Modeling the Geographic Distribution of *Bacillus anthracis*, the Causative Agent of Anthrax Disease, for the Contiguous United States using Predictive Ecologic Niche Modeling

Jason K. Blackburn,* Kristina M. McNyset, Andrew Curtis, and Martin E. Hugh-Jones

Spatial Epidemiology and Ecology Research Laboratory, Department of Geography, California State University, Fullerton, California; Environmental Protection Agency, Office of Research and Development/Western Ecology Division, Corvallis, Oregon; Department of Geography, College of Letters, Arts, and Sciences, University of Southern California, Los Angeles, California; Department of Environmental Studies, School of the Coast and Environment, Louisiana State University, Baton Rouge, Louisiana

**Abstract.** The ecology and distribution of *Bacillus anthracis* is poorly understood despite continued anthrax outbreaks in wildlife and livestock throughout the United States. Little work is available to define the potential environments that may lead to prolonged spore survival and subsequent outbreaks. This study used the genetic algorithm for rule-set prediction modeling system to model the ecological niche for *B. anthracis* in the contiguous United States using wildlife and livestock outbreaks and several environmental variables. The modeled niche is defined by a narrow range of normalized difference vegetation index, precipitation, and elevation, with the geographic distribution heavily concentrated in a narrow corridor from southwest Texas northward into the Dakotas and Minnesota. Because disease control programs rely on vaccination and carcass disposal, and vaccination in wildlife remains untenable, understanding the distribution of *B. anthracis* plays an important role in efforts to prevent/eradicate the disease. Likewise, these results potentially aid in differentiating endemic/natural outbreaks from industrial-contamination related outbreaks or bioterrorist attacks.

**INTRODUCTION**

Anthrax is a zoonotic disease that remains a problem in many countries worldwide, including the United States for herbivorous livestock and wildlife species, and secondarily humans. Despite being a disease of antiquity, little is known about the spatial ecology or the specific geography of environmental conditions that promote long-term survival of *Bacillus anthracis*, the causative agent of anthrax. Much of the literature addresses the ubiquitous nature of *B. anthracis* as a soil-borne bacteria and many of these studies recognize the environmental constraints on long-term survivability of *B. anthracis* spores in soils (e.g., soil pH, calcium levels). Although early literature argued that *B. anthracis* could replicate in soil, the current literature supports that *B. anthracis* only replicates in the animal host and can then survive long periods of dormancy in soil. Despite a body of literature defining the specifics of anthrax outbreaks regionally, and a body of work on the specific soil parameters required to maintain *B. anthracis* in the environment, few studies have evaluated the geographic distribution of these environmental characteristics. Some early work did map the distribution of anthrax outbreaks at the county level and through simple visualization suggested potential high risk areas based on soil type, but without any associated quantitative analysis. Although some spatio-temporal analyses have been used to relate specific soil conditions to anthrax outbreaks and spore persistence in the Kruger National Park in Africa, no such detailed spatial analyses exist for North America.

*Bacillus anthracis* is an old world bacterial species and was most likely introduced into the United States by early European colonists through cattle trading, bone meal production, and industrial hide tanning. However, the exact timing and specific introduction pathway remains unclear. Anthrax was a significant problem in domestic American livestock and wildlife through the 1950s. A series of surveys showed an annual increase in the number of counties affected by anthrax between 1915 and 1944, with 157 counties affected in 21 states in 1915 that expanded to 405 counties in 37 states by 1944. Continued surveys reported an increase in the spatial distribution and incidence of the disease up until the mid-1950s. With the introduction of a mass-produced vaccine, disease management improved and the number of counties affected decreased after the 1950s. Anthrax however, still remains a problem in both livestock and wildlife in certain parts of the United States. Although vaccination is inexpensive and readily available, it is often used in reaction to an outbreak rather than as a disease preventative. This reactive approach perpetuates the likelihood of future outbreaks by not preventing the continued introduction of viable spores into disease promoting environments. Further complicating disease eradication, the current vaccine is administered through injection and therefore only useful for livestock or farmed wild species that can be safely handled. As a result, in areas such as western Texas where the disease remains a problem in white-tailed deer *Odocoileus virginianus*, vaccination is ineffective for disease control until an oral vaccine can be developed.

