

# Potential Distribution of Six North American Higher-Attine Fungus-Farming Ant (Hymenoptera: Formicidae) Species

Sarah F. Senula,<sup>1,4,\*</sup> Joseph T. Scavetta,<sup>3,\*</sup> Joshua A. Banta,<sup>1,6</sup> Ulrich G. Mueller,<sup>2</sup> Jon N. Seal,<sup>1</sup> and Katrin Kellner<sup>1</sup>

<sup>1</sup>Department of Biology, University of Texas at Tyler, Tyler, TX 75799, USA, <sup>2</sup>Section of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA, <sup>3</sup>Department of Computer Science, Rowan University, Glassboro, NJ 08028, USA, and <sup>4</sup>Corresponding author, e-mail: [sfsenula@gmail.com](mailto:sfsenula@gmail.com)

\*These authors contributed equally.

Subject Editor: Sunil Kumar

Received 3 December 2018; Editorial decision 18 November 2019

## Abstract

Ants are among the most successful insects in Earth's evolutionary history. However, there is a lack of knowledge regarding range-limiting factors that may influence their distribution. The goal of this study was to describe the environmental factors (climate and soil types) that likely impact the ranges of five out of the eight most abundant *Trachymyrmex* species and the most abundant *Mycetomoellerius* species in the United States. Important environmental factors may allow us to better understand each species' evolutionary history. We generated habitat suitability maps using MaxEnt for each species and identified associated most important environmental variables. We quantified niche overlap between species and evaluated possible congruence in species distribution. In all but one model, climate variables were more important than soil variables. The distribution of *M. turrifex* (Wheeler, W.M., 1903) was predicted by temperature, specifically annual mean temperature (BIO1), *T. arizonensis* (Wheeler, W.M., 1907), *T. carinatus*, and *T. smithi* Buren, 1944 were predicted by precipitation seasonality (BIO15), *T. septentrionalis* (McCook, 1881) were predicted by precipitation of coldest quarter (BIO19), and *T. desertorum* (Wheeler, W.M., 1911) was predicted by annual flood frequency. Out of 15 possible pair-wise comparisons between each species' distributions, only one was statistically indistinguishable (*T. desertorum* vs *T. septentrionalis*). All other species distribution comparisons show significant differences between species. These models support the hypothesis that climate is a limiting factor in each species distribution and that these species have adapted to temperatures and water availability differently.

**Key words:** MaxEnt, attine, Texas, ecological niche modeling, temperature

Insects are the most abundant and diverse group of terrestrial animals on the planet, with ants (Formicidae) being one of the most successful in Earth's evolutionary history (Hölldobler and Wilson 1990, Ward 2014). With over 16,000 ant species spread throughout diverse ecological niches (Bolton 2016), it has been suggested that their symbiotic relationships with microorganisms may have been a major cause of their radiation and success (Akman Gunduz and Douglas 2008, Russell et al. 2009, Douglas 2015, Hu et al. 2018). While ants are among the most abundant and diverse group of insects, there is a lack of range-limiting data and readily available distribution surveys (Diniz-Filho 2010, Simões-Gomes et al. 2017), especially in the south-eastern United States (Tschinkel et al. 2012, King et al. 2013, Noss et al. 2015). To address this relative paucity of available data, there have been recent attempts to use distribution modeling to determine past, present, and future species distributions (e.g., Lobo 2016). Species distribution models, or ecological

niche models, are created using available species occurrence data, in conjunction with environmental characteristics, such as climate or soil datasets from public databases. Using these data, the target function,  $f: X \rightarrow Y$ , can be approximated, where  $X$  is the set of environmental conditions at a given location and  $Y$  is the probability of occurrence at that location, by finding the best fit for the model. The approximate function can then be applied across the entire study area (which are mostly unsampled areas) to estimate how suitable all locations on the landscape are at a given grain or pixel size (resolution) (Peterson et al. 2011). This allows one to make a forecast, across a wide area, of where the species is favored based on the environmental characteristics of the landscape, even if the entire landscape has not been sampled densely. In this way, putative range maps can be deduced from limited sampling, and these preliminary maps can then serve as a springboard to target certain areas for future sampling to corroborate and refine the range maps (Marcer et al.

