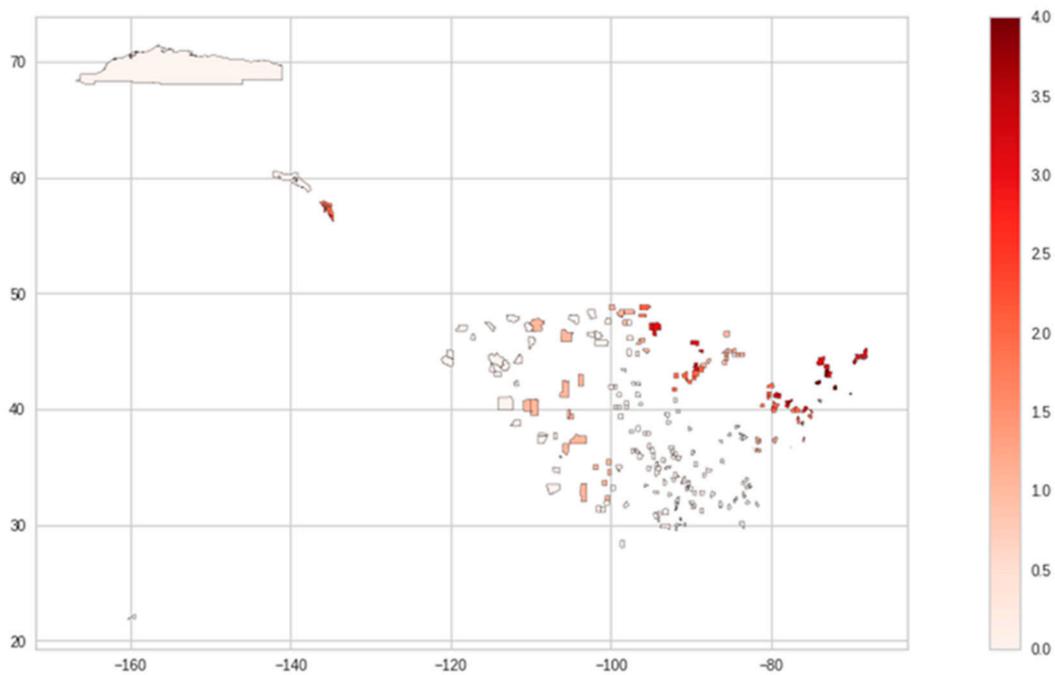


Figure 5. Counties that were misclassified from the multinomial logistic regression model. Lighter colors show no risk, and darker colors show high risk.

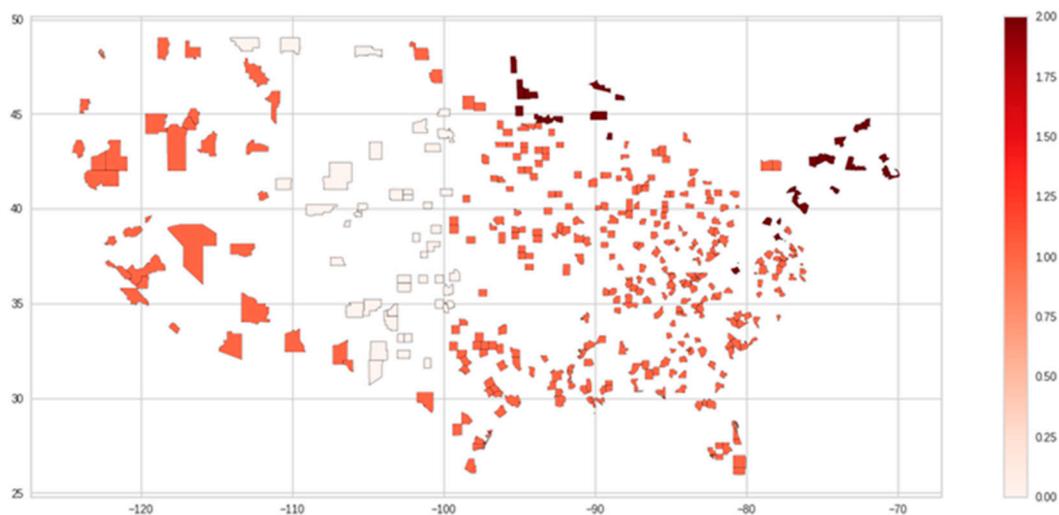


Figure 6. Counties that were correctly classified from the multinomial logistic regression model. Lighter colors show no risk, and darker colors show high risk.

3.2. Regional Analysis

Because of the regional nature of the disease, the original dataset was subset into different regions of the United States (Figure 7): Alaska, Midcontinent, Northeast, Northwest Pacific Islanders, Rocky Mountains, Southeast, and Southwest. The analyses were rerun at the regional scale. The different statistical analyses were conducted for each region in the same way to keep consistency throughout the study.

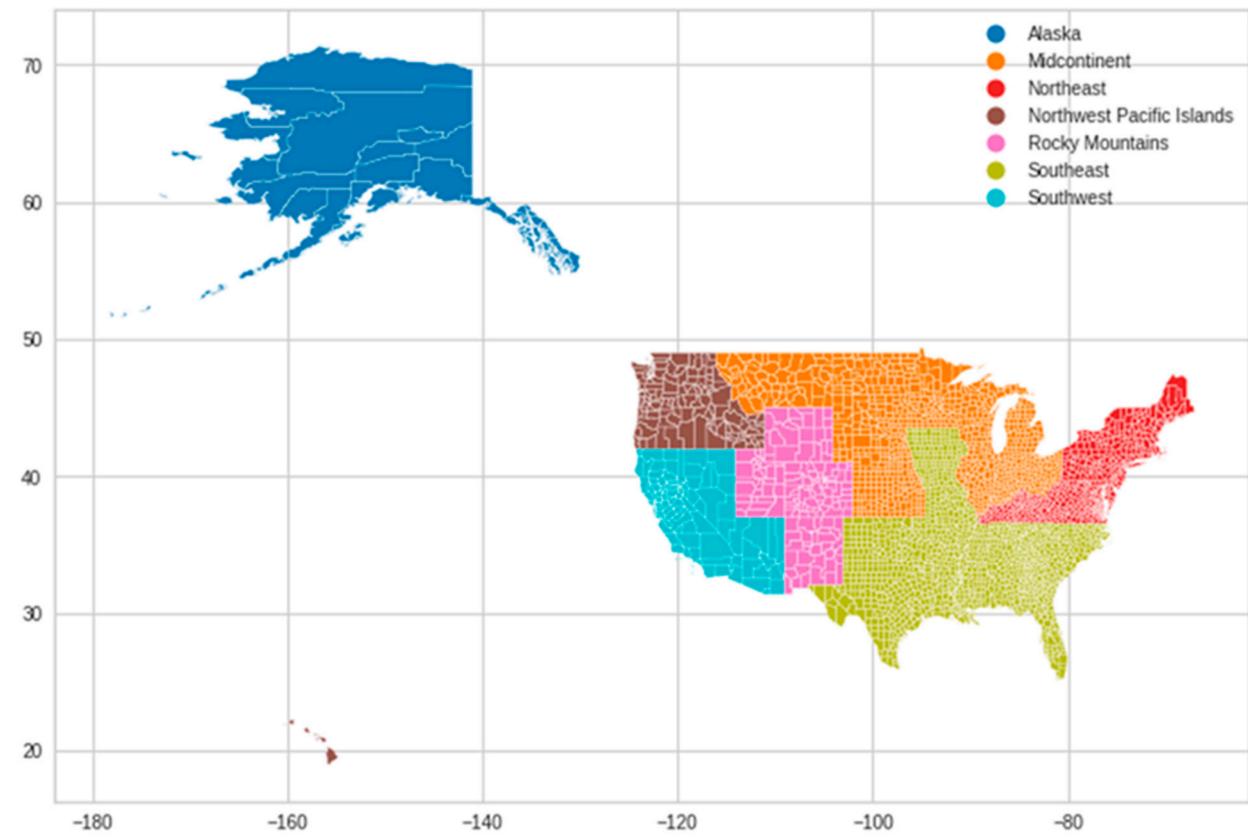


Figure 7. Map of different United States regions.

3.2.1. Descriptive Statistics

Most counties within each region of the United States also had no or low incidence rates (see Appendix A). Again, the data were subset before each cluster analysis to account for the abundance of areas with no incidence rates throughout the timeframe and any outliers of high incidence.

3.2.2. K-Means Clustering

The regional k-means clustering shows clusters of varying levels of Lyme disease risk compared to other counties within each region. These comparisons were merged into a single map of the nation (Figure 8). Areas shown in white have no Lyme disease risk, and darker areas show increasing levels of Lyme disease risk.

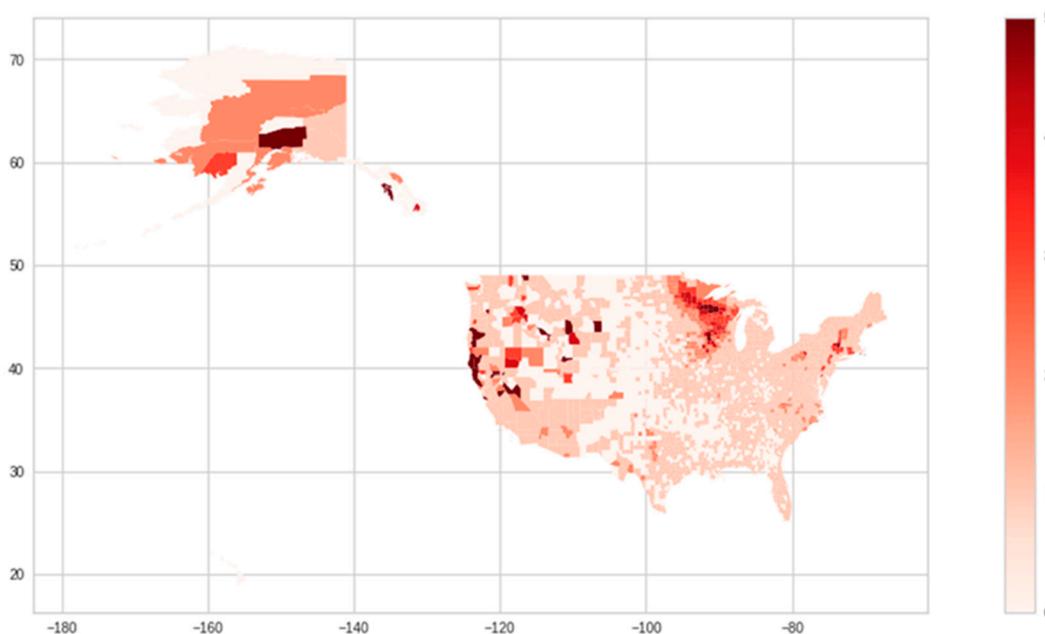


Figure 8. Map of k-means cluster groups per region for Lyme disease risk for the 18-year period. Lighter colors show no risk, and darker colors show high risk.

Table 2 shows the frequencies of counties in each category per region. For each region, most of the counties fell within the "no risk" cluster, "very low risk", or "low risk" cluster, which is consistent with results from the national scale analysis.

Table 2. Cluster frequencies for Lyme disease risk by county per region.

	Alaska	Midcontinent	Northeast	Northwest and Pacific Islands	Rocky Mountains	Southeast	Southwest
N	27	897	554	124	149	1300	90
No Risk	14 (51.85%)	244 (27.20%)	50 (9.02%)	39 (31.45%)	100 (67.11%)	451 (34.69%)	9 (10.00%)
Very Low Risk	2 (7.41%)	542 (60.42%)	470 (84.83%)	69 (55.64%)	42 (28.19%)	760 (58.46%)	54 (60.00%)
Low Risk	7 (37.03%)	52 (5.80%)	14 (2.53%)	3 (2.42%)	2 (1.34%)	80 (6.15%)	12 (13.33%)
Medium Risk	1 (3.70%)	38 (4.24%)	15 (2.71%)	6 (4.84%)	1 (0.67%)	1 (<0.01%)	3 (3.33%)
High Risk	1 (3.70%)	16 (1.78%)	4 (0.72%)	2 (1.61%)	1 (0.67%)	6 (0.46%)	1 (1.11%)
Very High Risk	2 (7.41%)	5 (0.11%)	1 (0.18%)	5 (4.03%)	3 (2.01%)	2 (0.15%)	11 (12.22%)

3.2.3. Random Forest

Model summaries—A total of 459 species distributions were downloaded from the USGS GAP website. Species whose presence varied across a specific region were included within that analysis, otherwise, the species was excluded from that specific model. A detailed summary of the random forest models for each region, including the number of species used and the cross-validation scores for train and test data, are included in Table 3. Of the 459 species distributions collected from the USGS GAP website, 0 mammal species were present throughout the counties in the Alaska region. Therefore, no regional analysis was conducted on this region. Model success ranged from 48.42% (Northwest and Pacific Islands) to 84.885 (Northeast) for the training data and 56.00% (Northwest and Pacific Islands) to 84.70% (Northeast) for the test data.

Table 3. Random forest model summaries for Lyme disease risk by region.

Model	Number of Species Used in the Model	Cross-Validation Score (Training Data)	Cross-Validation Score (Test Data)
Alaska	0	N/A*	N/A*
Midcontinent	163	72.56%	64.86%
Northeast	88	84.88%	84.70%
Northwest and Pacific Islands	180	48.42%	56.00%
Rocky Mountains	213	66.67%	69.33%
Southeast	217	67.73%	60.59%
Southwest	279	59.90%	61.67%

*N/A stands for not applicable.

Variable importance summaries—The variable importance graphs (Appendix B) were used to obtain the initial list of species retained for further analyses. Then, pairwise comparisons were conducted to exclude any species whose presence was highly correlated ($r > 0.8$). Heatmaps were created to allow for a quick visualization check for the autocorrelation (see Appendix C). As in the national analysis, *Peromyscus leucopus* (white-footed mouse) and *Odocoileus virginianus* (white-tailed deer) were retained for the multinomial regression analysis even if they were not found to be important within the random forest model unless the species was either present throughout the entire region or absent throughout the entire region.

Six species were kept for the Midcontinent region, four were kept for the Northeast region, seventeen were kept for the Rocky Mountains region, eleven were kept for the Southeast region, and twelve were kept for the Southwest region (see Table 4). Most of these species within each region were small terrestrial mammals except for the following: one medium-sized (*Marmota monax*) and one large-sized terrestrial (*Lynx canadensis*) mammal were kept in the Midcontinent region; one small aerial (*Myotis sodalis*) mammal was kept for the Northeast region; one large terrestrial mammal (*Odocoileus virginianus*) was kept in the Northwest and Pacific Islands region; two small aerial (*Euderma maculatum* and *Lasiurus blossevillii*), two large terrestrial (*Equus caballus* and *Odocoileus virginianus*), three small aerial (*Myotis lucifugus*, *Tadarida brasiliensis*, and *Myotis septentrionalis*), and one large aquatic (*Trichechus manatus*) mammals were retained for the Southeast region; and one medium-sized (*Aplodontia rufa*) mammal was retained for the Southwest region.

Table 4. Important mammalian species for Lyme disease risk by region.

Model	Species	Mammal Size
Alaska	N/A*	N/A*
Midcontinent	<i>Marmota monax</i>	medium terrestrial
	<i>Cynomys ludovicianus</i>	small terrestrial
	<i>Lynx canadensis</i>	large terrestrial
	<i>Microtus pinetorum</i>	small terrestrial
	<i>Sorex arcticus</i>	small terrestrial
	<i>Peromyscus leucopus</i>	small terrestrial
	<i>Microtus ochrogaster</i>	small terrestrial
Northeast	<i>Condylura cristata</i>	small terrestrial
	<i>Myotis sodalis</i>	small aerial
	<i>Peromyscus leucopus</i>	small terrestrial
Northwest and Pacific Islands	<i>Sorex benderii</i>	small terrestrial
	<i>Odocoileus virginianus</i>	large terrestrial
	<i>Equus caballus</i>	large terrestrial
Rocky Mountains	<i>Sorex preblei</i>	small terrestrial

	<i>Microtus mogollonensis</i>	small terrestrial
	<i>Euderma maculatum</i>	small aerial
	<i>Zapus hudsonius</i>	small terrestrial
	<i>Neotoma mexicana</i>	small terrestrial
	<i>Urocyon v. armatus</i>	small terrestrial
	<i>Peromyscus leucopus</i>	small terrestrial
	<i>Vulpes macrotis</i>	small terrestrial
	<i>Notiosorex crawfordi</i>	small terrestrial
	<i>Xerospermophilus spilosoma</i>	small terrestrial
	<i>Dipodomys merriami</i>	small terrestrial
	<i>Lasiurus blossevillii</i>	small aerial
	<i>Tamiasciurus hudsonicus</i>	small terrestrial
	<i>Tamias rufus</i>	small terrestrial
	<i>Onychomys torridus</i>	small terrestrial
	<i>Odocoileus virginianus</i>	large terrestrial
	<i>Myotis lucifugus</i>	small aerial
	<i>Zapus hudsonius</i>	small terrestrial
	<i>Microtus pennsylvanicus</i>	small terrestrial
	<i>Blarina hylophaga</i>	small terrestrial
	<i>Trichechus manatus</i>	large aquatic
Southeast	<i>Tadarida brasiliensis</i>	small aerial
	<i>Neotoma floridana</i>	small terrestrial
	<i>Spilogale putorius</i>	small terrestrial
	<i>Perognathus flavescens</i>	small terrestrial
	<i>Myotis septentrionalis</i>	small aerial
	<i>Peromyscus leucopus</i>	small terrestrial
	<i>Rattus rattus</i>	small terrestrial
	<i>Aplodontia rufa</i>	medium terrestrial
	<i>Dipodomys heermanni</i>	small terrestrial
	<i>Microtus californicus</i>	small terrestrial
	<i>Microtus longicaudus</i>	small terrestrial
Southwest	<i>Tamias amoenus celeris</i>	small terrestrial
	<i>Dipodomys deserti</i>	small terrestrial
	<i>Sorex vagrans</i>	small terrestrial
	<i>Sylvilagus audubonii</i>	small terrestrial
	<i>Ochotona princeps</i>	small terrestrial
	<i>Neotoma macrotis</i>	small terrestrial
	<i>Peromyscus leucopus</i>	small terrestrial

*N/A stands for not applicable.