Because anthrax remains a problem in both livestock and wildlife, a clear understanding of the spatial ecology and corresponding geographic distribution of the *B. anthracis* is essential. Given that the disease remains enzootic, surveillance efforts must be targeted on areas of greatest risk of infection, and cover all ecologic components of the disease (hosts, reservoirs, potential vectors, outbreak triggering climatic events). However, disease surveillance is expensive, requires multi-agency networks, and a multi-disciplinary approach. These networks must be clearly identified and the goal of the surveillance explicit. This study aims to identify the geographic potential of *B. anthracis* in the United States, further the current understanding of the disease, improve our under-
standing of its ecology, and target areas of risk that require further research or surveillance.

Ecologic niche modeling (ENM) is one tool for evaluating the potential distribution of species. Its approaches define the niche of the target species and predict its potential geographic and ecologic distribution through the analysis of relationships between combinations of environmental variables (e.g., temperature, precipitation, and elevation derived from digital maps or satellite data) and species' locality data. The ecologic niche follows the Hutchinsonian definition as the hyper-volume of ecologic parameters that allow a species to maintain populations without immigration.\textsuperscript{13-15}

We evaluated the specific geography of \textit{B. anthracis} in the continental United States using ENM to clearly delineate the areas most suitable for spore survival in the 48 contiguous states. Defining the potential spatial extent of \textit{B. anthracis} can be useful for generating new hypotheses about disease persistence and for targeting surveillance efforts to areas at greatest risk of potential disease presence, with the ultimate goal of providing sound methods for improving disease control. At the same time, a great deal of biologic information can be extracted from the modeling process, providing insights into the ecologic requirements, biogeography, and evolution of an important zoonotic disease agent.

\textbf{METHODS}

\textbf{Ecologic niche modeling.} This study was completed using the Genetic Algorithm for Rule-Set Prediction (GARP) to develop an ENM for anthrax in the United States. For this study, the DesktopGARP version 1.1.3 [DG] application was used to develop all GARP models (available from http://www.lifemapper.org/desktopgarp). GARP is a presence-only modeling technique that determines non-random associations between point localities (anthrax outbreak locations) and environmental parameters (environmental coverages),\textsuperscript{16,17} such as satellite-derived data and interpolated field measurements. It generates presence/absence predictions on the basis of a set of heterogeneous rules (rule-set) derived from a series of rule types in an iterative process. GARP uses four specific IF/\textsc{then} rule types in model development: 1) atomic rules, where predicted locations are defined by a specific environmental variable (e.g., \textsc{if} temperature = [22°C] \textsc{and} precipitation = [380 mm] \textsc{then} species = present/absent); 2) range rules, where predicted locations are defined by a range of variables (e.g., \textsc{if} temperature = [18–22°C] \textsc{and} precipitation = [350–380 mm]); 3) negated range rules, where prediction locations are defined as values outside of a defined range (e.g., \textsc{if} range not temperature = [18–22°C] \textsc{and} precipitation = [350–380 mm]); and 4) logit rules, where predicted locations are fit to a logistic regression model with the environmental variables.\textsuperscript{16} It can be considered a super-set of individual modeling approaches because range rules are similar to bioclimatic rules and other approaches use only logistic regression, and should have higher predictive accuracy than any single modeling approach.\textsuperscript{16}

GARP modeling is stochastic in nature, due in part to both the genetic algorithm for building models and the random partitioning of input locality data. Because of this characteristic, GARP can generate multiple solutions across multiple model runs. To evaluate this potential inter-model variation, it is critical to develop multiple models. Optimal models are those that compromise between omission (exclusion of known locations from the model) and commission (inclusion of areas with no known cases).\textsuperscript{18} DesktopGARP uses a best subset procedure to optimize model outputs by selecting models with user defined omission and commission thresholds. GARP outputs are rasterized coverages of the study area representing presence and absence pixels that can be manipulated in a geographic information system (GIS). These individual models can be summed to identify geographic areas where none, some, or all of the models predict presence or absence.\textsuperscript{19} The greater the number of models that agree, the more certainty there is in the prediction classification.\textsuperscript{20} Likewise, similarity across models indicates strong associations between the input data set and the environmental coverages. A number of studies have confirmed the usefulness, success, and applicability of GARP to a wide range of species across terrestrial\textsuperscript{21-24} and aquatic taxa.\textsuperscript{19,25,26} Additionally, a number of studies have applied the GARP modeling approach to disease systems.\textsuperscript{15,20,27-30}