2013). Species distribution models can also give insight as to which environmental variables, used to create the model, are most influential in determining the range of a species, i.e., temperature, precipitation, or soil properties. In this way, researchers may be able to gain a better understanding of the evolutionary history of a species and may be able to predict how said species may be impacted by changing climate or other anthropogenic affects.

Fungus-farming ‘attine’ ants present a unique study system to investigate range expansions and distributions. Higher-attine ants (genera *Trachymyrmex*, *Mycetomoellerius*, *Paratrachymyrmex*, *Sericomyrmex*, *Acromyrmex*, and *Atta*) cultivate gardens of fungal monocultures (leptotaceous basidiomycetes in the family Agaricaceae) as their primary food source, while the fungal garden is protected, fed, and maintained in ideal environmental conditions by the ants (Weber 1972, Hölldobler and Wilson 1990, Ward et al. 2015). While there are basal attine lineages that cultivate fungi that have been found to be free-living, higher-attine ants cultivate fungi that are only found to be grown by ants, thus forming an obligate symbiotic relationship. The presence of such obligate symbiotic microorganisms may allow ants to consume food sources not previously digestible, thus allowing the species to be able to take advantage of niches that were previously uninhabitable (Aylward et al. 2012, Brune 2014, Oliver and Martinez 2014, DeMillo et al. 2017), which may have profound effects on the fitness, adaptation, and range distribution (Douglas 2010, Engel and Moran 2013, Russel et al. 2016, Muhammad et al. 2017). Attine ants are found only in the New World and are thought to have evolved about 55–65 mya (Mueller et al. 2005, Schultz and Brady 2008, Ward et al. 2015, Branstetter et al. 2017) in South America, then expanded Northward across Central America to North America (Rabeling 2007, Mueller et al. 2017). The environmental factors that determine the distributions of attine ants may be unlike those in other ant species distributions, as attine distributions depend on the environmental needs of both the ant host and their fungal symbionts.

In the present study, our goal was to investigate the possible drivers of the ecological distributions of six North American higher-attine nonleaf-cutting ant species from the genera *Trachymyrmex* and *Mycetomoellerius*: *T. arizonensis* (Wheeler, W.M., 1907), *T. carinatus*, *T. desertorum* (Wheeler, W.M., 1911), *T. septentrionalis* (McCook, 1881), *T. smithi* Buren, 1944, and *M. turrifex* (Wheeler, W.M., 1903). *Trachymyrmex* is the most species-rich genus of higher-attine ants in North America, with a total of seven species found in the conterminous United States. *Mycetomoellerius* is a newly classified genus consisting of former *T. turrifex* and *T. jamaicensis*, which are the only species found in the United States, as most species are found in the New World tropics (Solomon et al. 2019). *Trachymyrmex* is also primarily a tropical genus, but in the United States most of the species are found in the arid southwestern states of Arizona, New Mexico, and Texas. It is hypothesized that North American *Trachymyrmex* species were originally adapted to survive in dry, arid environments (Seal and Tschinkel 2006, 2010, Rabeling et al. 2007, Branstetter et al. 2017). *T. septentrionalis*, however, has a distribution that extends northward into the temperate zone and thus lives in wetter, cooler climates such as central Illinois and Long Island, New York (approximately 40° N) (Rabeling et al. 2007, Seal et al. 2015).

Soil and climate are known to be important environmental variables in determining the distribution of ant species (Diehl-Fleig and Rocha 1998, Cardoso and Cristiano 2010, Cardoso et al. 2010, Meyer et al. 2011) with temperature, rainfall, and humidity affecting abundance and distribution the most (Seal and Tschinkel 2010, Savopolou-Soultani et al. 2012). We were interested in whether soil