3.2.4. Multinomial Logistic Regression

Model summaries—A detailed summary of the random forest models for each region, including the number of species used and the cross-validation scores for train and test data, is included in Table 5. Model successes ranged from 54.67% (Rocky Mountains) to 84.70% (Northeast). However, no models were successful at predicting counties in high or very high risk accurately (see Table 6). Furthermore, only eight counties were accurately classified as medium Lyme disease risk, and these were all located within the Midcontinent region.

Table 5. Random forest model summaries for Lyme disease risk by region.

Model	Cross-Validation Score (Training Data)	Cross-Validation Score (Test Data)
Alaska	N/A*	N/A*
Midcontinent	72.98%	69.87%
Northeast	84.88%	84.70%
Northwest and Pacific Islands	61.68%	72.00%
Rocky Mountains	65.83%	54.67%
Southeast	66.67%	62.17%
Southwest	58.67%	56.67%

*N/A stands for not applicable.

Table 6. Frequencies of regional clusters that were misclassified and correctly classified from the multinomial logistic regression model by region.

Region	Regional Cluster	Errors	Accuracies
Alaska	N/A*	N/A*	N/A*
	No risk	17	32
	Very low risk	17	91
Midcontinent	Low risk	10	0
	Medium risk	0	8
	High risk	3	0
	Very high risk	1	0
	No risk	10	0
	Very low risk	0	94
Northeast	Low risk	3	0
	Medium risk	3	0
	High risk	1	0
	Very high risk	0	0
	No risk	7	1
	Very low risk	2	12
Northwest and Pacific Islands	Low risk	1	0
	Medium risk	1	0
	High risk	0	0
	Very high risk	1	0
	No risk	5	15
	Very low risk	6	2
Rocky Mountains	Low risk	0	0
	Medium risk	0	0
	High risk	0	0
	Very high risk	1	0
	No risk	36	54
	Very low risk	32	120
Southeast	Low risk	15	1
	Medium risk	0	0
	High risk	1	0
	Very high risk	0	0
	No risk	1	1
	Very low risk	0	11
Southwest	Low risk	2	0
	Medium risk	1	0
	High risk	0	0
	Very high risk	2	0

*N/A stands for not applicable.

Correlation estimates—Table 7 gives a detailed list of each mammal species with their correlation estimates per region. Appendix D gives a visual representation of each of them. A few species were present in multiple regional analyses, including *Peromyscus leucopus*, *Odocoileus virginianus*, and *Zapus hudsonius*. *Peromyscus leucopus* had conflicting results based on the region with a positive relationship with Lyme disease risk in the Midcontinent region, negative relationships with Lyme disease risk in the Rocky Mountains and Southeast regions, and neutral relationships with Lyme disease risk in the Northeast and Southwest regions. *Odocoileus virginianus* had a negative relationship in the Northwest and Pacific Islands region but a neutral relationship in the Rocky Mountains region. There was also a negative relationship between *Zapus hudsonius*, a small rodent, and Lyme disease risk within the Rocky Mountains and Southeast regions.

Table 7. Correlations between mammal presence and Lyme disease clusters by region. Blue indicates positive associations greater than 0.100, and red represents negative associations less than 0.100.

Model	Species	Estimate
Alaska	N/A*	N/A*
	<i>Marmota monax</i>	-1.439
Midcontinent	<i>Cynomys ludovicianus</i>	1.209
	<i>Lynx canadensis</i>	-2.323
	<i>Microtus pinetorum</i>	-1.352
	<i>Sorex arcticus</i>	-2.381
	<i>Peromyscus leucopus</i>	0.281
	<i>Microtus ochrogaster</i>	1.233
Northeast	<i>Condylura cristata</i>	-1.876
	<i>Myotis sodalis</i>	-0.131
	<i>Peromyscus leucopus</i>	0.015
Northwest and Pacific Islands	<i>Sorex benderii</i>	-1.229
	<i>Odocoileus virginianus</i>	-0.745
	<i>Equus caballus</i>	-0.332
	<i>Sorex preblei</i>	-0.695
	<i>Microtus mogollonensis</i>	-0.035
	<i>Euderma maculatum</i>	0.076
	<i>Zapus hudsonius</i>	-0.135
	<i>Neotoma mexicana</i>	0.458
	<i>Urocyon v. armatus</i>	-0.652
	<i>Peromyscus leucopus</i>	-0.408
Rocky Mountains	<i>Vulpes macrotis</i>	0.072
	<i>Notiosorex crawfordi</i>	0.591
	<i>Xerospermophilus spilosoma</i>	0.630
	<i>Dipodomys merriami</i>	-0.100
	<i>Lasiurus blossevillii</i>	0.149
	<i>Tamiasciurus hudsonicus</i>	-0.551
	<i>Tamias rufus</i>	-0.007
	<i>Onychomys torridus</i>	-0.234
	<i>Odocoileus virginianus</i>	-0.055
	<i>Myotis lucifugus</i>	0.748
<i>Zapus hudsonius</i>	-0.577	
Southeast	<i>Microtus pennsylvanicus</i>	-1.230
	<i>Blarina hylophaga</i>	0.857
	<i>Trichechus manatus</i>	-1.208
	<i>Tadarida brasiliensis</i>	1.146

Southwest	<i>Neotoma floridana</i>	0.774
	<i>Spilogale putorius</i>	-0.060
	<i>Perognathus flavescens</i>	0.311
	<i>Myotis septentrionalis</i>	-0.414
	<i>Peromyscus leucopus</i>	-0.480
	<i>Rattus rattus</i>	-0.712
	<i>Aplodontia rufa</i>	-0.632
	<i>Dipodomys heermanni</i>	-0.403
	<i>Microtus californicus</i>	-0.440
	<i>Microtus longicaudus</i>	-0.057
	<i>Tamias amoenus celeris</i>	-0.212
	<i>Dipodomys deserti</i>	-0.872
	<i>Sorex vagrans</i>	0.192
	<i>Sylvilagus audubonii</i>	0.533
	<i>Ochotona princeps</i>	0.677
	<i>Neotoma macrotis</i>	-0.386
<i>Peromyscus leucopus</i>	0.030	

*N/A stands for not applicable. Red filled cell means there is a negative relationship. Blue filled cell means there is a positive relationship. No fill means there is a neutral relationship.

4. Discussion

Cluster analyses showed that most areas at high risk were in the upper Midwest and Northeastern United States with sporadic areas of varying levels of Lyme disease risk spread throughout the rest of the United States, which is consistent with previous research [1]. Most of the counties within the United States were clustered into the low risk level followed by areas of no risk, which makes sense because Lyme disease has been known to occur in limited areas throughout the nation. These limited areas should make it easier to predict where Lyme disease would occur.

Five of the six mammal species excluded from the national multinomial logistic regression model were found on small islands off the coast of the United States but did not have counties associated with these islands. Therefore, there were no data about Lyme disease cases in these areas, if any cases exist. There have been cases of Lyme disease reported on some treeless islands off the coast of Scotland [14], as well as Fire Island [15] and Monhegan [16], which are two islands off the northeastern coast of the United States. Furthermore, in this study, there were no mammal species distributions found for the Alaskan region. However, there are 115 mammal species found in this region [17]. Future work could obtain incidence rates for these islands and the mammalian species distributions for the Alaskan region to assess whether mammal species have an impact on Lyme disease occurrence in these areas.

The overlap of species used among the different scales and regions includes six species. *Condylura cristata*, the star-nosed mole, was found to be negatively associated with Lyme disease risk both on the national scale and within the Northeast region. This is a semiaquatic mammal that prefers moist areas within mesic forests with friable soils in which it can burrow [18]. *Zapus hudsonius*, the meadow jumping mouse, also was found to have a negative relationship with Lyme disease risk, but this was only found in two regions: the Rocky Mountains and the Southeast. This burrowing species also prefers moist areas as well as oak–hickory upland forests that provide plenty of acorns and nuts for food [19]. Both species are burrowing species, but they are in areas that have moist soil. The combination of soil moisture and type in these areas may not be suitable areas for vectors of Lyme disease to flourish. Future work could include soil moisture and type as valuable variables in predicting Lyme disease risk.

Interestingly, *O. virginianus* (*O. virginianus* (white-tailed deer) and *P. leucopus* (white-footed mouse) had a negative relationship with Lyme disease risk at the national scale. *O.*

virginianus. *O. virginianus* also had negative associations with Lyme disease risk when looking at the Northwest Pacific regions, whereas *P. leucopus* (white-footed mouse) had negative relationships with Lyme disease risk throughout the Rocky Mountains and Southeast regions, while having a positive relationship throughout the Midcontinent region. Previous research focused on the presence of a few mammals, such as *P. leucopus* and *O. virginianus*, in Lyme disease risk [20]. However, the results of this study show that the presence of these mammal species alone may not be suitable for determining areas of Lyme disease risk. One reason why these species did not have the effect that was hypothesized could be that they are both generalist species that are found throughout much of the nation [21,22]. This would mean that they would be found in areas of varying levels of Lyme disease risk and are not suitable for a solid predictive model.

Most species whose presence or absence was associated with Lyme disease risk were small terrestrial mammals throughout all analyses. When thinking of the feeding habits of ticks, it makes sense. Small mammals, especially from the order Rodentia, are the main hosts for the pathogen, and ticks feed on small mammals in their nymphal stage, ingesting the bacterium to pass along to other species [23]. However, there were some medium- and large-sized terrestrial mammals as well as small aerial and a large aquatic mammal whose presence or absence was associated with Lyme disease risk clusters within this study. These species could be more related to adult tick rather than nymphal tick survival. There were three medium-sized terrestrial mammals, two large terrestrial mammals, and one large aquatic mammal whose presence influenced Lyme disease risk. On a national scale, *Vulpes velox* (swift fox) had a positive influence on Lyme disease risk. The swift fox is heavily dependent upon burrowing to create dens, usually in areas with little vegetation, low slopes, and clay-type soils found within shortgrass, midgrass, and sandhill prairies [24]. These areas are not consistent with known tick habitat preferences. Therefore, more work is needed to understand why this association could have been observed within the study. The other two medium-sized mammals as well as the larger mammal species within regional analyses, had negative relationships with Lyme disease risk. Perhaps adult ticks in these areas prefer to feed on other large- or medium-sized mammals for survival. Future work involving these species could include live-trapping and serology testing techniques to calculate pathogen infection rates, as well as comparing ectoparasite density and composition of the specimens to determine if and how these species could be inhibiting the spread of the pathogen.

As mentioned before, one or two species may not be suitable for predicting Lyme disease risk. The presence of mammal, reptile, bird, and tick species, along with environmental factors such as climate data and habitat type [9], human population estimates, and human social behaviors [4,20], may also be needed to develop an accurate model for predicting Lyme disease risk. Ticks may prefer some species in their feeding habits, which could promote (i.e., *P. leucopus*) or inhibit the spread of the bacterium to the tick. It has been reported that ticks feed from some reptiles in the south versus small mammals, which may be the cause of fewer human cases of Lyme disease in the area [25]. Furthermore, tick density and infection rates have been found to be key factors in Lyme disease risk [4,26]. Future work includes refining this model utilizing the species found to contribute to Lyme disease risk along with *O. virginianus*, *O. virginianus* and *P. leucopus*, which did not seem to have a significant effect on the model but has been shown to be related to Lyme disease risk [27], as well as reptile species [25] and birds [14,28] which have been found to impact on the spread of *Borrelia* spp.

One major limitation of this study is that it is very laborious and time-consuming. This is because of the time it takes to update and maintain risk prediction maps, including keeping up-to-date species presence data. Data collection and clean-up is very time-consuming, and the knowledge of not only statistical languages but also computer languages such as Python is necessary to reduce errors and time in gathering and working with data. Furthermore, there were a large number of criterion variables used in the analysis, many of which were probably unnecessary. Future work could include dividing the species into

the categories of competent vs. incompetent reservoirs. However, there is no current comprehensive list of competent reservoir species for each region or throughout the nation. Future work could include determining what makes a competent reservoir for the pathogen. Ticks themselves can live with many infections because of their highly efficient immune system containing akirins, antimicrobial peptides, caspases, defensins, etc., which deter pathogens from taking over their system, allowing them to thrive when infected and infect many other organisms during their lifespan [29]. Perhaps comparing immunity-related genes between species to find mammals that have considerable overlap with ticks could aid in finding competent mammalian species that have more influence on the spread of *Borrelia* spp.

The sensitivity of the model for high risk clusters was low, and this is the main concern with this project. Future research must be conducted to develop a better model for better prediction. Human Lyme disease cases may not be a reliable source for determining areas of high risk because where the person is diagnosed with the disease may not be where they encountered the tick that gave them the disease. However, some of these errors could be alleviated by including human social behavior in the model, as planned for future work.

Tickborne diseases constitute most of the vector-borne diseases in the United States, and there is a need to understand these diseases better. Lyme disease has increased over the years, and actions for prevention are necessary. Vaccine deployment, which could help reduce Lyme incidence rates in humans, has been a neglected area of research. This project was a first step in understanding how to determine areas of Lyme disease occurrence and where to administer vaccines via oral baits to wildlife. It also provides a baseline for future research. The multi-scale analysis focused on both national and regional levels, which allowed us to focus on the limited areas of the Northeast. However, further area reduction for the Midcontinent and Southwest regions could help improve our understanding of how Lyme disease spreads in these areas.

Author Contributions: Conceptualization, R.M.B.-B. and E.O.; methodology, R.M.B.-B. and E.O.; software, R.M.B.-B. and E.O.; validation, R.M.B.-B. and E.O.; formal analysis, R.M.B.-B. and E.O.; investigation, R.M.B.-B. and E.O.; resources, R.M.B.-B. and E.O.; data curation, R.M.B.-B. and E.O.; writing—original draft preparation, R.M.B.-B.; writing—review and editing, R.M.B.-B. and E.O.; visualization, R.M.B.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study did not require ethical approval due to its reliance on publicly available data.

Informed Consent Statement: This study did not require ethical approval due to its reliance on publicly available data.

Data Availability Statement: Included in the "data collection" section of the methods.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Descriptive Statistics

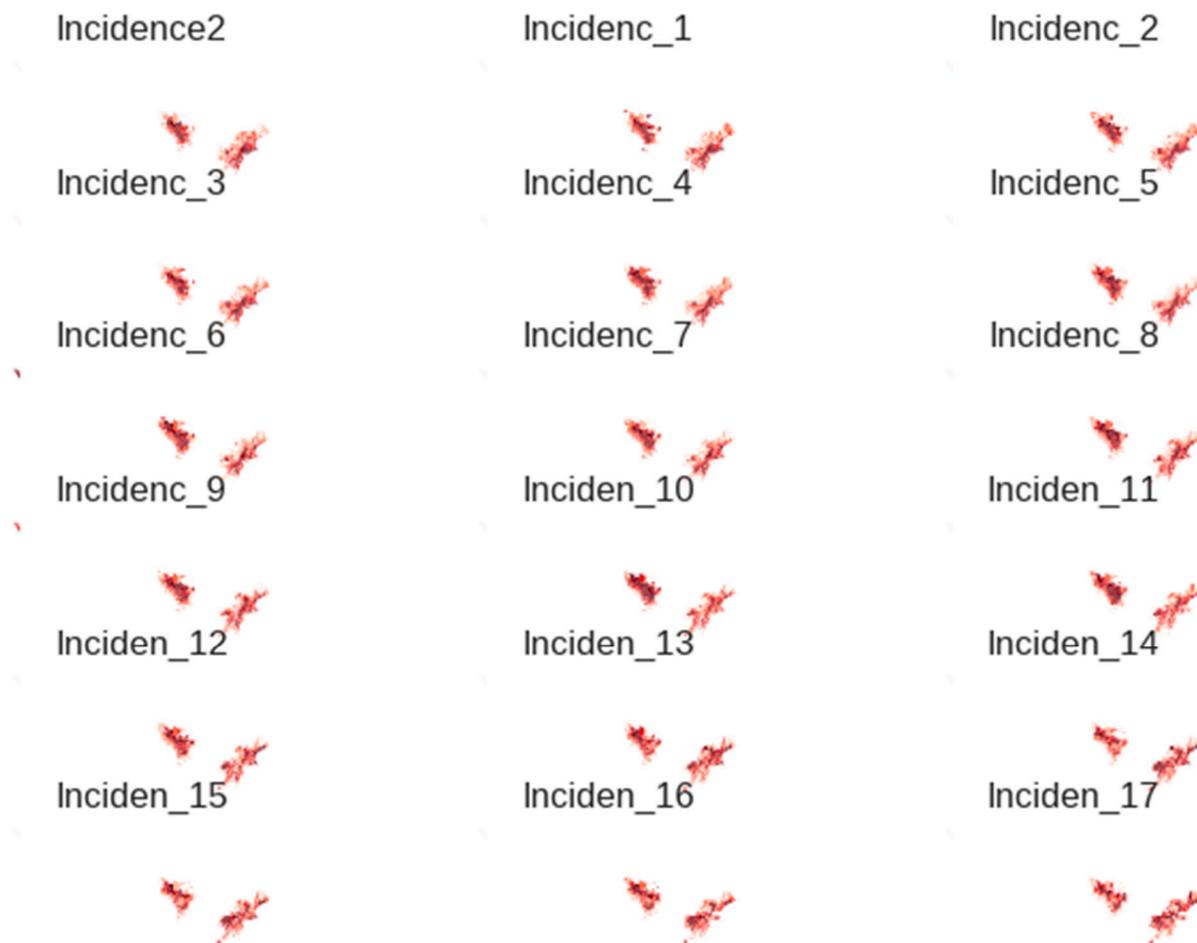


Figure A1. National Lyme disease incidence rates for years 2000 (upper left) to 2017 (bottom right). Lighter colors show no incidence and darker colors show high incidence.