\textbf{Input data.} A GIS database of specific anthrax outbreak localities within the 48 contiguous states was developed from a variety of data sources to represent the distribution of \textit{B. anthracis} for modeling. These data were assimilated during 2000–2005, with the exception of a 1957 outbreak report that could be mapped at the point level for Oklahoma and a 1968 outbreak for east central California. Point level maps were derived from coordinate pairs (latitude/longitude) collected by field personnel, address-matched using a GIS to farm front gates from diagnostic laboratory records, or heads-up digitized (capturing data by clicking the computer mouse on-screen at appropriate spatial locations) over high resolution satellite imagery using field reports and paper maps as guides to case locations. Point data representing both confirmed wildlife and livestock outbreaks were available from six states representing three regions of anthrax outbreaks for the contiguous 48 states: 1) the Dakotas Region (North Dakota, South Dakota, Minnesota), 2) the Southern Region (Oklahoma and Texas), and 3) the Western Region (Nevada, California; Figure 1). Table 1 summarizes the sample sizes and methods of data collection for each of the states used in this analysis.

A set of environmental coverages was constructed from publicly available satellite-derived climatic and biophysical parameters. Nineteen variables were downloaded from the WorldClim data set representative of various temperature and precipitation measurements (www.worldclim.org).\textsuperscript{31} Thirteen additional environmental variables, including temperature and vegetation measures (mean normalized difference vegetation index [NDVI]), were provided by the Trypanosomiasis and Land Use in Africa (TALA) research group at Oxford University (Oxford, United Kingdom).\textsuperscript{32} Two continuous soil parameters from the state soil geographic (STATSGO) data set (soil moisture, soil pH; www.nrcs.usda.gov/products/datasets/statsgo) were used to incorporate measures of known ecologic factors that impact or promote spore survival. Both soils variables were rasterized for inclusion in the ENM. All environmental coverages were re-sampled to 0.10 degree$^2$ (~8 x 8 km) and clipped to the boundary of the 48 contiguous states. All data sets were prepared using ERDAS Imagine version 8.7 (Leica Geosystems, St. Gallen, Switzerland), ArcGIS 9.0, and ArcView 3.2a (Environmental Systems Research Institute, Redlands, CA).
All coverages were subjected to a culling procedure prior to inclusion in modeling, first eliminating variables that represented similar parameters from the two data sources. Combinations of coverages were then evaluated using a jackknife procedure in DG (N-1 variables are used to build models iteratively until all N-1 combinations of variables have been used). The jackknife procedure is useful for eliminating variables that lead to over-fitting. To evaluate jackknife results a correlation matrix was derived from a set of models using the N-1 procedure and the measure of omission error for a 20-model set. Environmental variables were excluded from the final variable set if they increased omission between model outputs. A combination of jackknife evaluations and systematic model development and omission evaluation lead to the selection of the environmental coverage set. A final coverage set of six environmental variables was used that captured temperature, precipitation, elevation, soil moisture and pH, and mean NDVI (Table 2).

**Model building and evaluation.** For this study, 177 spatially unique anthrax outbreak locations were available for model building (Table 1). Prior to initiating DG, a randomly selected, independent hold-out sample of ~25% (n = 47) of the original data was withheld for later calculation of accuracy metrics. The remaining ~75% of the data (n = 130) were used for model building. A training/testing partition (50% and 50%, respectively) internal to DG was used for model building. To maximize DG performance, model runs were set to a maximum of 1,000 models and the best subset procedure was used to select the 20 best models under a 10% hard omission threshold and a 50% commission threshold for a final 10-model best subset. The final 10 models were summed within the GIS to visualize the geographic areas of presence/absence predicted across the best subsets.