and/or climate influences the distribution of these fungal-gardening species. These species are found in rocky and hard soil (Arizona and western Texas), clay to sandy soils (central and east Texas) and pure sandy soils (east Texas and along the entire Southeastern Coastal Plain). For example, two southeastern species *M. turrifex* and *T. septentrionalis* are thought to prefer different soil types. *M. turrifex* occurs mainly in clay soils whereas *T. septentrionalis* is almost exclusively found in sandy soils (Seal and Tschinkel 2006, Rabeling et al. 2007). However, the location of sand, clay, and rocky soils is not evenly distributed in the southern United States; where these ants occur, sand occurs primarily along the coastal plains, clay further inland and rocky soils in arid mountains of southwestern North America (Noss et al. 2015). Southern North America is characterized by a profound rainfall gradient that range from true deserts in the southeast to subtropical rainforest-like conditions in the southeast (Soltis et al. 2006, Noss et al. 2015, Seal et al. 2015, Chapman and Bolen 2018). Thus, climate and the distribution of soil types may confound each other. By modeling these two broad sets of possible drivers, it might be possible to determine which of these variable(s) might explain the differences in distribution of these species. Additionally, the subterranean nests of fungus-gardening ants are thought to exert ecological impacts because the ants move and displace soil nutrients (Tschinkel and Seal 2016, Swanson et al. 2019). While these ants are likely important in building soil-based ecosystems, we lack good models that could determine where they are found. Thus, our understanding of ecosystem ecology could be hindered by not knowing which ants might be found in which ecosystems. The goal of this study was to determine and describe the environmental factors that likely explain the range distributions of each species and compare the distributions among them.

## Materials and Methods

### Study Design and Modeling

MaxEnt modeling is a useful method for creating species distribution models, because it requires only the locations of known occurrences for a species (in the form of global positioning system coordinates) and environmental data, often available from public repositories (Phillips et al. 2006). The number of locations of known occurrence for a species can be quite small and still be used to make species distribution models covering a very large area, though this method is not without potential statistical artifacts (van Proosdij et al. 2016). This allows one to make forecasts of the probable areas of occurrence of a species based on very limited information, as is the case when working with endangered species living in fractured landscapes, where obtaining landowner permissions for surveys is difficult (Marcer et al. 2013). MaxEnt is used to estimate a species’ probability of presence in a given area by creating a raster map, where each pixel contains an estimation of the relative habitat suitability (ranging from 0, unsuitable, to 1, highly suitable) for the modeled species. A score will be higher when the environmental variables assigned to that pixel are more similar to those where the species is known to occur (Phillips and Dudik 2008). Species distribution models are typically used to model the distributions of one species at a time, though in the present study, each model represents two species simultaneously: the ant species and its obligate symbiont, their fungal garden.

We performed MaxEnt species distribution modeling on five out of the seven *Trachymyrmex* species and one out of the two *Mycetomoellerius* species that are found within the conterminous United States: *T. arizonensis*, *T. carinatus*, *T. desertorum*, *T. septentrionalis*, *T. smithi*, and *M. turrifex*. Three species were not included in this study: *M. jamaicensis*, *T. nogalensis*, and

*T. pomonae* Rabeling & Cover, 2007 were excluded because of sparse collection records and presumed limited distributions in the United States (Rabeling et al. 2007). *T. arizonensis*, *T. carinatus*, *T. desertorum*, and *T. smithi* are broadly sympatric with one another. *M. turrifex* and *T. septentrionalis* are sympatric in certain areas of their ranges, i.e., Texas and Oklahoma. Species occurrence records were obtained from published and unpublished data from collections on private and public land within the known ranges of each species (Fig. 1, See Supp Table 4 [online only] for data sources). The species varied greatly in the number of known locations of occurrence (see Table 1). This study investigated conterminous United States *Trachymyrmex* and *Mycetomoellerius* species distributions using average climatic variables from the years 1970–2000 (WorldClim) (Hijmans et al. 2005) and soil variables. We obtained soil data from the State Soil Geographic (STATSGO) Data Base (United States Department of Agriculture 1995a), and the data processing steps to create this dataset can be found from Wolock (1997). The STATSGO dataset was captured as 1:250,000 scale USGS topographic quadrangle units by generalizing soil survey maps (United States Department of Agriculture 1995a,b, Mednick 2010); but since the soil survey maps were not always available at specific locations, the STATSGO dataset interpolates across these gaps based on broad physiographic characteristics (Mednick 2010). A single STATSGO map unit may contain up to 21 different component soils (USDA NRCS 1994). Environmental and species occurrence data were processed using GRASS GIS Version 7.2 (GRASS Development Team 2017). All rasters were resampled to a common resolution of 1,000 m × 1,000 m and projected into the North American Datum of 1983 horizontal datum reference system in an Albers Equal Area projection.