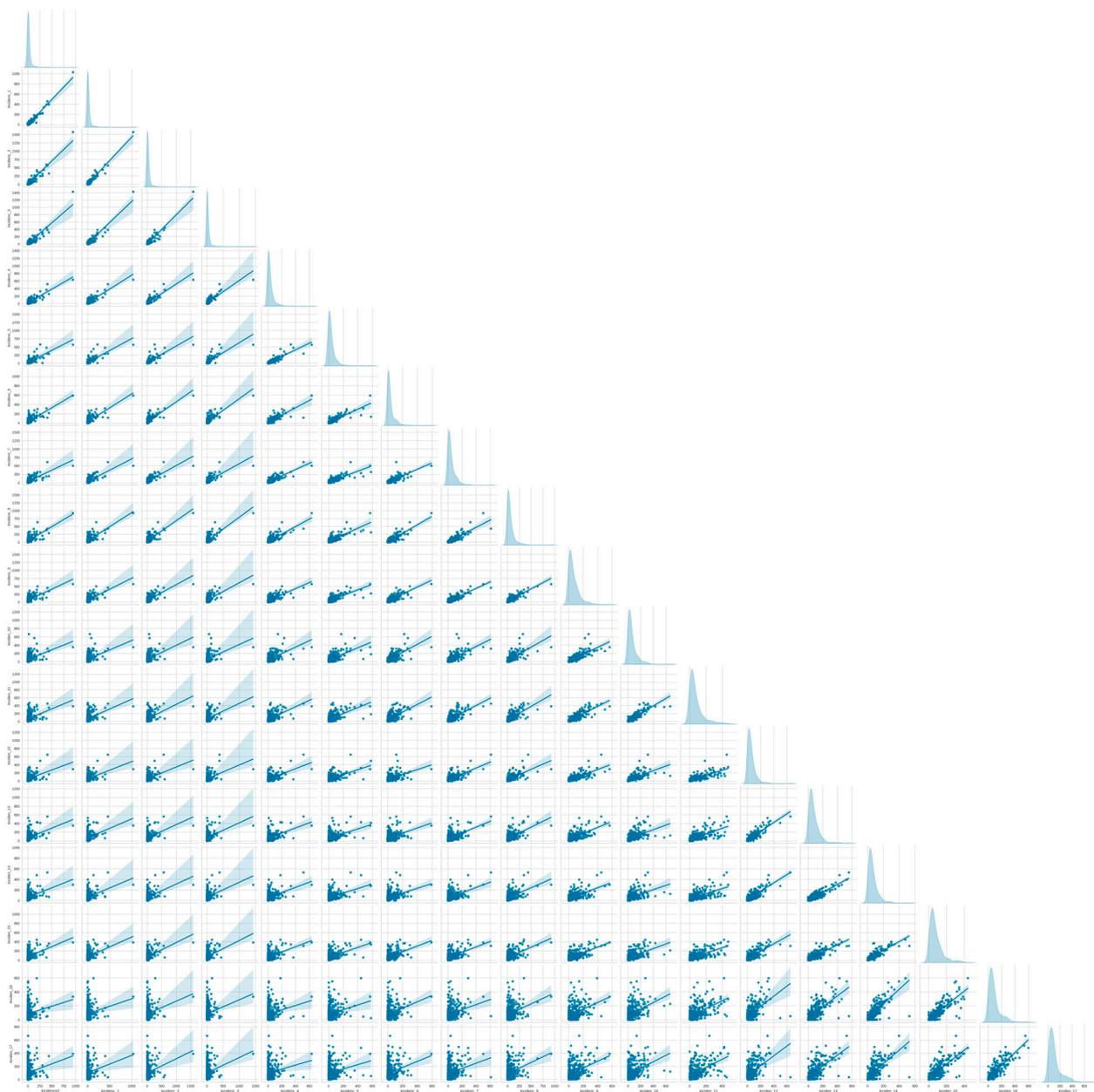


Figure A2. Pairwise annual Lyme disease incidence rates.

Table A1. Descriptive statistics for Lyme disease incidence by county for all areas, areas with very low incidence, and areas prone to Lyme disease.

Statistic	All Data	Very Low Incidence	Areas Prone to Lyme Disease
N	3141	1836	398
Mean	8.03	1.21	57.83
Median	0.42	0.61	39.04
Standard Deviation	29.07	1.55	61.82
Minimum	0.00	0.01	8.54
Maximum	641.17	8.42	641.17

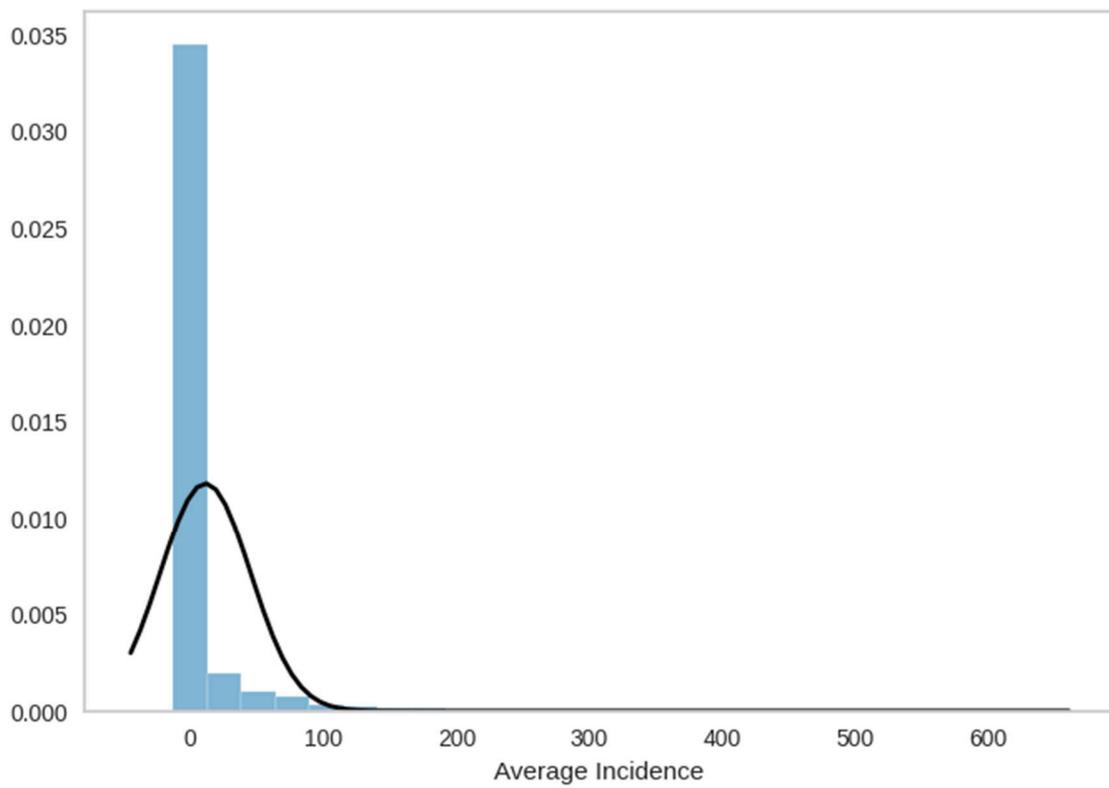


Figure A3. Frequency histogram of the average incidence of Lyme disease per county throughout the US. Blue bars indicate frequency of Lyme disease incidence rates. Black line represents the distribution curve.

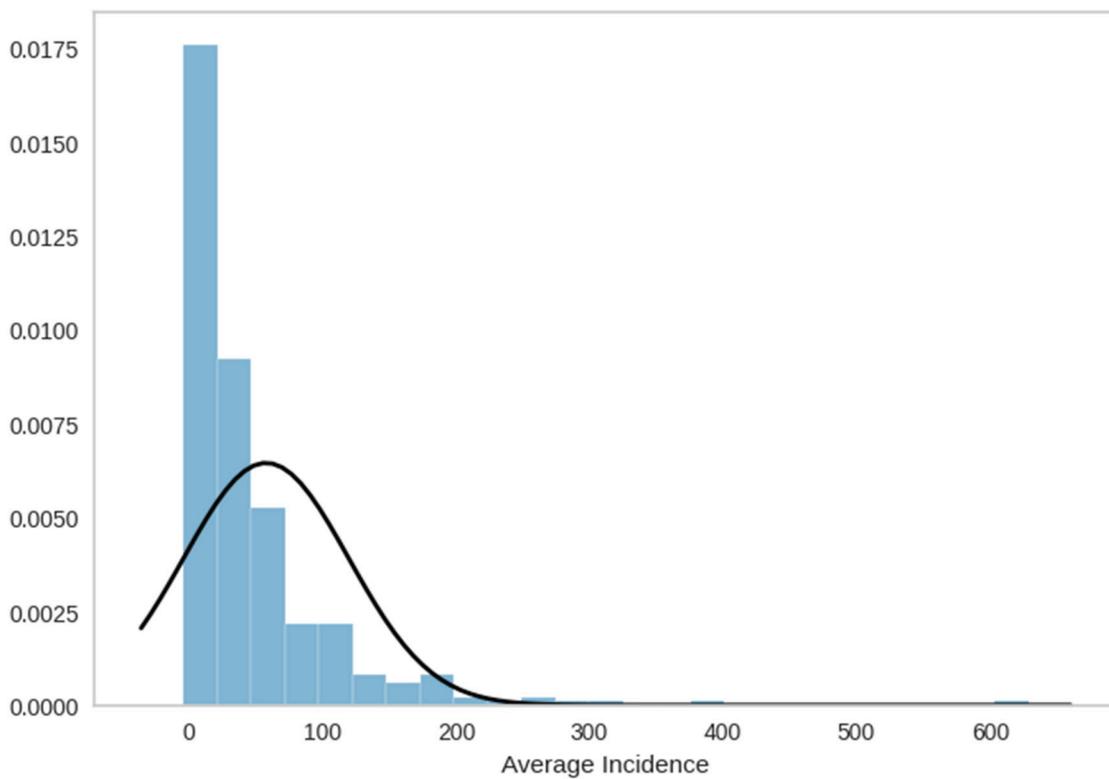


Figure A4. Frequency histogram of the average incidence of Lyme disease per county for areas prone to Lyme disease. Blue bars indicate frequency of Lyme disease incidence rates. Black line represents the distribution curve.

Table A2. Descriptive statistics for Lyme disease incidence by county per region.

	Nation	Alaska	Midcontinent	Northeast	Northwest and Pacific Islands	Rocky Mountains	Southeast	Southwest
N	3141	27	897	554	124	149	1300	90
Mean	8.03	1.15	8.99	29.05	0.54	0.12	0.69	0.74
Median	0.42	0.00	0.60	6.86	0.30	0.00	0.24	0.29
Standard Deviation	29.07	2.69	27.66	54.17	0.73	0.35	2.12	1.26
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	641.17	13.44	260.77	641.17	4.10	2.88	44.57	6.54

Appendix B

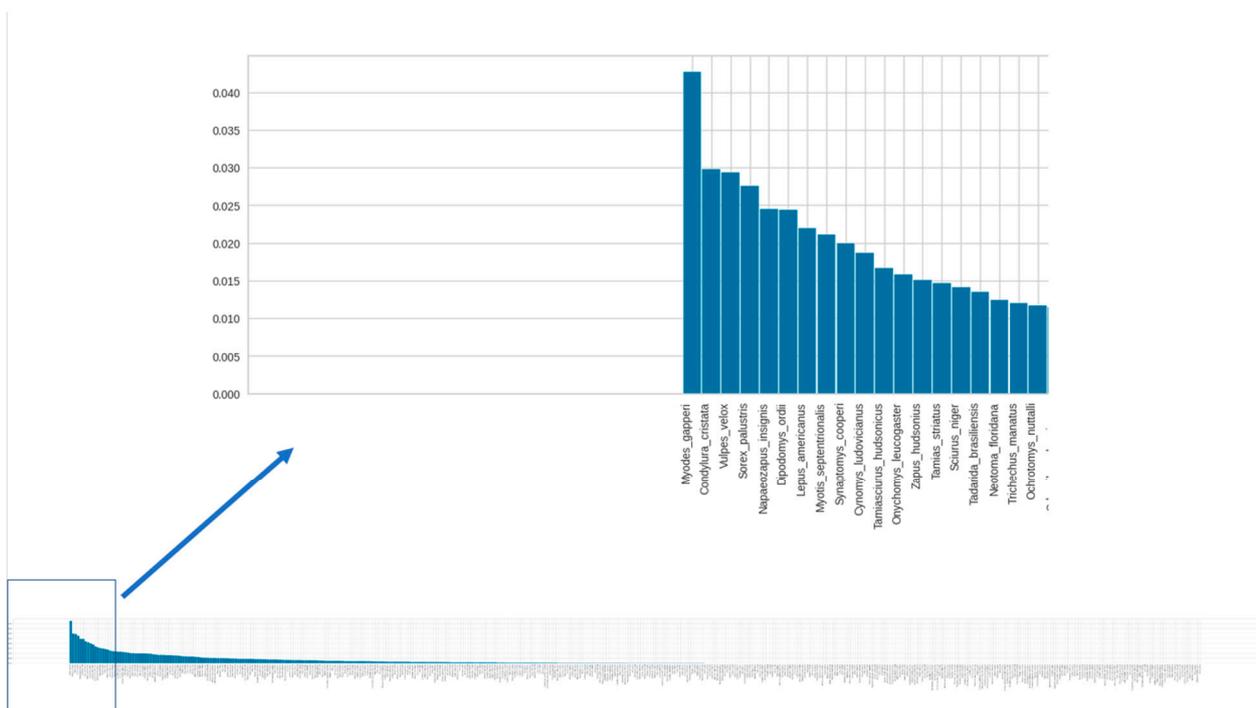


Figure A5. Histogram showing variable importance in classifying Lyme disease risk obtained from the random forest model conducted on a national scale.

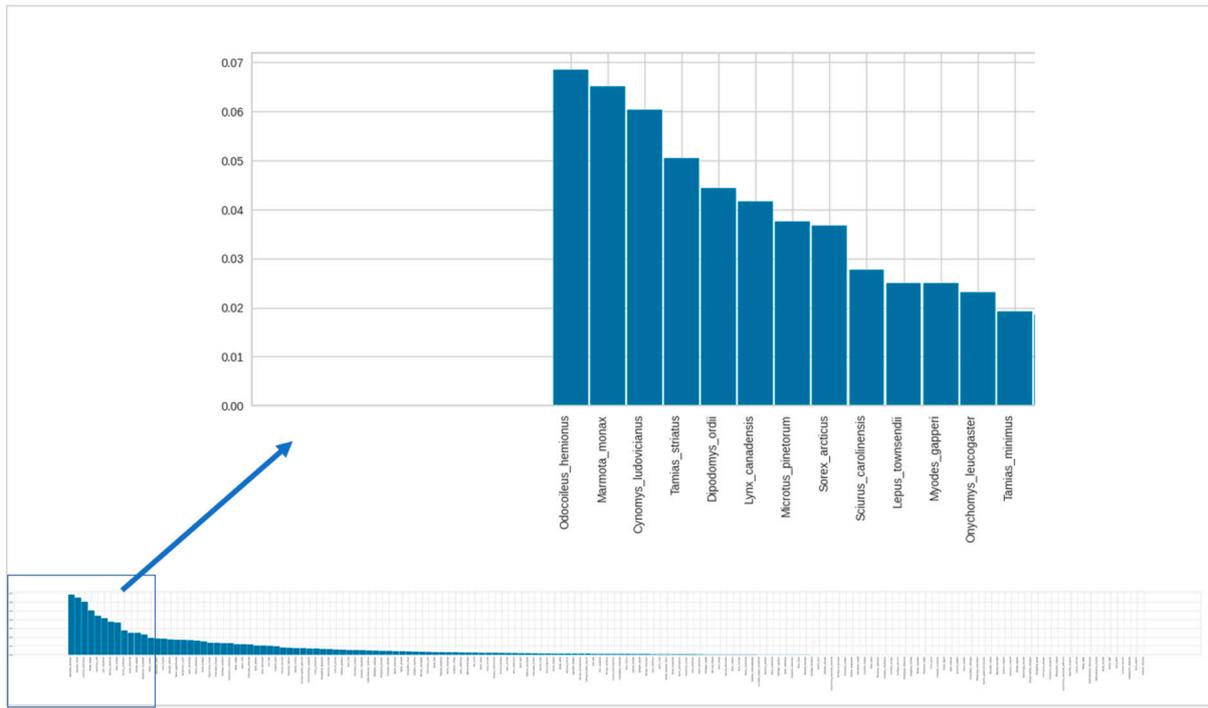


Figure A6. Histogram showing variable importance in classifying Lyme disease risk for the Mid-continent region.

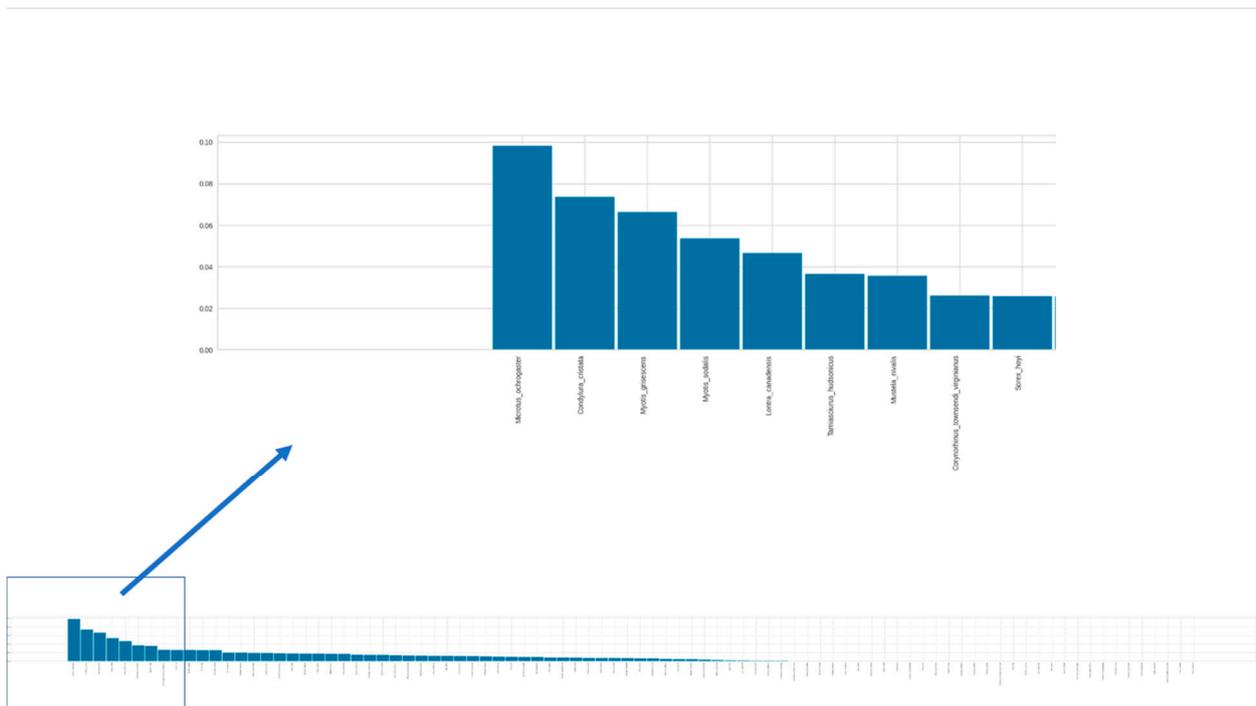


Figure A7. Histogram showing variable importance in classifying Lyme disease risk for the North-east region.