An area under the curve (AUC) in a receiver operating characteristic (ROC) analysis was used to evaluate the predictive performance of the 10 best model subset using measures of specificity (absence of commission error) and sensitivity (absence of omission error) following other GARP studies. The ROC analysis is a threshold independent assessment of model quality derived from a plot of sensitivity (true positive rate; y-axis) versus 1 - specificity (error or true negative rate; x-axis) constructed from the best subset to de-

---

**FIGURE 1.** Distribution of the 177 spatially unique anthrax localities used for genetic algorithm for rule-set prediction model building. Print: Black dots indicate the data used for model building (n = 130) and gray dots indicate the independent hold-out sample used for model validation (n = 47). Online: Yellow dots indicate the data used for model building (n = 130) and green dots indicate the independent hold-out sample used for model validation (n = 47). This figure appears in color at www.ajtmh.org.
termine if models are predicting better than random.\textsuperscript{35,36} Likewise, AUCs, as used here, are based on all pixels of presence and all pixels of absence. The AUC of a given model set is compared with the AUC of a random prediction using a z-test. Successful models have AUC scores approaching 1.0 (a perfect model or a measure of reality), the higher the AUC the better the model is predicting presence/absence. Models predicting no better than random will have an AUC approaching 0.5.\textsuperscript{25} The ROC was derived from the 25\% independent test data points withheld from the original GARP model building data sets.\textsuperscript{26}

Two measures of omission were calculated from the 10-best model subset and the independent test data.\textsuperscript{26} First, total omission was calculated as the total number of independent test points predicted absent by the summated grid of all 10 best models. Second, an average omission was calculated as the average omission across each of the 10 best models. Omission indices are useful for evaluating the success of GARP at predicting known localities not included in model building. Two commission indices were also developed. First, total commission was calculated as the total number of pixels predicted present across all ten models divided by the total number of pixels in the study area. Second, an average commission was calculated as the average of the total number of cells predicted present divided by the total number of pixels within the study area on a model-by-model basis for each of the 10 models in the best subset. Little difference between these two measures indicates little variation in the rule-sets across the models; whereas a large difference indicates high variation across the models.

\textbf{RESULTS}

The modeling process reached convergence of accuracy (0.01) prior to the maximum iteration setting of 1,000 models. The AUC score from the ROC analysis was 0.7916 and was significantly different from a line of no information ($P < 0.01$). Average omission was 23.2\% and total omission was 6.8\%. The geographic prediction for \textit{B. anthracis} in the United States is shown in Figure 2 as the summation (or overlap) of the 10-model best subset. Table 3 summarizes the accuracy metrics and AUC score for this analysis.

The predicted distribution of \textit{B. anthracis} is primarily restricted to a narrow corridor from southwest Texas northward through western Oklahoma, central Kansas, central Nebraska, and into the Dakotas and Minnesota (Figure 2). This north/south corridor expands eastward in South Dakota into western Minnesota. From North Dakota, the distribution is predicted westward through Montana into Idaho, especially along the Snake River drainage basin merging with the Columbia River drainage into central Oregon and Washington. There are also small areas predicted in eastern Michigan and northwestern Ohio along the shorelines of Lake Huron and Lake Erie, respectively. \textit{Bacillus anthracis} is also predicted to occur from western Texas across southern New Mexico into Arizona. The southwestern half of Arizona was also identified as suitable as well as western California and a narrow strip of the state east of the central valley. The distribution is patchy in Nevada, but predicted along the central California/Nevada border and appears to be a continuation south from the Snake River drainage.

A review of the individual rule-sets (logic strings of atomic, range, negated range, and logit regression rules) developed during the modeling process indicate that narrow envelopes of mean NDVI, precipitation, and elevation define the ecological niche for \textit{B. anthracis} in this analysis. Rule sets from this modeling process were dominated by range rules (rules defining presence or absence based on the occurrence of points within a specific range of values from a given variable). Table 4 summarizes the primary presence-predicting rules from a single model experiment to demonstrate the biologic information contained within the GARP model output.

\begin{table}[h]
\centering
\caption{Environmental coverages used to develop the genetic algorithm for rule-set prediction model*}
\begin{tabular}{ll}
\hline
Environmental variable & Reference or Source \\
\hline
Mean annual temperature (°C) & 31 \\
Annual precipitation (mm) & 31 \\
Elevation (meters above mean sea level) & 32 \\
Soil moisture (lowest liquid limit as % of weight) & STATGO U.S. soil database \\
Soil pH (lowest reaction, no units) & STATGO U.S. soil database \\
Mean NDVI (no units) & 32 \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Specific locality data used to develop genetic algorithm for rule-set prediction (GARP) models*}
\begin{tabular}{|l|l|l|l|l|}
\hline
Outbreak location & Outbreak date & Resolution & No. & GARP no.† \\
\hline
\textbf{Dakotas region} & & & & \\
Minnesota‡ & 2000–2005 & Farm location & 60 & 23 \\
North Dakota§ & 2005 & Farm location & 88 & 54 \\
South Dakota¶ & 2005 & Farm location & 49 & 40 \\
\textbf{Southern region} & & & & \\
Oklahoma† & 1957 & Pasture locations & 21 & 10 \\
Texas** & 2001–2005 & Carcass locations & 122 & 28 \\
\textbf{Western Region} & & & & \\
Nevada†† & 2002 & Carcass locations & 32 & 2 \\
California‡‡ & 1968, 1984, 1991 & Farm locations & 42 & 20 \\
\hline
\end{tabular}
\end{table}