If any two continuous variables were found to be highly correlated with each other, according to a Pearson's moment correlation coefficient (Sokal and Rohlf 1995) of 0.75, then one of the two variables was removed from the dataset until no two variables remaining in

the dataset were highly correlated with each other. At each iteration, we removed the variable that correlated more with other variables, allowing us to retain the most unique predictors. This methodology allows for a quicker runtime when creating models without a loss of environmental information and can simplify interpretation of the results (Elith and Leathwick 2009). Eight out of nineteen climatic variables and nine out of seventeen soil variables were incorporated into the model for each species: annual mean temperature (BIO1), isothermality (BIO3), minimum temperature of the coldest month (BIO6), temperature annual range (BIO7), mean temperature of wettest quarter (BIO8), precipitation seasonality (BIO15), precipitation of warmest quarter (BIO18), precipitation of coldest quarter (BIO19), available water capacity (inches per inch), annual flood frequency, calcium carbonate in soil layer (%), cation exchange capacity, share of map unit with hydric soils, erodibility, average depth of bedrock (inches), slope of map unit (%), and depth of soil (inches).

### Selecting Background Points and Correcting for Sampling Bias

To find the best function to predict a species occurrence, we require a set of features that occurs where the species is present, and a baseline set of features that occurs in the landscape of interest. Using this data, MaxEnt can find the feature distribution for the species that is closest to the baseline landscape distribution, while constraining the species distribution such that it closely resembles the feature averages found amongst the occurrence points (Elith et al. 2010). To create a feature distribution for the landscape we can randomly sample the background environment. The simplest approach is to randomly sample background points uniformly across a study extent; however, this approach may yield statistically flawed results if occurrence point sampling was not truly random (Kramer-Schadt et al. 2013). To combat bias in sampling, occurrence data can be filtered spatially allowing fewer points to be used within an area of

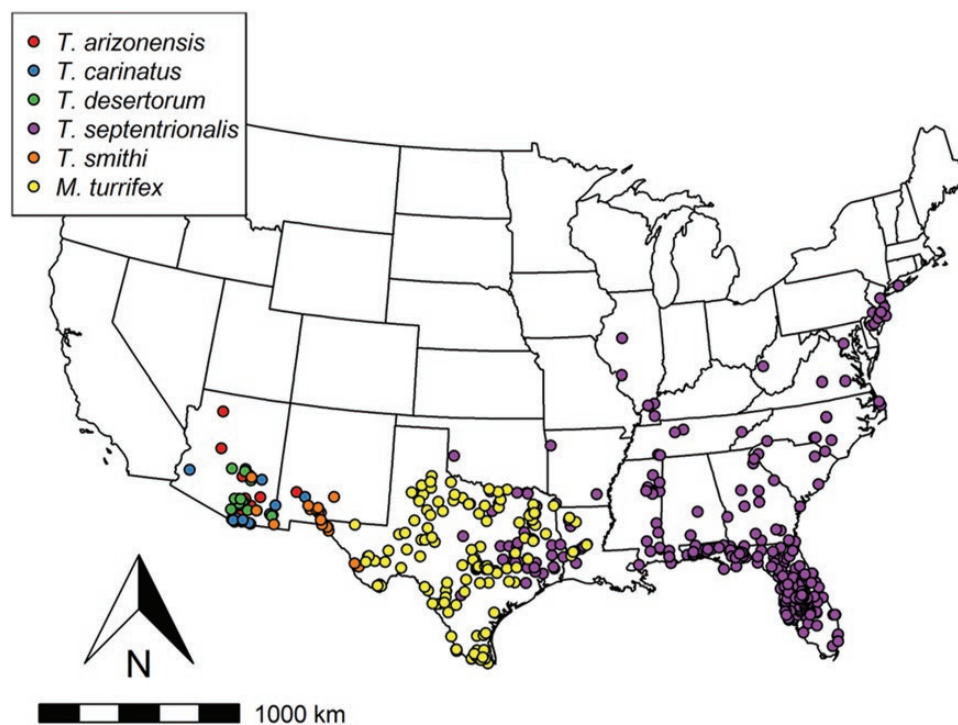


Fig. 1. Species occurrence points used to create MaxEnt models.