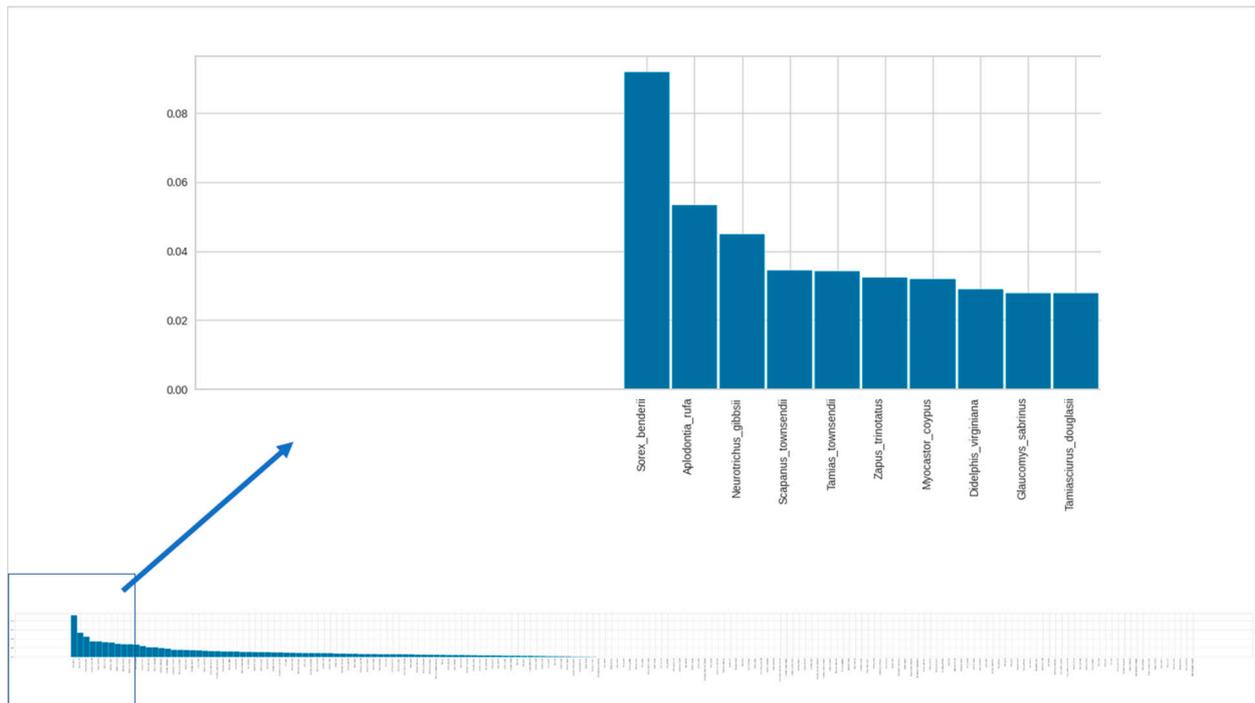


Figure A8. Histogram showing variable importance in classifying Lyme disease risk for the North-west and Pacific Islands region.

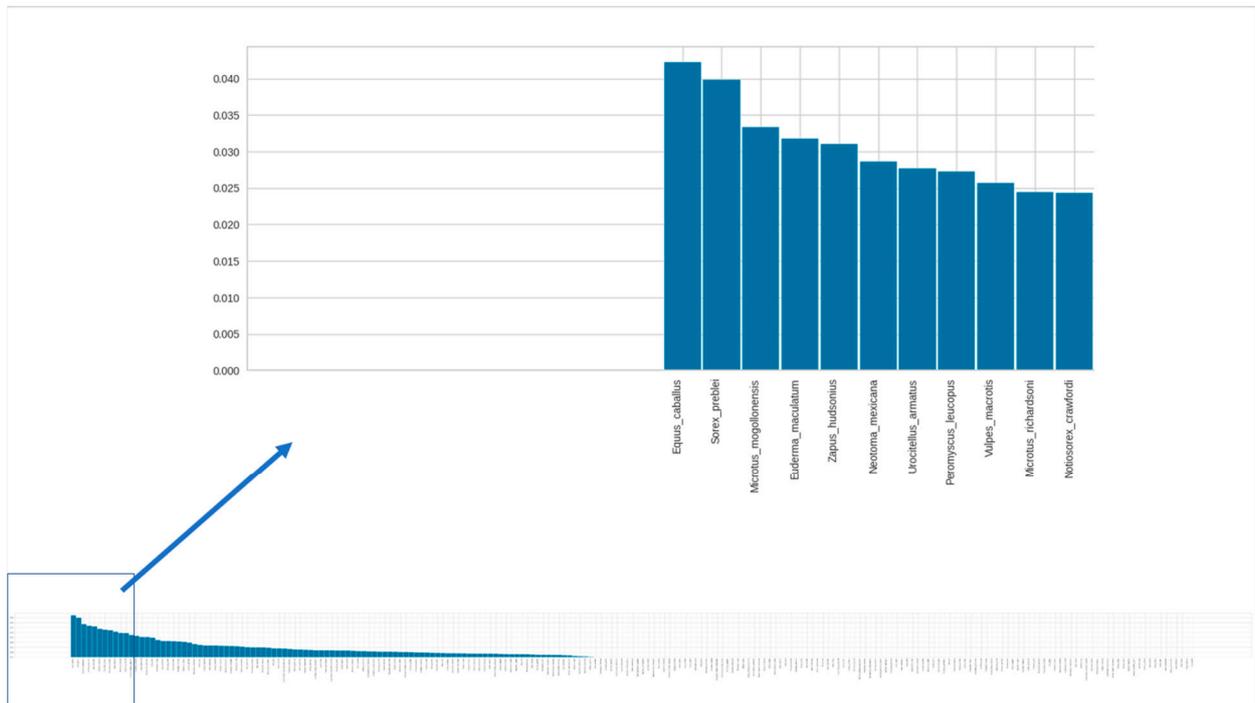


Figure A9. Histogram showing variable importance in classifying Lyme disease risk for the Rocky Mountains region.

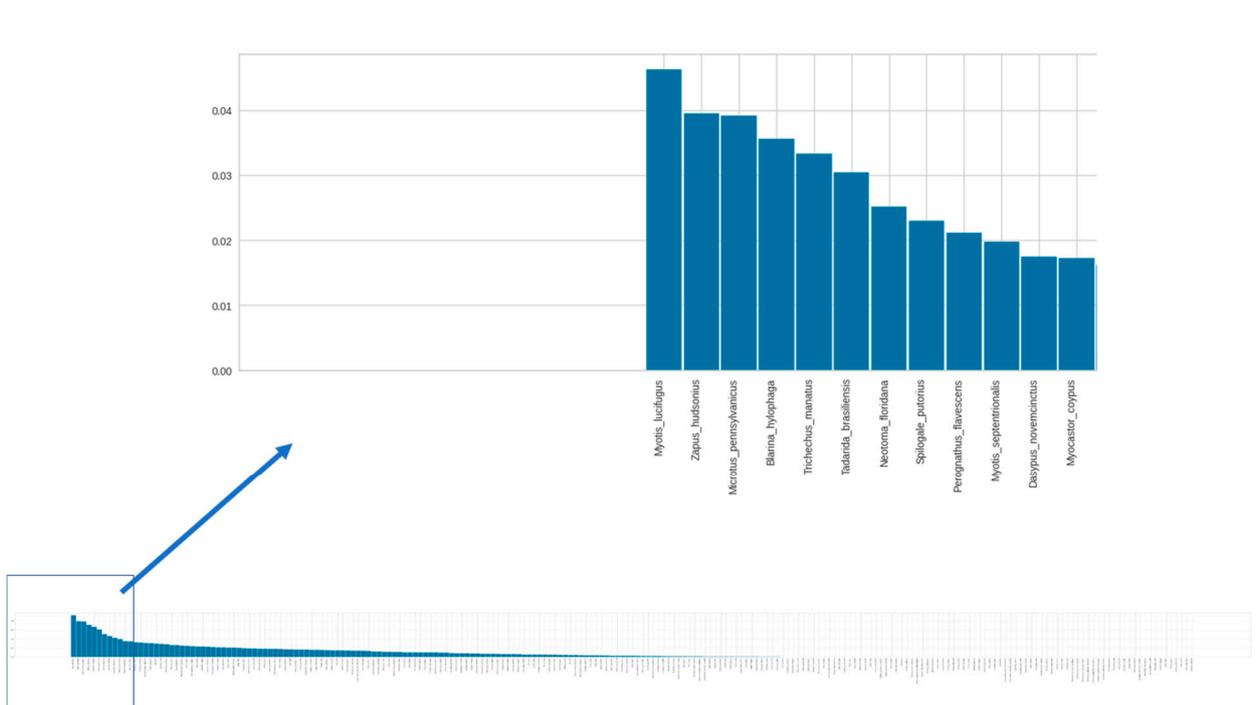


Figure A10. Histogram showing variable importance in classifying Lyme disease risk for the South-east region.

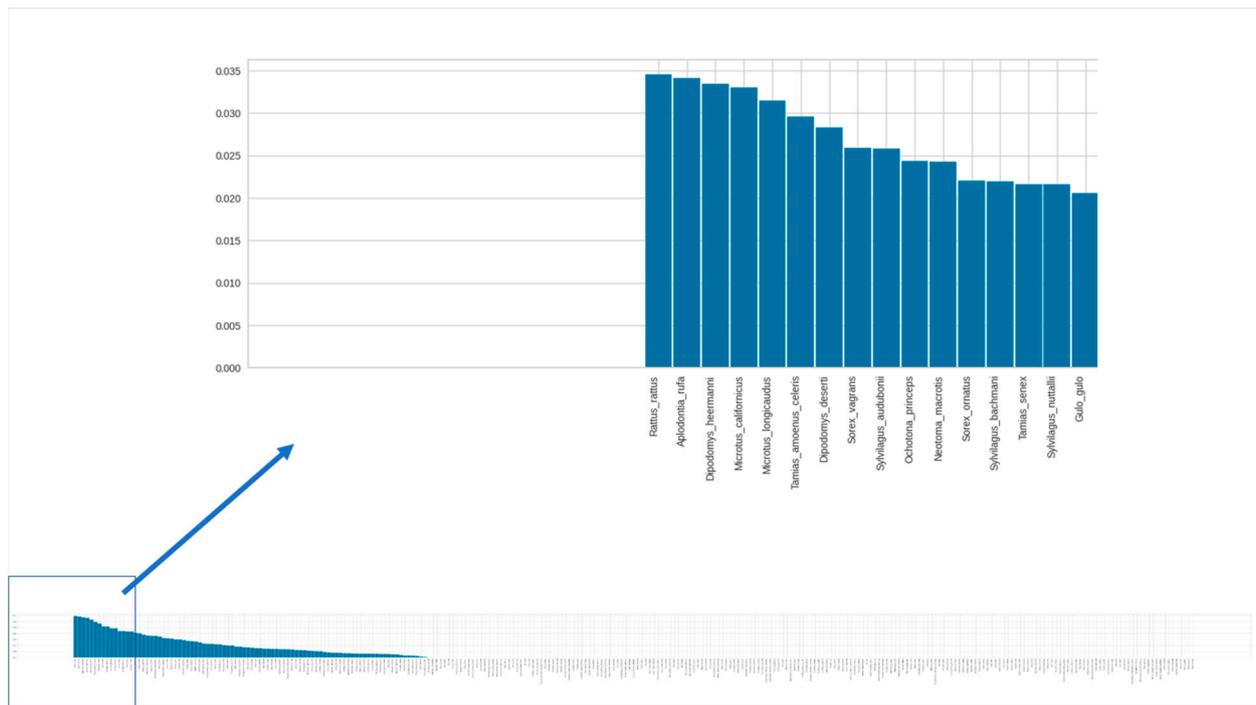


Figure A11. Histogram showing variable importance in classifying Lyme disease risk for the South-west region.

Appendix C

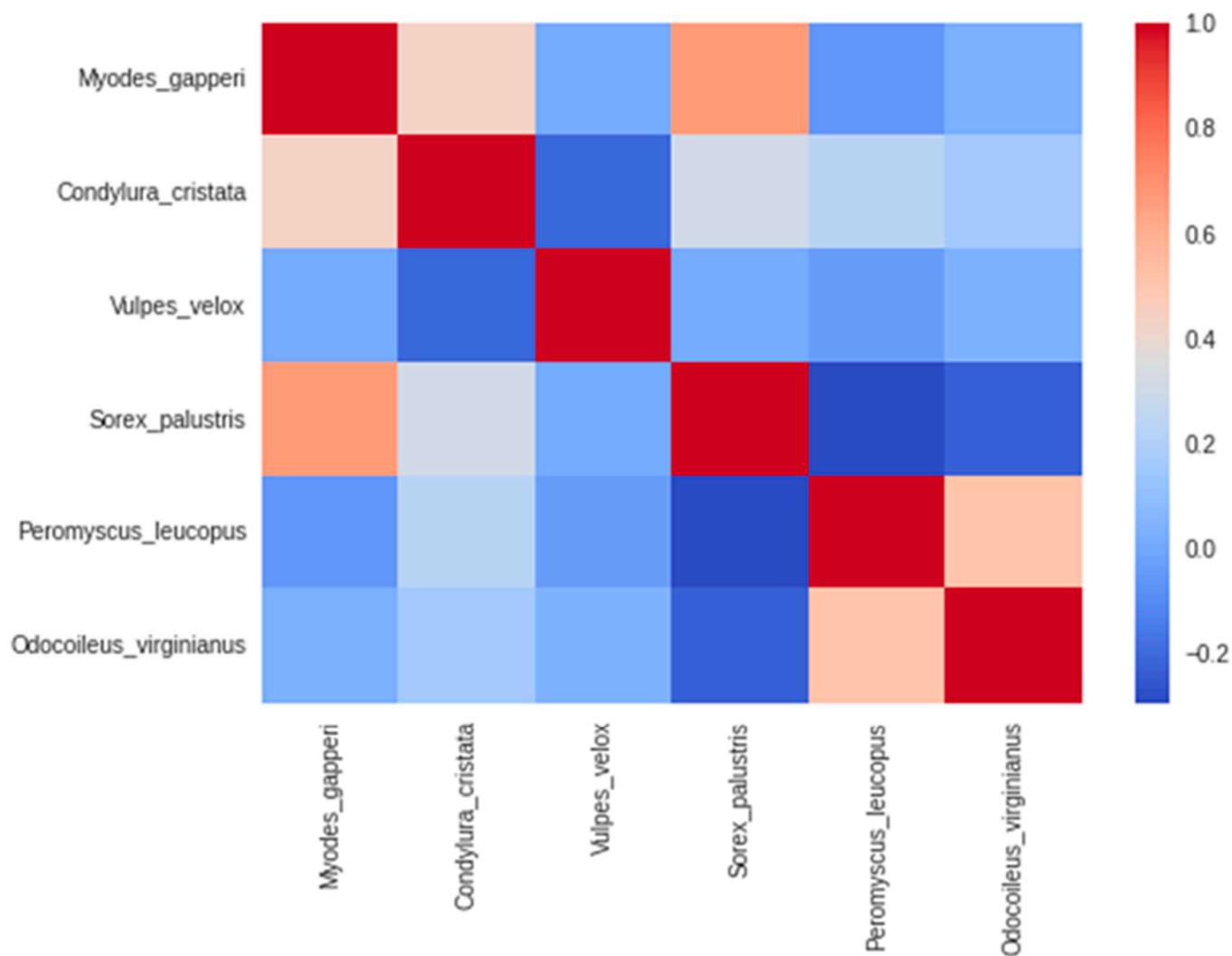


Figure A12. Heatmap showing pairwise correlations for selected mammal species from the random forest model in the national analysis.