* GPS = global positioning system.
† Sample size of spatially unique anthrax localities in each state.
‡ Minnesota Board of Animal Health.
§ North Dakota State University Veterinary Diagnostic Laboratory.
¶ South Dakota State University Agriculture Extension Office and Geographic Information System Center for Excellence.
† Oklahoma Department of Agriculture.
** Centers for Disease Control and Prevention and Louisiana State University-World Health Organization Collaborating Center field investigations.

* STATSGO = state soil geographic; NDVI = normalized difference vegetation index.
DISCUSSION

This study presents the first estimate of the geographic potential for *B. anthracis* in the contiguous United States on the basis of the modeled ecological niche. According to the models, there is an important north-south corridor from west Texas up to the Canadian border and two east-west corridors traversing Arizona and Montana into California and Idaho/Washington, respectively. Although the specifics of disease introduction into the United States are unclear,\(^\text{10}\) it is hypothesized that *B. anthracis* was imported into Mexico by the Spanish and by colonial activities eastwards into what is now Texas and Louisiana\(^\text{10}\) and then disseminated across the central and western states through the movement of livestock via cattle trails and railways. The historical cattle trails and railways for the United States have been mapped (Hugh-Jones ME, unpublished data), and this mapping has confirmed that routes of cattle movement most likely traversed the regions of high model agreement from Texas up through the Dakotas and west through Arizona and Montana into the west and Pacific Northwest, which suggests that animals were moved through areas that could support long-term spore survival. Additional genetic evidence from a global analysis of the *B. anthracis* genome suggests a close relationship between a North American sub-lineage and the dominant European sub-group, which supports a European introduction of the species.\(^\text{38}\)

**Table 3**  
Model sample sizes and accuracy metrics for genetic algorithm for rule-set prediction model development and validation*  

<table>
<thead>
<tr>
<th>Metric</th>
<th>Model specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>N to build models</td>
<td>130†</td>
</tr>
<tr>
<td>N to test models (independent)</td>
<td>47</td>
</tr>
<tr>
<td>Total omission</td>
<td>6.8%</td>
</tr>
<tr>
<td>Average omission</td>
<td>23.2%</td>
</tr>
<tr>
<td>Total commission</td>
<td>66.5%</td>
</tr>
<tr>
<td>Average commission</td>
<td>41.6%</td>
</tr>
<tr>
<td>AUC</td>
<td>0.7916‡§</td>
</tr>
</tbody>
</table>

* AUC = area under curve.  
† N was divided into 50% training/50% testing at each model iteration.  
‡ z = 10.503 (P < 0.01).  
§ SE = 0.0394.
Previous studies have also indicated that industry-related outbreaks, those associated with the processing of imported, contaminated hides, hair, and wool and resulting subsequent downstream livestock outbreaks that lead to production of contaminated bone meal, were concentrated in the New England area and across states east of the Mississippi River. However, none of these eastern states have reported cases associated with naturally occurring spores in the past several decades. This suggests that the close of these industries lead to a decrease in industrial outbreaks. Additionally, this suggests that once these sources of spores were removed, the environments of these areas were unable to sustain long-term spore survival. This is supported by the ecological niche models because little area is predicted as suitable north of the Ohio or east of the Mississippi River drainages. Although this may in part be caused by a lack of locality data from the eastern states, the overall geographic extent is predicted to occur from central Texas north and west. The limited areas of predicted distribution in Michigan and Ohio are consistent with soil maps produced in the 1950s to delineate potential outbreak areas based solely on soil conditions, but the GARP models are more restrictive in the geographic extent of area predicted in both states. Likewise, the same soil-based map suggests that Arizona and New Mexico may contain adequate soil pH to maintain spores (Figure 1), but lack adequate soil moisture. The GARP models predict parts of both states with high model agreement. These differences may be caused by the additional environmental signals captured with improved environmental data, improved techniques in deriving soil measurements, and the more sophisticated analyses provided here. Although anthrax has not been reported in Arizona, there have been historical cases in New Mexico.