**Table 1.** Number of species occurrence points used to create distribution models

Species	No. of localities	Unique localities
<i>T. arizonensis</i>	88	40
<i>T. carinatus</i>	40	17
<i>T. desertorum</i>	21	12
<i>T. septentrionalis</i>	389	330
<i>T. smithi</i>	29	26
<i>M. turrifex</i>	174	147

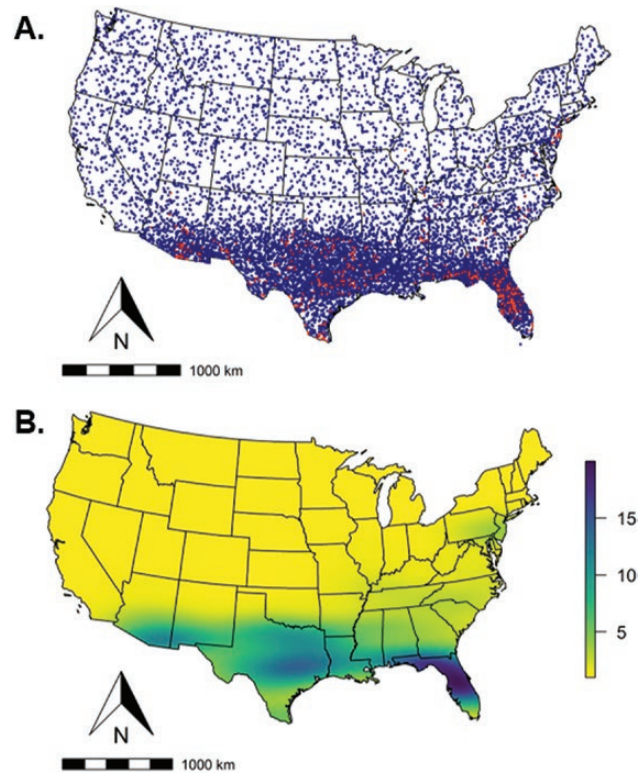
high sampling to better balance the overall sampling distribution. Though, when lacking a sufficient sample size, removing the points may not be possible. Instead, background sampling can be altered to better represent the sampling bias found in the occurrence dataset. While this approach is often better than leaving unaccounted spatial biasing, it may introduce weaker predictions (Kramer-Schadt et al. 2013). Because of this, we present models both with and without spatial bias accounted for (see supplemental figures for models with spatial bias).

To obtain background points in models that do not account for spatial biasing, we simply select points across the United States extent uniformly (Supp Fig. 1 [online only]). In models that account for spatial biasing, we adjust our background point sampling so that it better represents the biasing in the occurrence sampling efforts (Fig. 2). Because it is unlikely that one fungal-gardening species would be missed while surveying for the other species, we assume that the spatial biases are the same for all species. Combining all species occurrence data, we generate a probability density using an axis-aligned bivariate normal kernel and scale the density values from 1 to 20, similar to previous studies (Elith et al. 2010, Fourcade et al. 2014). A set of 10,000 background points was obtained once for models created with uniformly sampled background points, and once for models created with background points sampled from the biased probability distribution (Fig. 2 and see Supp Fig. 1 [online only] for uniformly sampled background points).

### Model Hyperparameter Optimization

The MaxEnt algorithm relies on settings, or hyperparameters, that must be set before models are trained, and the values for these settings may affect a model's performance greatly depending on the dataset; however, the right values are for a given task is unknown and often hard to estimate (Muscarella et al. 2014). Specifically, the permitted feature types and the beta regularization multiplier ( $\beta_M$ ) are important settings to adjust. Permitted features refer to the transformations MaxEnt can employ in the model function, for example, whether to use quadratic terms. Regularization adds a cost to overly complex models, as they tend to fail in generalizing to new data; the use of  $\beta_M$  controls how much of a cost is employed. To achieve models with the greatest predictive power, we specifically tuned those hyperparameters. In addition, we set add samples to background as false; all other parameters were left as default.