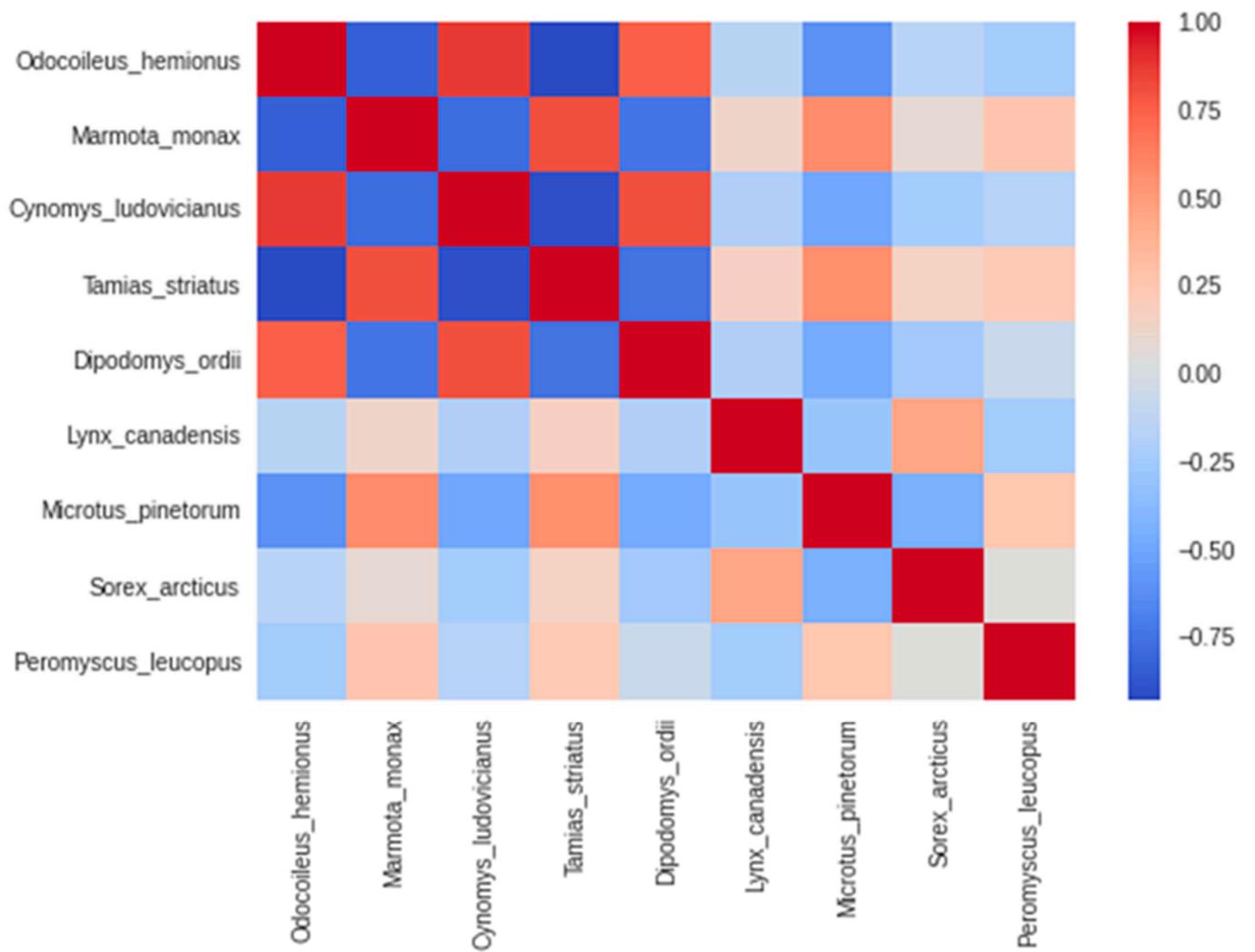


Figure A13. Heatmap showing pairwise correlations for selected mammal species from the random forest model for the Midcontinent region.

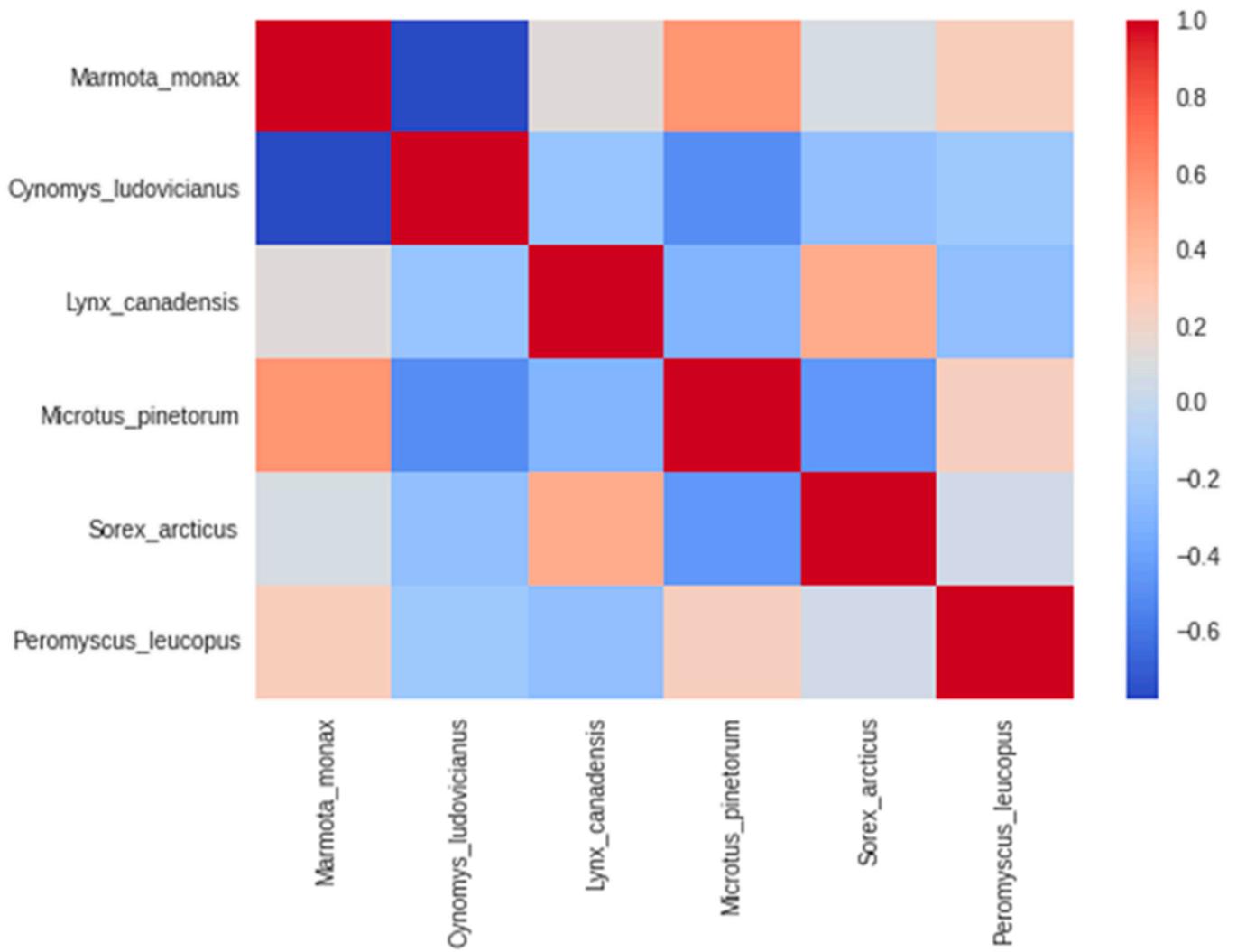


Figure A14. Heatmap showing pairwise correlations for selected mammal species for multinomial regression analysis for the Midcontinent region after autocorrelated species were removed.

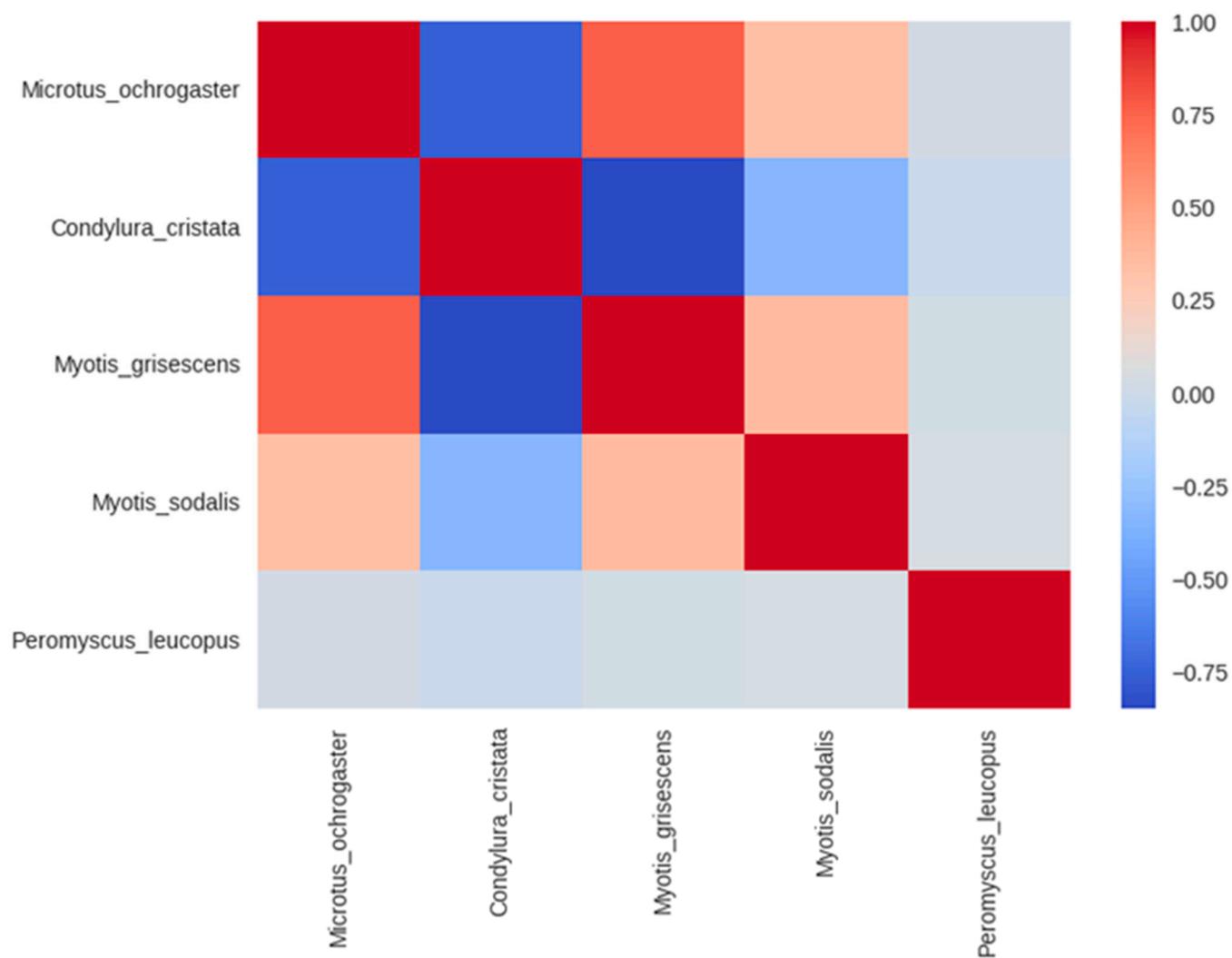


Figure A15. Heatmap showing pairwise correlations for selected mammal species from the random forest model for the Northeast region.

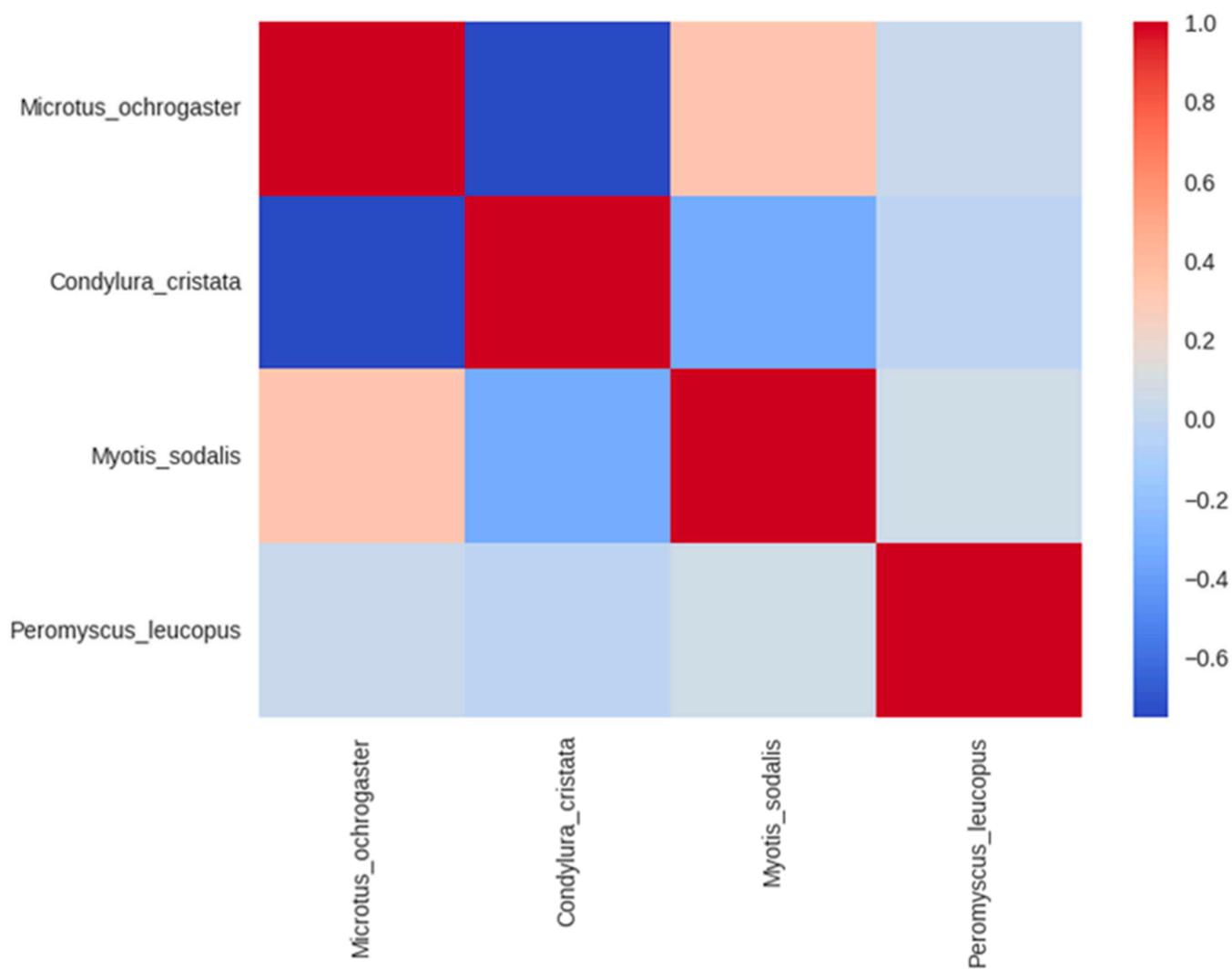


Figure A16. Heatmap showing pairwise correlations for selected mammal species for multinomial regression analysis for the Northeast region after autocorrelated species were removed.

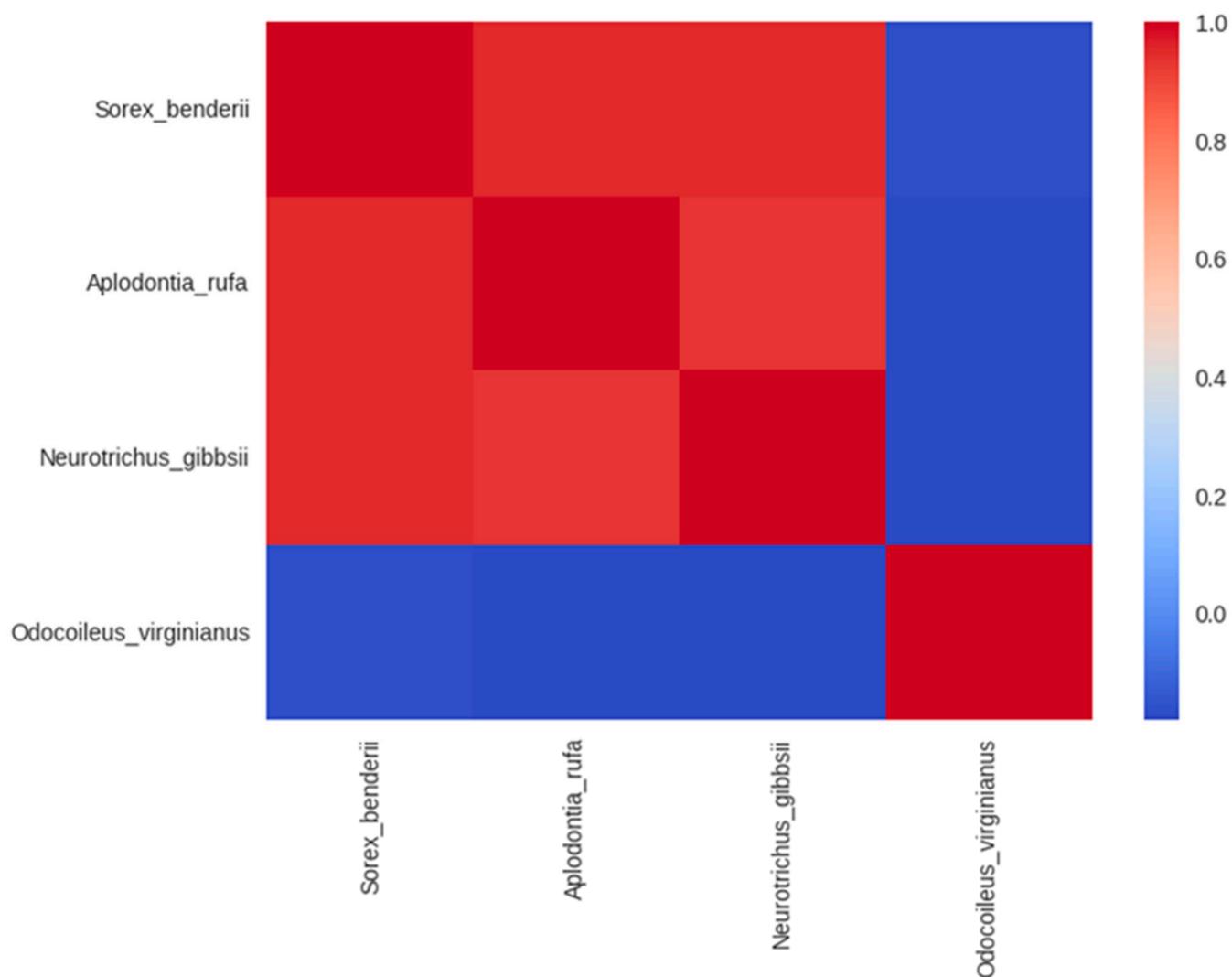


Figure A17. Heatmap showing pairwise correlations for selected mammal species from the random forest model for the Northwest and Pacific Islands region.