The accuracy metrics (Table 3) show that these models successfully predicted 76.8% of the independent test points (average omission = 23.2%). Apart from successfully predicting the validation data, the models predict well in regions from which no explicit validation data were included. For example, the northwestern corner of Montana, which had anthrax cases on at least two ranches in 2005 (specific localities not available for modeling) was predicted present by the models. Likewise, cases reported near Billings, Montana in 1999 also occurred within the areas of greatest model agreement. Ozona and Sonora, Texas also reported severe outbreaks (estimates > 1,000 deer) in the summer of 2005, and both of these areas are within the areas of highest model agreement (Figure 2). As for those points in the validation data set that the models seemingly failed to correctly predict, closer examination shows that they were randomly selected from the 1957 eastern Oklahoma breakout. This outbreak occurred in a region of eastern Oklahoma that has reported little to no anthrax since 1957. One study described the event in detail and suggested that the environmental conditions for the region of the state where cases occurred should not normally sustain B. anthracis spores. The study suggested that anthropogenic activities around the time of the outbreak, such as limestone mining efforts and road building, lead to a temporary surface soil environment that could support spores. Although it is difficult to confirm that the 1957 outbreak was caused by these superficial environmental conditions, or potentially undocumented industrial feed contamination common at the time, we had no expectation that the models would predict this area. This was confirmed by a lack of prediction in the eastern part of the state across the best model subset. When those independent validation points from Oklahoma were removed from the calculation of average omission, only 16.7% of the independent validation points were omitted. Likewise, total omission improved from 6.8% to 0% because all independent points are predicted by at least one of the 10 best subset models. This indicates that the GARP modeling process can be robust to the inclusion of small amounts of outlier data. This is important because it is not always apparent a priori if a given outbreak is derived from naturally occurring populations or was artificial, as may

<table>
<thead>
<tr>
<th>Task 9</th>
<th>1 range rule</th>
<th>4 logit rule</th>
<th>6 range rule</th>
<th>12 negated range rule</th>
<th>13 negated range rule</th>
<th>14 range rule</th>
<th>16 range rule</th>
<th>16 range rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>If precipitation = (219.69, 1,078.77) and elevation = (221.29, 682.60) and soil pH 1 = (0.00, 7.90) and NDVI = (0.24, 0.38) Then sp = presence</td>
<td>If temperature = 0.0039 – precipitation × 0.0156 – elevation × 0.0078 – soil pH 1 × 0.0000 + NDVI × 0.0117 Then sp = presence</td>
<td>If precipitation = (117.11, 2,784.11) and elevation = (778.71, 3,777.24) and soil moisture = (0.00, 0.90) and NDVI = (0.01, 0.65) Then sp = absence</td>
<td>If not temperature = (2.90, 23.20) and elevation = (144.40, 2,277.97) and NDVI = (−0.99, 0.70) Then sp = absence</td>
<td>If not temperature = (0.32, 22.20) and precipitation = (53.00, 2,271.23) and elevation = (−9.37, 2,969.94) and soil moisture = (0.00, 0.90) and NDVI = (0.12, 0.55) Then sp = absence</td>
<td>If temperature = (1.21, 21.42) and precipitation = (1,091.59, 2,335.34) and elevation = (−566.79, 3,815.68) and soil moisture = (0.00, 0.90) and soil pH 1 = (0.00, 9.04) and NDVI = (−0.92, 0.69) Then sp = absence</td>
<td>If temperature = (2.32, 22.09) and elevation = (221.29, 1,432.23) and soil pH 1 = (0.00, 7.90) and NDVI = (0.27, 0.46) Then sp = presence</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 4</th>
<th>Primary presence rules (logic strings) from the output of a single model from best model subset*</th>
</tr>
</thead>
<tbody>
<tr>
<td>* NDVI = normalized difference vegetation index.</td>
<td></td>
</tr>
</tbody>
</table>
be the case in Oklahoma. The models were re-run excluding the Oklahoma localities included in the initial GARP calculation. An evaluation of the best subset showed that the overall geographic extent of the predicted distribution of anthrax did not change.