To find the optimal values for both the permitted features and  $\beta_M$ , we ran many models for each species separately for both background sets. We chose to test five sets of permitted features: {linear}, {linear, quadratic}, {linear, quadratic, hinge}, {linear, quadratic, hinge, product}, and {linear, quadratic, hinge, product, threshold}. For each permitted feature set, we combined 10 selected  $\beta_M$  values within the domain of [0, 2.5]; therefore, each final model came from a set of 50 models with varying settings. We selected  $\beta_M$  values



**Fig. 2.** Background point selection (blue) with occurrence points (orange) for all species. In total, 10,000 background points was sampled. (A) Background points were randomly selected from a generated probability density (see B) across the United States. (B) Probability density distribution generated from all species occurrence points using a normal kernel; probability density values were scaled from 1 to 20.

using the adaptive LIPO algorithm, which performs an informed random search across an unknown function, for more details on the algorithm, see Malherbe and Vayatis (2017). The adaptive LIPO approach is based on uniform sampling across the function domain, however, a Lipschitz constant is estimated such that the function value maximums can be determined before evaluation. This knowledge can be used to avoid running models at poor  $\beta_M$  inputs, such that a high performing setting can be found in less time. For specific implementation details, see the open-source R implementation (Scavetta 2019).

### Metrics for Model Performance

To determine how well a model performs, both for selecting the best models and for providing tuning information for hyperparameters optimization, mathematical metrics must be calculated. The results of a presence-background model can be summarized as the number of occurrence points accurately predicted as occurrence points (true positive, TP), the number of background points wrongly predicted as occurrence points (false positive, FP), the number of background points accurately predicted as such (true negative, TN), and the number of occurrence points wrongly predicted as background points (false negative, FN). As the output of the MaxEnt algorithm is a probability, these values will depend on the threshold that is set to separate occurrence from background. Using these results, various metrics can be calculated, giving an impression of how well the model performed.

Many metrics were calculated for our models to give an overview of performance. A common metric used is the area under the receiver

operating characteristic curve (area under curve; AUC), which measures the ratio between the true-positive rate (TPR) ( $TP/(TP + FN)$ ) and the false-positive rate (FPR) ( $FP/(FP + TN)$ ) at all thresholds (Peterson et al. 2011). While AUC can be representative of performance when true absences are present, in general it is a poor metric for presence-background studies as well as when modeling the potential distribution of a species (Jiménez-Valverde 2012). Another popular metric, the kappa statistic, compares the model's output to what would be expected by chance. Though kappa has been shown to be less likely to score a model overoptimistically (Fernandes et al. 2018), it has also been criticized for its sensitivity to prevalence, i.e., the amount of true occurrence points present in the model ( $TP + FN$ ) (Allouche et al. 2006). The kappa value selected for a model is the maximum across all thresholds.

To provide a metric with the advantages that kappa has, without the dependence on prevalence, the true skill statistic (TSS) can be used (Allouche et al. 2006). TSS measures the difference between the TPR and the FPR at a given threshold. While TSS can generally give a good representation of model performance, it may be susceptible to large study extents as a large value for any of the results (TP, FP, FN, and TN) can cause the statistic to prefer overpredictions (Wunderlich et al. 2019). Two additional metrics, the odds ratio skill score (ORSS) and the symmetric extremal dependence index (SEDI) can be used in place of TSS as they do not converge toward overpredictions; ORSS tends to predict performance better with true absences while SEDI tends to predict performance better when using background points as in this study (Wunderlich et al. 2019). To calculate SEDI, we must calculate the FPR as  $FP/(FP + TN)$  and the TPR as  $TP/(TP + FN)$ .

We then calculate SEDI as  $(\ln(FPR) - \ln(TPR) - \ln(1 - FPR) + \ln(1 - TPR))/(\ln(FPR) + \ln(TPR) + \ln(1 - FPR) + \ln(1 - TPR))$ . In the case that any of the confusion matrix elements (TP, FP, TN, and FN) were 0, this metric would become undefined. To account for this, we consider SEDI equal to 1 if  $FP + FN = 0$  as there are no false predictions. We consider SEDI equal to -1 if  $TP + TN = 0$  as there are no true predictions. We consider SEDI equal to 0 if  $TP + FP = 0$  or if  $TN + FN = 0$ , though this case could not happen if the test set has both presence and background samples. In all other cases, we set the element with 0 predictions as  $1e-05$  so that we can get a close approximation to the true SEDI score. We report AUC, TSS, and SEDI for comparison purposes, however, we applied SEDI as the main metric for model selection.