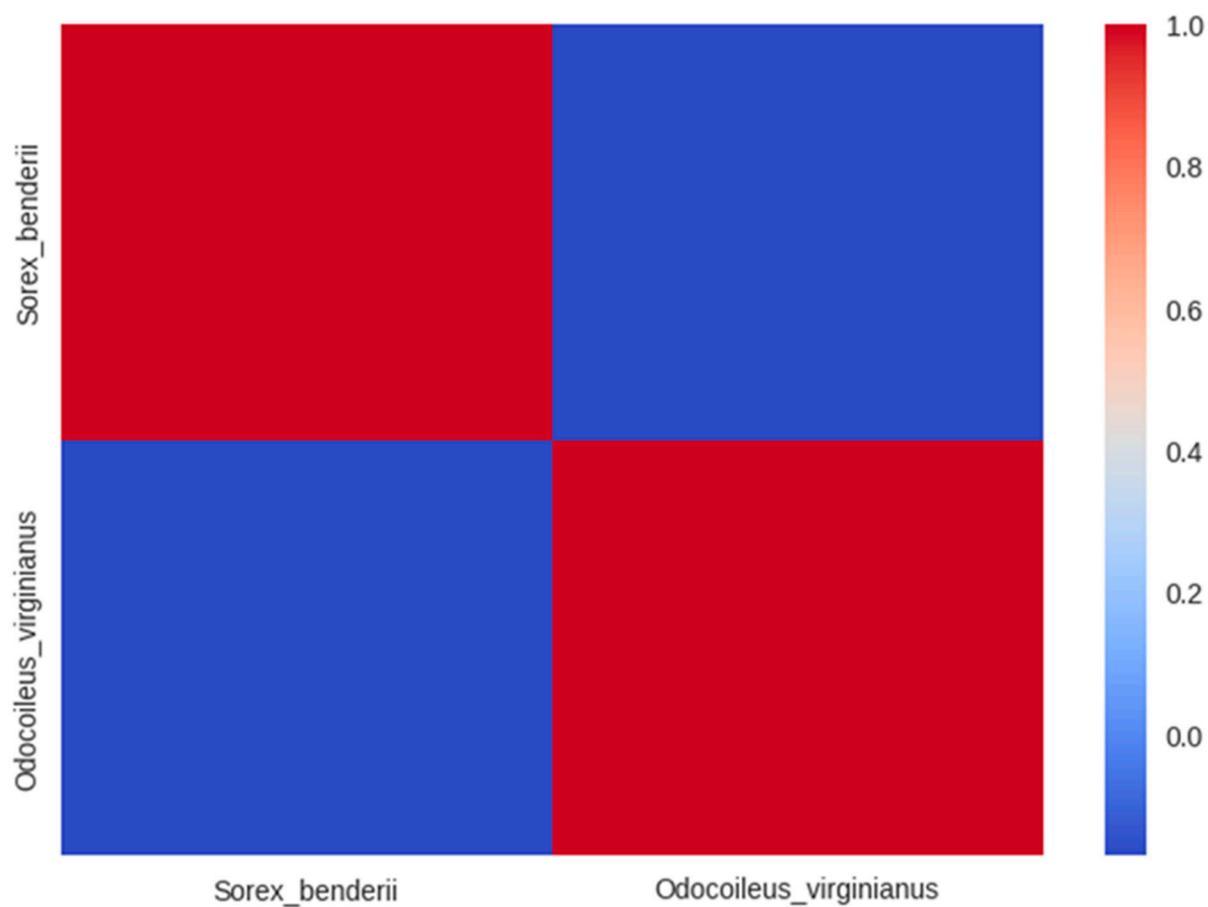


Figure A18. Heatmap showing pairwise correlations for selected mammal species from the random forest model for the Northwest and Pacific Islands region after autocorrelated species were removed.

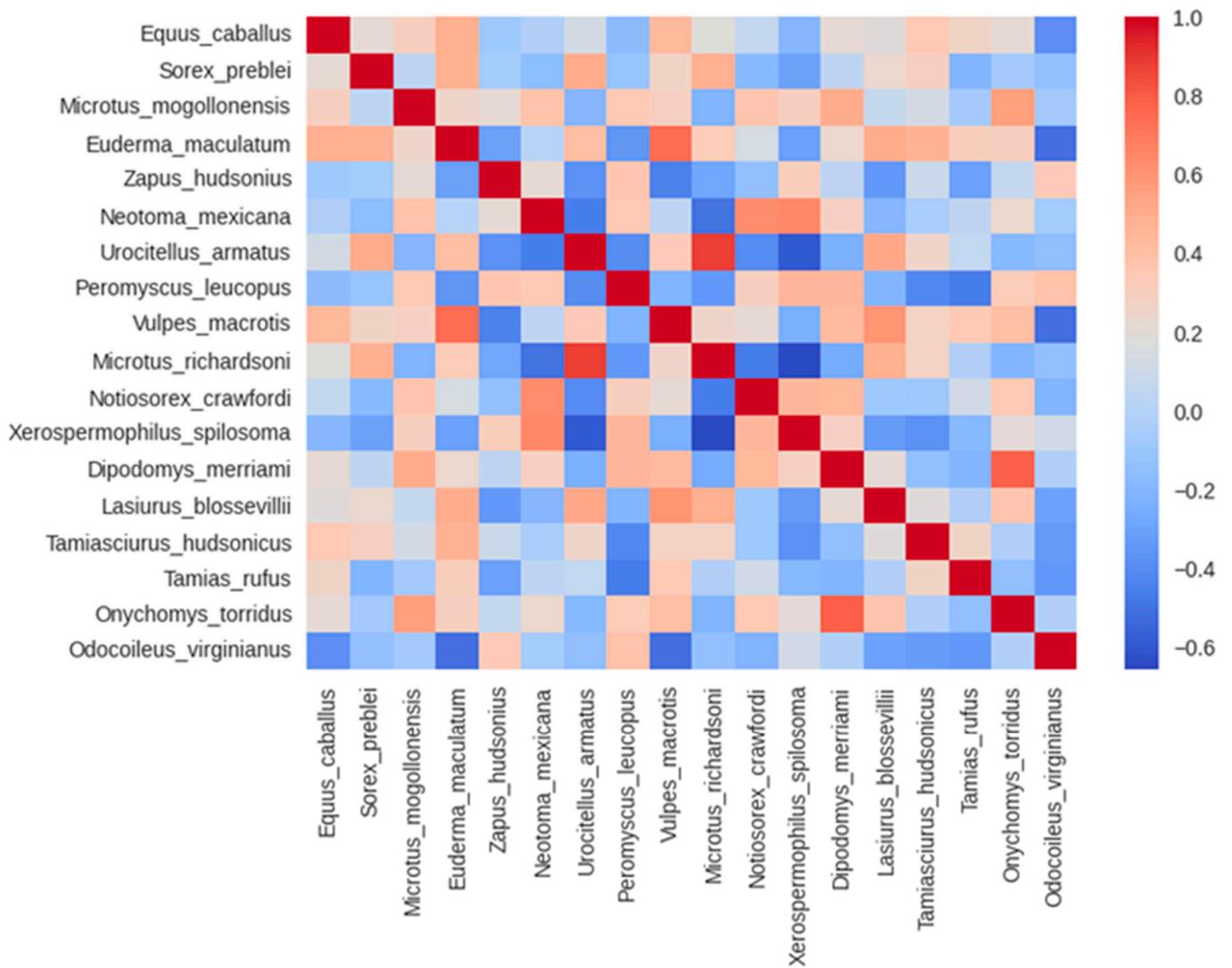


Figure A19. Heatmap showing pairwise correlations for selected mammal species from the random forest model for the Rocky Mountains region.

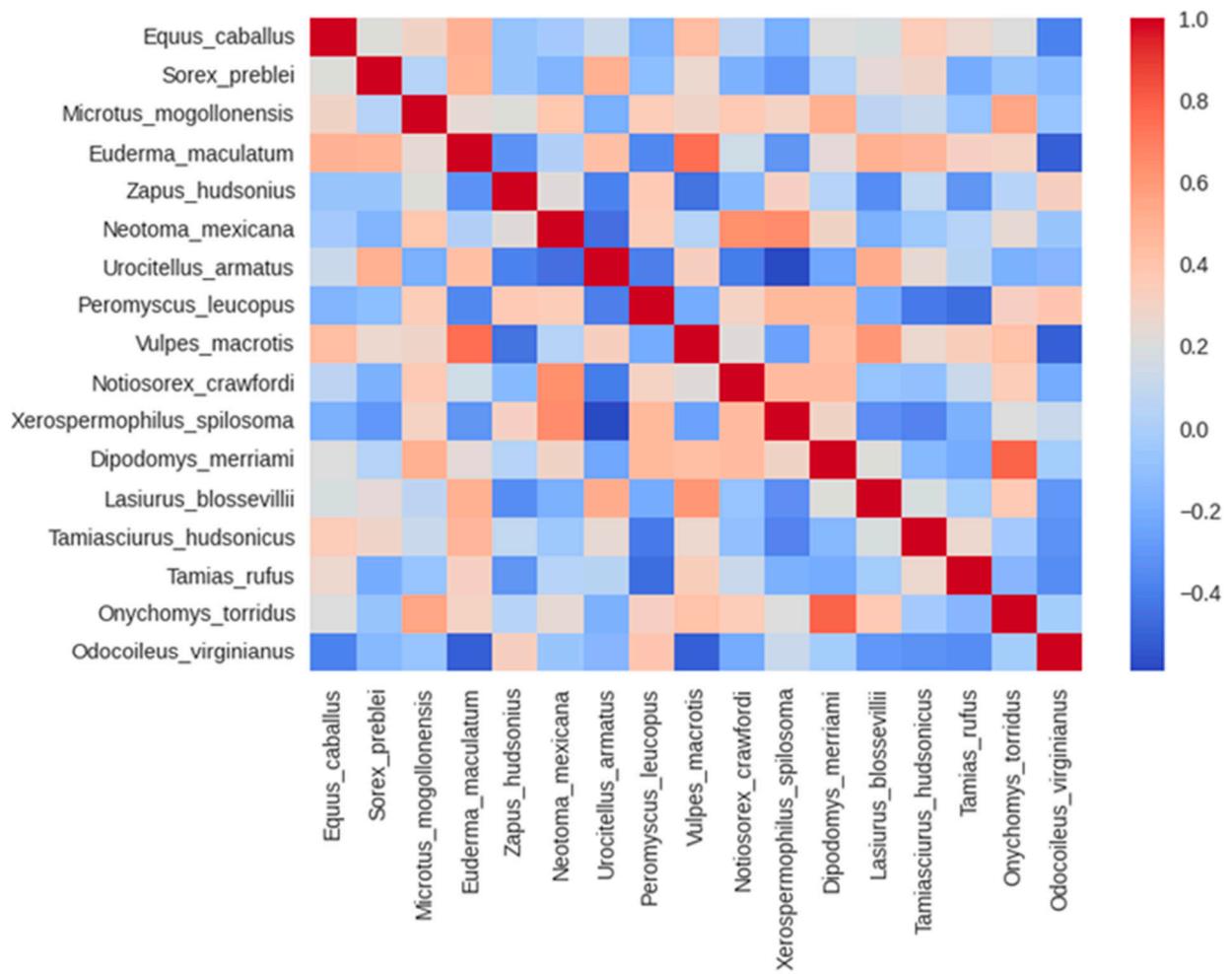


Figure A20. Heatmap showing pairwise correlations for selected mammal species for multinomial regression analysis for the Rocky Mountains region after autocorrelated species were removed.

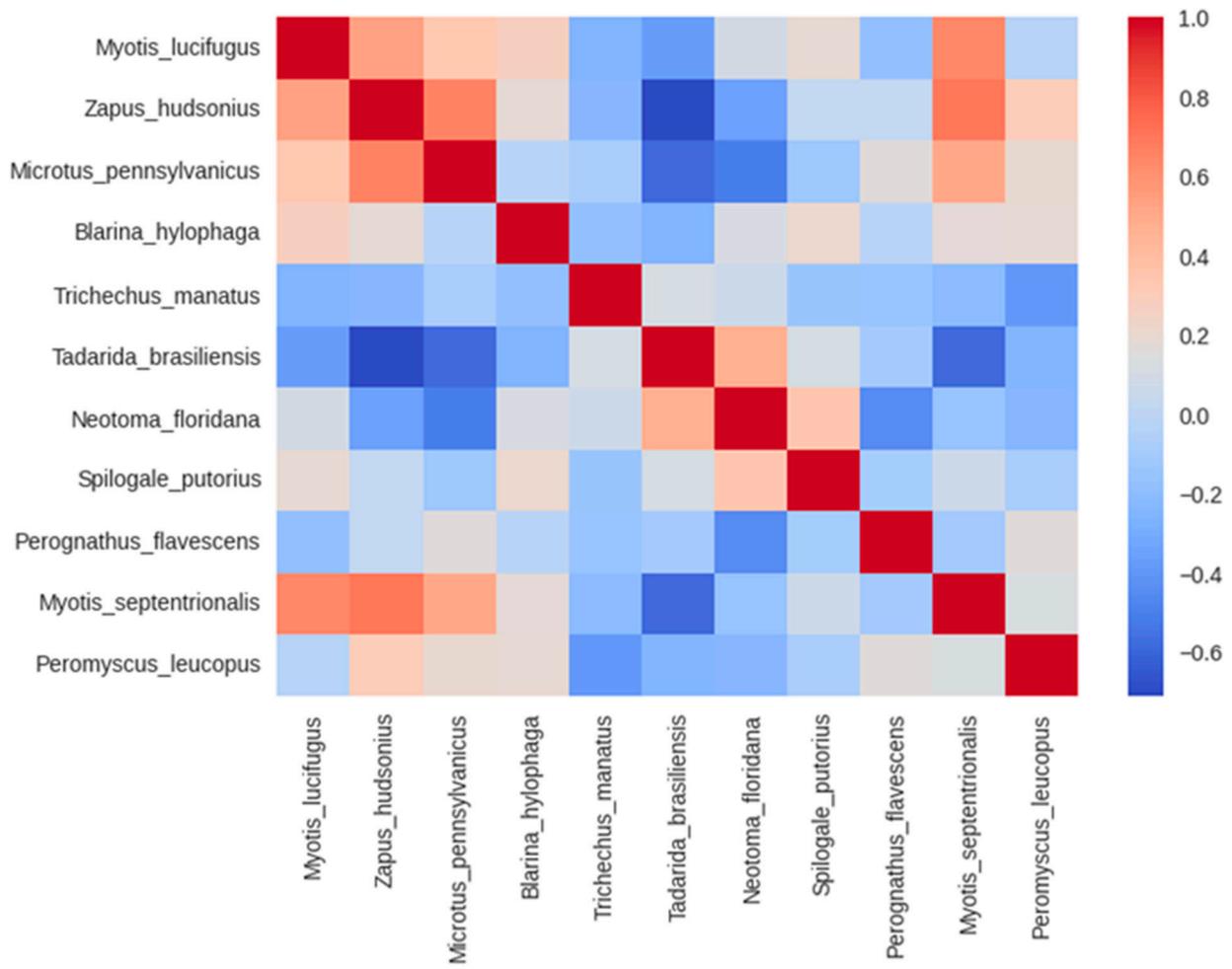


Figure A21. Heatmap showing pairwise correlations for selected mammal species from the random forest model for the Southeast region.

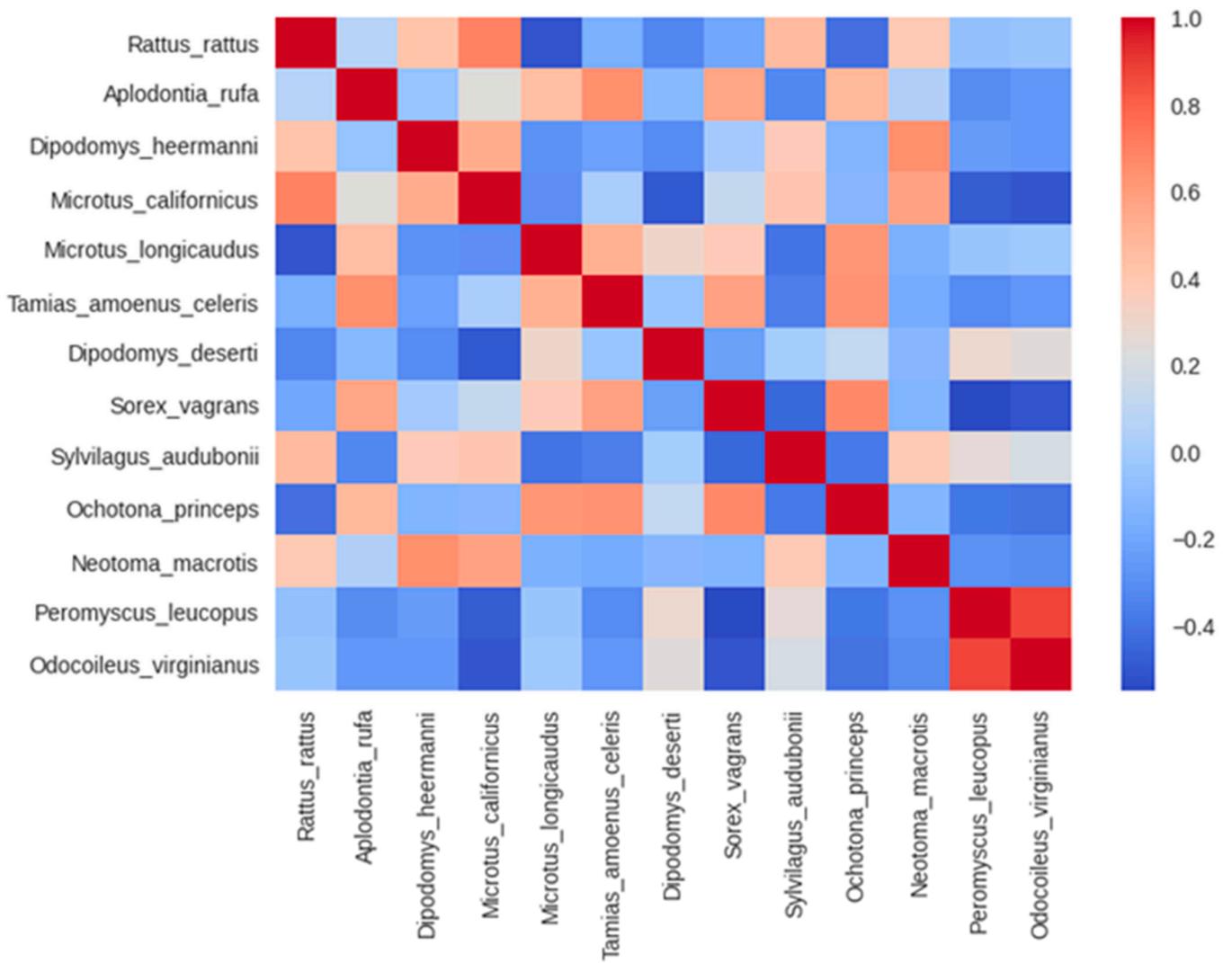


Figure A22. Heatmap showing pairwise correlations for selected mammal species from the random forest model for the Southwest region.