The accuracy of the predicted distribution of B. anthracis and the distribution of recent outbreaks within the predicted distribution indicates that B. anthracis has established a natural ecology in North America, despite the introduction of vaccination efforts to control the disease (and therefore potentially restrict the survival of B. anthracis). Although there are vaccination efforts in livestock and some farmed wildlife, the disease is still prevalent over a large geographic area in both groups. In general, anthrax vaccination has become a reactionary practice in the United States used in response to outbreaks once they have begun. Integrating modeling approaches such as these can improve our understanding of where vaccination should be a priority before, during, and after outbreak events. Additionally, the location of outbreaks in free ranging wildlife within the predicted distribution suggests a mechanism for maintaining the disease in unvaccinated wildlife populations, even in areas where livestock herds may be vaccinated. In this way, ENM is useful for identifying areas where wildlife populations should be monitored for the disease more regularly, and possible post-outbreak control such as carcass disposal should be used (e.g., western Texas, the eastern Dakotas, and northwestern Minnesota).

Although our results are presented at the continental scale, with relatively large pixel size (8 × 8 km), GARP is scalable and not limited to broad-scale data. Investigating local scale niche parameters, perhaps on a state-by-state basis using higher resolution data (i.e., with the exception of the STATSGO soils data, all of the parameters discussed in this paper are available at ~1.2 × 1.2 km) is possible. Additionally, more specific filtering techniques may be useful for subdividing the predicted distribution into areas that may or may not currently support B. anthracis transmission cycles or animal/spore interactions. For example, future work is planned to assess the usefulness of using land cover/land use data to determine areas within the predicted distribution that do not support livestock or wildlife populations. However, even at lower resolutions, these models provide a geographically explicit understanding of the natural ecology of B. anthracis, and should help shape future studies and control measures for anthrax. Future modeling efforts should also incorporate data from genetic studies to develop individual model experiments for each strain in the United States. Such an analysis could be useful to determine if clade or strain-specific ecological affinities can be found to improve the niche definition and enhance predictive power. One other local-level study from Africa has provided evidence for strain-specific differences in physiologic tolerances for certain soil conditions and associated differences in those strains’ geographic distributions. Ecological niche modeling, when coupled with genetic methods, provides a tool for determining continental-scale differences in ecologic affinities and corresponding geographies for diverse genetic lineages of B. anthracis.

These results also provide a preliminary tool that could be useful for discriminating between endemic/naturally occurring outbreaks, contaminated feed or industrial-related outbreaks, and potential bioterrorist attacks. Although continued efforts to refine and validate these models are certainly required, outbreaks occurring outside of predicted geographic regions (unsuitable areas for spore survival) may indicate the intentional release of spores or target epidemiologists to investigate sources of contamination, such as feed. When modeling and genetic analysis are combined, there is potential for identifying areas of risk for subsequent natural outbreaks and improving trace back analyses for epidemiologists or forensic scientists in the case of intentional release.

Received July 5, 2007. Accepted for publication August 24, 2007.

Acknowledgments: We thank R. Scachetti-Pereira for developing the DesktopGARP implementation of the GARP algorithm; N. Dyer, R. Daly, M. Wimmerly, and L. Glaser for providing outbreak data; D. Wilkund for providing data from the Dakotas and Minnesota; and D. Rogers and S. Hay for developing and providing the TALA environmental data set. We also thank the editor of the American Journal of Tropical Medicine and Hygiene and two anonymous reviewers for strengthening this manuscript. The Texas data set was a combination of data from Louisiana State University field investigations from 2000–2005 and the Centers for Disease Control and Prevention outbreak response in 2001.

Financial support: Texas field data was supported by the RV Ranch and staff. Other support was provided by Louisiana State University.

Authors’ addresses: Jason K. Blackburn, Spatial Epidemiology and Ecology Research Laboratory, Department of Geography, California State University, Fullerton, 800 North St. College Drive, Fullerton, CA 92834. Kristina M. McNyset, Environmental Protection Agency, Office of Research and Development/Western Ecology Division, Corvallis, OR 97333. Andrew Curtis, Department of Geography, College of Letters, Arts, and Sciences, University of Southern California, Kaprielian Hall, Room 416, Los Angeles, CA 90089. Martin E. Hugh-Jones, Department of Environmental Studies, School of the Coast and Environment, Louisiana State University, Baton Rouge, LA 70803.

REFERENCES