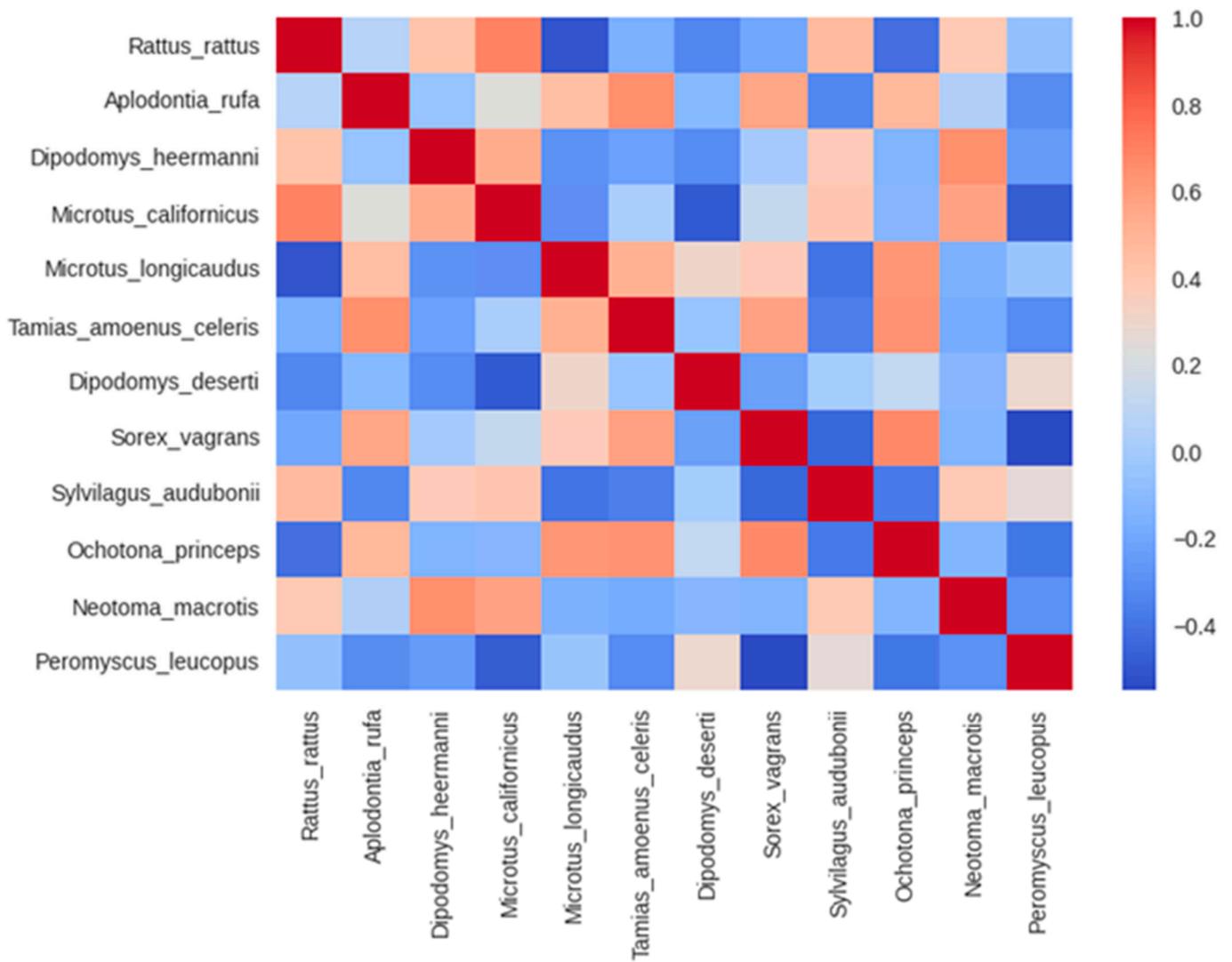


Figure A23. Heatmap showing pairwise correlations for selected mammal species for multinomial regression analysis for the Southwest region after autocorrelated species were removed.

Appendix D

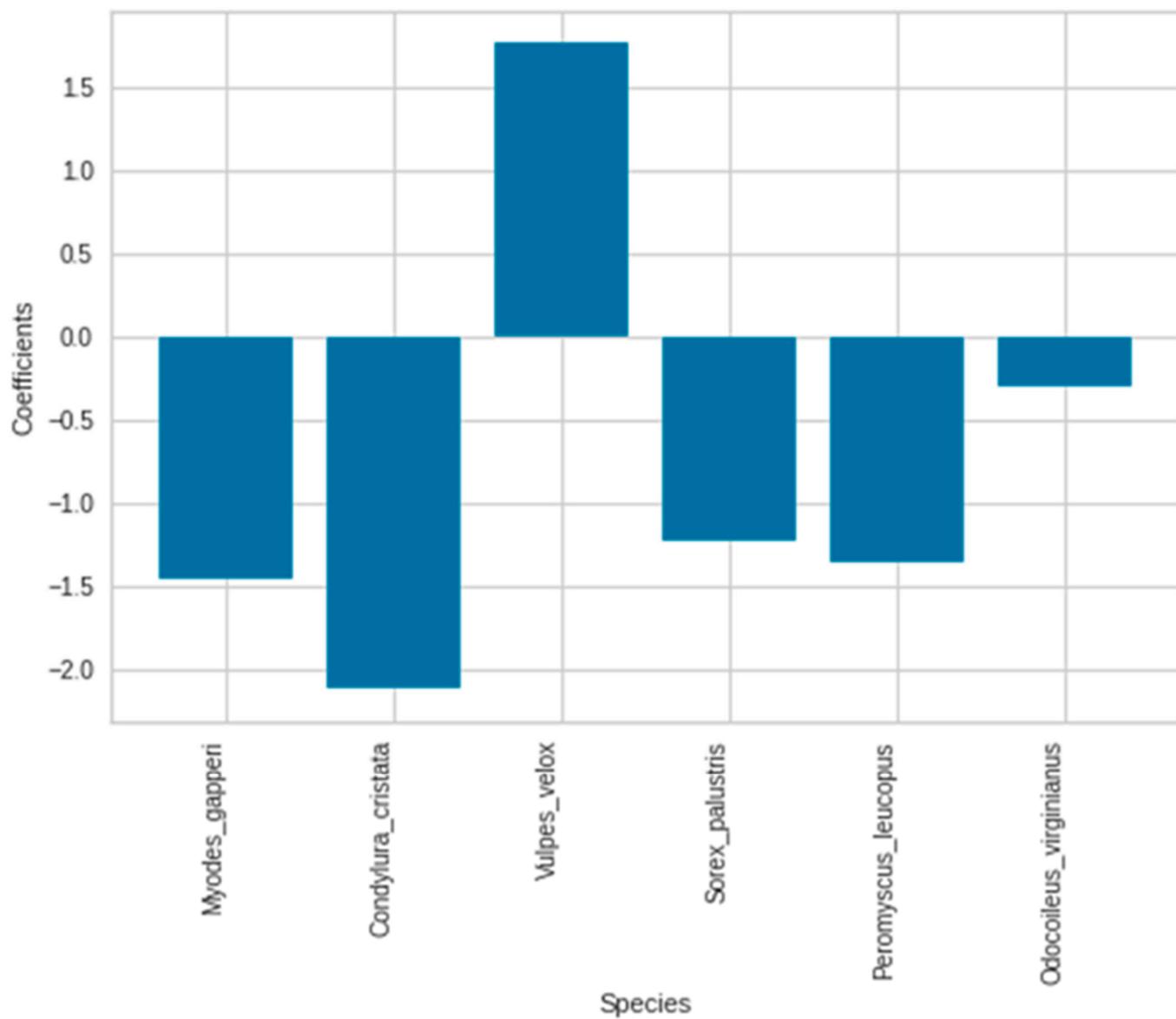


Figure A24. Species correlation coefficients obtained from multinomial logistic regression for the national analysis.

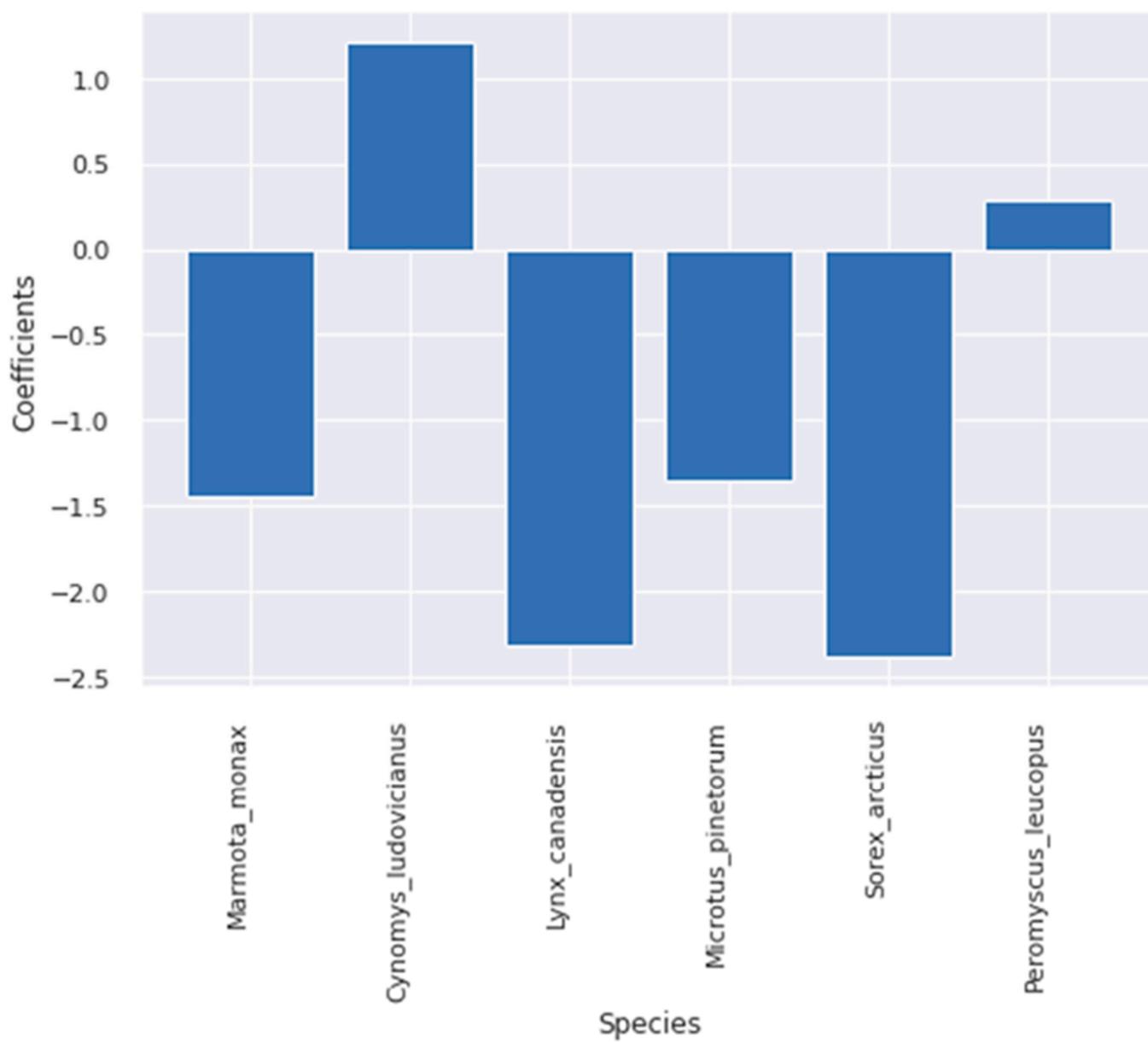


Figure A25. Species correlation coefficients obtained from multinomial logistic regression of the Midcontinent region.

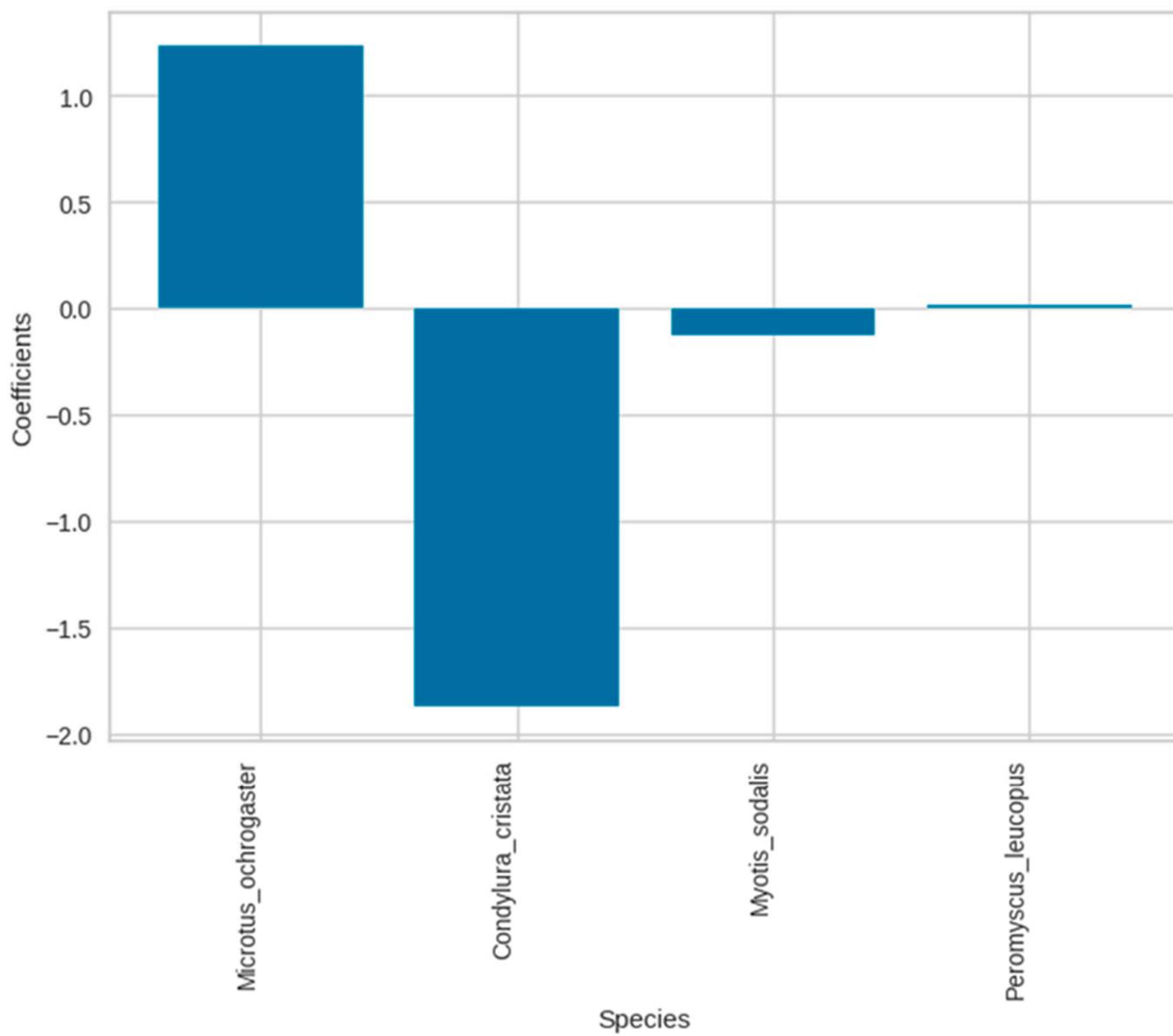


Figure A26. Species correlation coefficients obtained from multinomial logistic regression of the Northeast region.

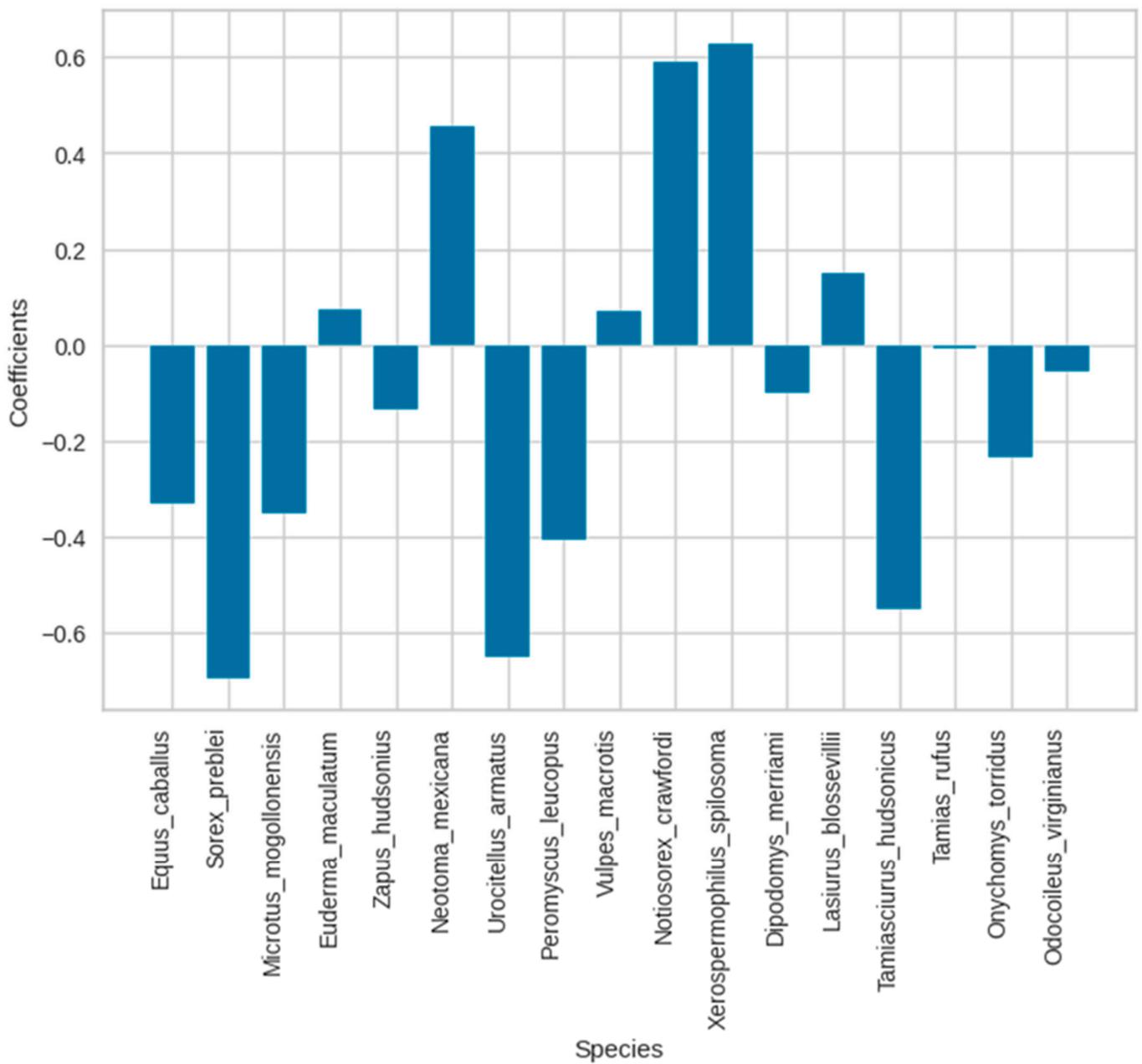


Figure A27. Species correlation coefficients obtained from multinomial logistic regression of the Rocky Mountains region.

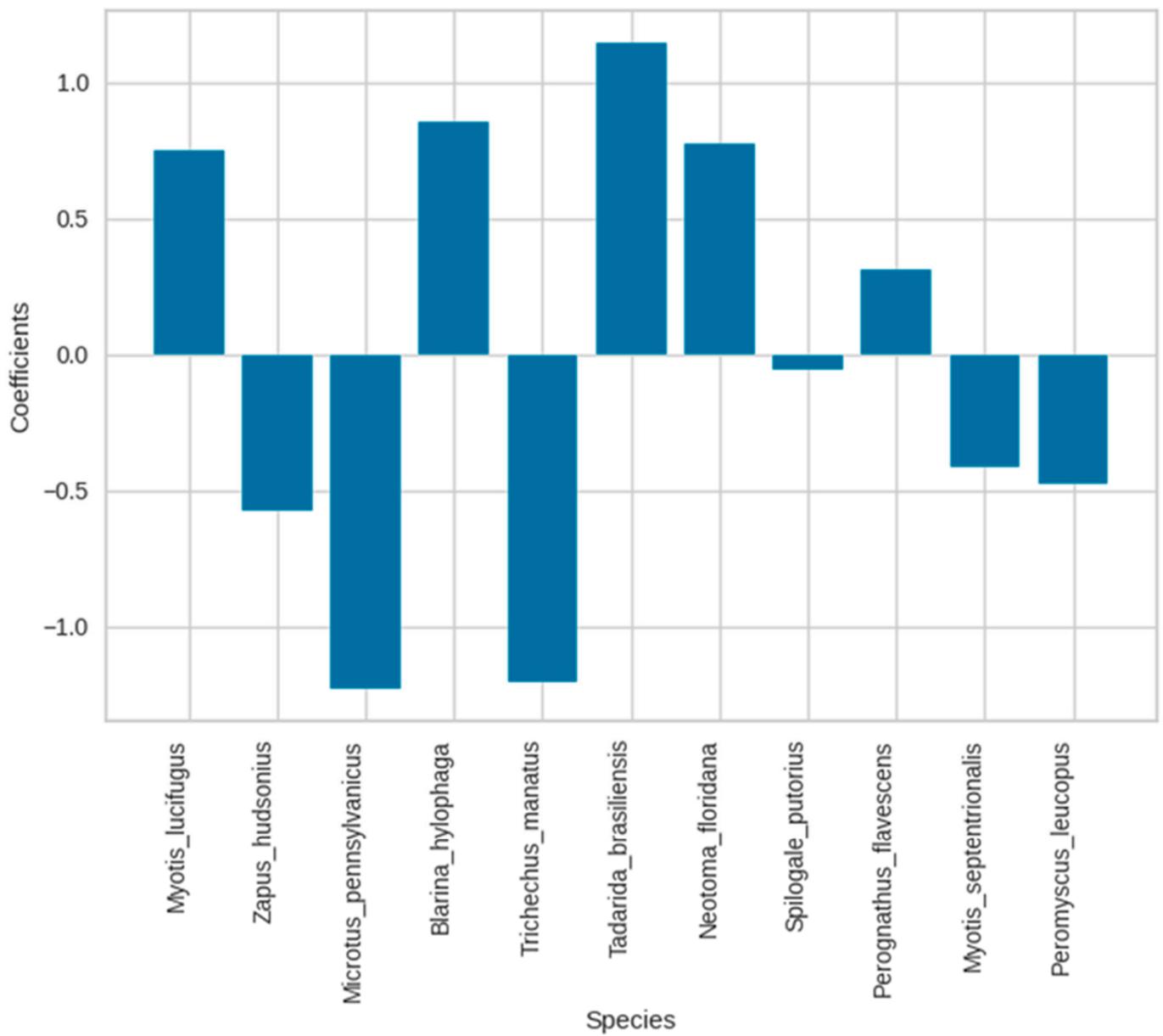


Figure A28. Species correlation coefficients obtained from multinomial logistic regression of the Southeast region.

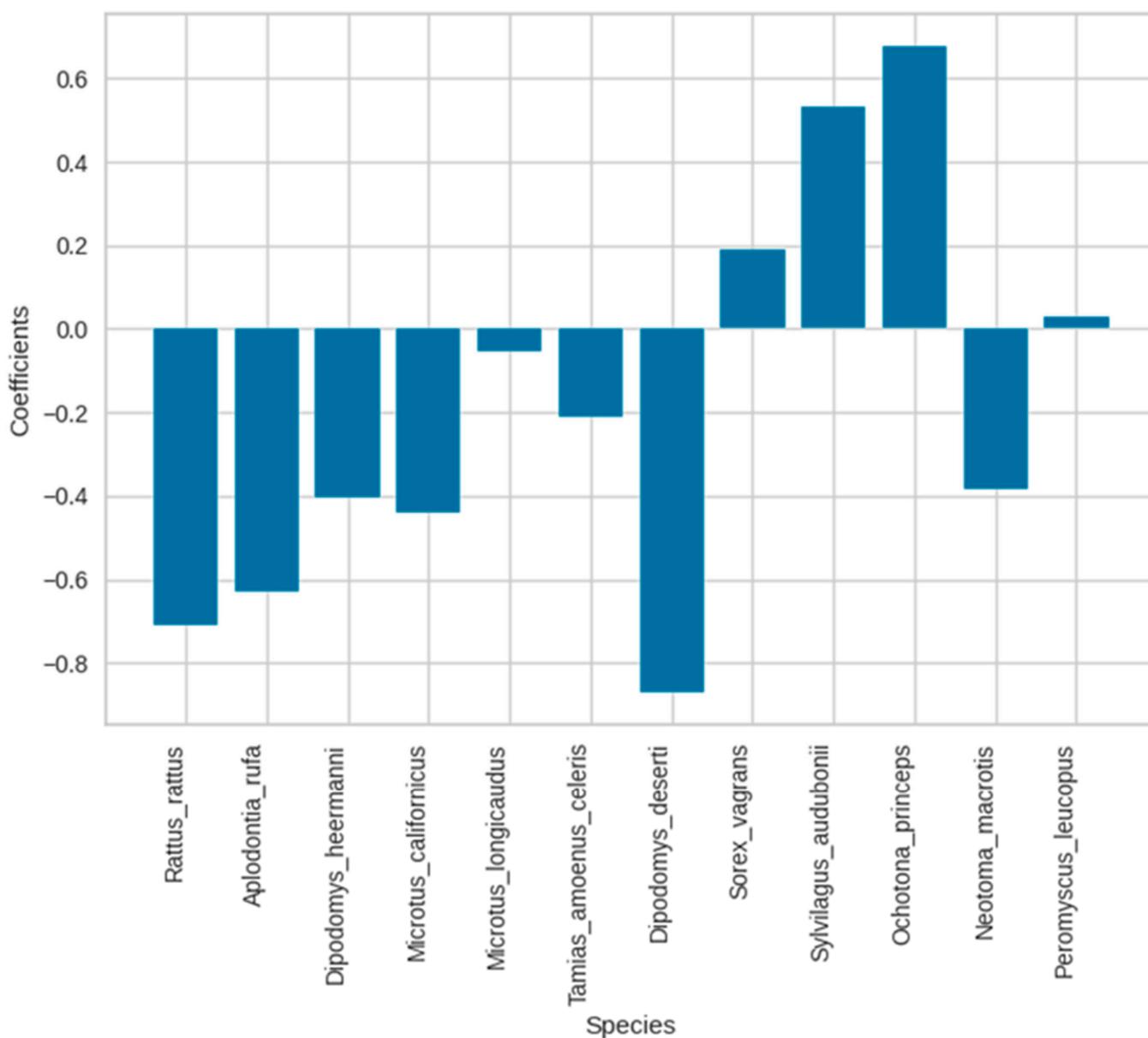


Figure A29. Species correlation coefficients obtained from multinomial logistic regression of the Southwest region.

References

- Centers for Disease Control and Prevention. Overview of Tickborne Diseases. Centers for Disease Control and Prevention. 2018. Available online: <https://www.cdc.gov/ticks/tickbornediseases/overview.html> (accessed on 30 April 2024).
- Arahal, D.R.; Bull, C.T.; Busse, H.; Chuvochina, M.; Dedys, S.N.; Fournier, P.E.; Konstantinidis, K.T.; Parker, C.T.; Rossello-Mora, R.; Ventosa, A. Judicial Opinions 123–127. *Int. J. Syst. Evol. Microbiol.* **2022**, *72*, 005708.
- Voordouw, M.J.; Lachish, S.; Dolan, M.C. The Lyme disease pathogen has no effect on the survival of its rodent reservoir host. *PLoS ONE* **2015**, *10*, e0118265.
- Bouchard, C.; Aenishaenslin, C.; Rees, E.E.; Koffi, J.K.; Pelcat, Y.; Ripoche, M.; Milord, F.; Lindsay, L.R.; Ogden, N.H.; Leighton, P.A. Integrated social-behavioral and ecological risk maps to prioritize public health responses to Lyme disease. *Environ. Health Perspect.* **2018**, *126*, 047008.
- Gomes-Solecki, M.; Arnaboldi, P.M.; Backenson, P.B.; Benach, J.L.; Cooper, C.L.; Dattwyler, R.J.; Diuk-Wasser, M.; Fikrig, E.; Hovius, J.W.; Laegreid, W.; et al. Protective Immunity and New Vaccines for Lyme Disease. *Clin. Infect. Dis.* **2020**, *70*, 1768–1773.
- Fessenden, J. FDA Approves Investigational New Drug Application for Lyme PrEP, a Pre-Exposure Prophylaxis to Prevent Lyme Disease. *UMASS Chan Medical School News*, 24 February 2021. Available online: <https://www.umassmed.edu/news/news-archives/2021/02/preventative-shot-for-lyme-disease-developed-at-umass-medical-school-enters-clinical-trial/> (accessed on 15 August 2021).

7. Cross, M.L.; Buddle, B.M.; Aldwell, F.E. The potential of oral vaccines for disease control in wildlife species. *Vet. J.* **2007**, *174*, 472–480.
8. Van Buskirk, J.; Ostfeld, R.S. Controlling Lyme disease by modifying the density and species composition of Tick hosts. *Ecol. Appl.* **1995**, *5*, 1133–1140.
9. Artois, M.; Blancou, J.; Dupeyroux, O.; Gilot-Fromont, E. Sustainable control of zoonotic pathogens in wildlife: How to be fair to wild animals? *Sci. Tech. Rev.* **2011**, *30*, 733–743.
10. Williams, S.C.; Stafford III, K.C.; Molaei, G.; Linske, M.A. Integrated control of nymphal *Ixodes scapularis*: Effectiveness of white-tailed deer reduction, the entomopathogenic fungus *Metarhizium anisopliae*, and Fipronil-based rodent bait boxes. *Vector-Borne Zoonotic Dis.* **2018**, *18*, 55–64.
11. Keesing, F.; Mowry, S.; Bremer, W.; Duerr, S.; Evans, A.S.; Fischhoff, I.R.; Hinckley, A.F.; Hook, S.A.; Keating, F.; Pendleton, J.; et al. Effects of tick-control interventions on tick abundance, human encounters with ticks, and incidence of tickborne diseases in residential neighborhoods, New York, USA. *Emerg. Infect. Dis.* **2022**, *28*, 957–966.
12. Ozdenerol, E. GIS and remote sensing use in the exploration of Lyme disease epidemiology. *Int. J. Environ. Res. Public Health* **2015**, *12*, 15181–15203.
13. Salomon, J.; Lawrence, A.; Crews, A.; Sambado, S.; Swei, A. Host infection and community composition predict vector burden. *Oecologia* **2021**, *196*, 305–316.
14. Millins, C.; Leo, W.; MacInnes, I.; Ferguson, J.; Charlesworth, G.; Nayar, D.; Davison, R.; Yardley, J.; Kilbride, E.; Huntley, S.; et al. Emergence of Lyme disease on treeless islands, Scotland, United Kingdom. *Emerg. Infect. Dis.* **2021**, *27*, 538–546.
15. Ginsberg, H.S.; Buckley, P.A.; Balmforth, M.G.; Zhioua, E.; Mitra, S.; Buckley, F.G. Reservoir competence of native North American birds for the Lyme disease spirochete, *Borrelia burgdorferi*. *J. Med. Entomol.* **2005**, *42*, 445–449.
16. Rand, P.W.; Lubelczyk, C.; Holman, M.S.; Lacombe, E.H.; Smith, R.P. Abundance of *Ixodes scapularis* (Acari: Ixodidae) after the complete removal of deer from an isolated offshore island, endemic for Lyme disease. *J. Med. Entomol.* **2004**, *41*, 779–784.
17. MacDonald, S.O.; Cook, J.A. *Recent Mammals of Alaska*; University of Alaska Press: Fairbanks, AK, USA, 2010; p. 399.
18. Learn, J.; Chapman, B.R.; Ford, W.M. Star-nosed mole: *Condylura cristata*. In *The Land Manager's Guide to Mammals of the South*; Trani, M.K., Ford, W.M., Brian, R., Eds.; Forest Service, United States Department of Agriculture: Washington, DC, USA, 2007; pp. 113–116.
19. Getz, L.L. Notes on the local distribution of *Peromyscus leucopus* and *Zapus hudsonius*. *Am. Midl. Nat.* **1961**, *65*, 486–500.
20. Ozdenerol, E.; Bingham-Byrne, R.M.; Seboly, J.D. The Effects of Lifestyle on the Risk of Lyme Disease in the United States: Evaluation of Market Segmentation Systems in Prevention and Control Strategies. *Int. J. Environ. Res Public Health* **2021**, *18*, 12883.
21. Adler, G.H.; Wilson, M.L. Demography of a habitat generalist, the white-footed mouse, in a heterogeneous environment. *Ecology* **1987**, *68*, 1785–1796.
22. Andstedt, E. Development and Evaluation of a Habitat Suitability Model for White-Tailed Deer in an Agricultural Landscape. Master's Thesis, Minnesota State University, Mankato, MN, USA, 2016. Available online: <https://cornerstone.lib.mnsu.edu/cgi/viewcontent.cgi?article=1645&context=etds#:~:text=Development%20and%20Evaluation%20of%20a%20Habitat%20Suitability%20Model,Minnesota%20State%20University%2C%20Mankato%20Mankato%2C%20Minnesota%20July%202016> (accessed on 15 September 2021).
23. Kurokawa, C.; Lynn, G.E.; Pedra, J.H.F.; Pal, U.; Narasimhan, S.; Fikrig, E. Interactions between *Borrelia burgdorferi* and ticks. *Nat. Rev. Microbiol.* **2020**, *18*, 587–600.
24. Jackson, V.L.; Choate, J.R. Dens and den sites of the swift fox, *Vulpes velox*. *Southwest. Nat.* **2000**, *45*, 212–220.
25. Ginsberg, H.S.; Hickling, G.J.; Burke, R.L.; Ogden, N.H.; Beati, L.; LeBrun, R.A.; Arsnoe, I.M.; Gerhold, R.; Han, S.; Jackson, K.; et al. Why Lyme disease is common in the northern US, but rare in the south: The roles of host choice, host-seeking behavior, and tick density. *PLoS Biol.* **2020**, *19*, e3001066.
26. Werden, L.; Barker, I.K.; Bowman, J.; Gonzales, E.K.; Leighton, P.A.; Lindsay, L.R.; Jardine, C.M. Geography, deer, and host biodiversity shape the pattern of Lyme disease emergence in the Thousand Islands Archipelago of Ontario, Canada. *PLoS ONE* **2014**, *9*, e85640.
27. Chen, D.; Wong, H.; Belanger, P.; Moore, K.; Peterson, M.; Cunningham, J. Analyzing the Correlation between Deer Habitat and the Component of the Risk for Lyme Disease in Eastern Ontario, Canada: A GIS-Based Approach. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 105–123.
28. Morshed, M.G.; Scott, J.D.; Fernando, K.; Beati, L.; Mazerolle, D.F.; Geddes, G.; Durden, L.A. Migratory songbirds disperse ticks across Canada, and first isolation of the Lyme disease spirochete, *Borrelia burgdorferi*, from the avian tick, *Ixodes auritulus*. *J. Parasitol.* **2005**, *91*, 780–790.
29. Gulia-Nuss, M.; Nuss, A.; Meyer, J.M.; Sonenshine, D.E.; Roe, R.M.; Waterhouse, R.M.; Sattelle, D.B.; de la Fuente, J.; Ribeiro, J.M.; Megy, K.; et al. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat. Commun.* **2016**, *7*, 10507.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.