### Metrics for Model Validation and Evaluation

To avoid overinflation of evaluation metrics in our models resulting from learning the training data, we split our total data into two sets, training data and testing data. A model is trained (fitted) with the training data, while the metrics are calculated using the results of the testing data. To achieve a test–train split, we subdivided our data equally across four ( $k = 4$ ) spatially independent bins using the block method (Muscarella et al. 2014). We chose to use the block method as it performs the best in distinguishing between poor and good fits and it implicitly tests a model's ability to transfer to another region (Fourcade et al. 2018). Each bin can in turn act as test data, omitted from the data used to fit the model, while the other three bins act as training data used to fit the model. This can be repeated for each bin such that a total of four tests are performed over the entire dataset. All metrics reported are an average of all data partition evaluations, i.e., all bins are treated as test data and evaluated once for a given model.

While the metrics selected give an impression of model performance on the test data, it may also be worth knowing how well the

model is generalizing (to the test data) compared to how it performed on the training data. Rather than computing all metrics twice, once for training and once for test, we can use the omission rate at the minimum training presence ( $OR_{MTP}$ ). A value above zero shows that there was an occurrence point in the test set with lower suitability than the minimally suitable training occurrence point (Muscarella et al. 2014).

### Quantifying Importance of Environmental Variables

In addition to creating distribution models, we examined which variables were the most important in determining each species' distribution model. To determine the relative importance of each individual environmental variable to the models, the fit of each model was compared to reduced univariate models (Phillips 2006). If an environmental variable made up a substantial portion of the model fit when it was the only variable used, compared to when all environmental variables were used, then that particular variable was considered important in creating the model for that species (Phillips 2006). This was done for each species (Table 3). Response curves of the most important variable to each species in each model were also created.

Model fit was measured with the gain statistic. Gain is a likelihood (deviance) statistic that measures the model performance compared to a model that assigns equal habitat suitability to all areas of the landscape (Walters et al. 2017). Taking the exponent of the final gain gives the (mean) probability of the presence sample(s) compared to the pseudoabsences. For instance, a gain of 3 means that an average presence location has a habitat suitability of  $e^3 = 20.1$  times higher than an average pseudoabsence site (Walters et al. 2017).

### Comparisons of Distributions Between Species

We tested whether the ant species differed significantly in their habitat associations, which presumably reflects ecological differentiation among the ants. The habitat associations of a species are quantified in the species distribution models by the habitat suitability scores of each individual pixel (i.e., spatial grain), because the habitat suitability scores are functions of the environment across the landscape. The observed levels of differentiation in habitat suitability scores across the landscape for pairs of species were calculated using the *I* statistic (Warren et al. 2008), as this value has been shown to be highly correlated with other measures of niche similarity (Warren et al. 2008). To test significance of the *I* statistic, we 1) pooled the occurrence data of the two species and obtained random subsamples to create two new samples with the same amount of observations used to create the original distribution maps, 2) modeled the distributions of the subsampled datasets in MaxEnt using the best hyperparameters obtained from each species, 3) calculated the amount of overlap in the habitat suitability scores of the two subsampled datasets, and 4) repeated the above steps 100 times to generate a nonparametric distribution of the *I* statistics (Warren et al. 2010). Two species were considered to have significantly different habitat associations if the observed (nonpermuted) *I* statistic for those species was below the empirically derived 5% permuted distribution of *I* statistics, corresponding to a 5% likelihood that the observed differentiation in habitat associations among the two species was merely artifactual (Walters et al. 2017).

### Software

All models were created in the java implementation of MaxEnt version 3.4.1 (Phillips 2006). To perform hyperparameter tuning, calculate metrics, and calculate niche overlap, an R package was

created that is publicly available (Scavetta 2019). The created R package makes use of functions from the *dismo* (Hijmans et al. 2017), *raster* (Hijmans 2019), and *ENMeval* (Muscarella et al. 2014) packages.

## Results

### Model Validation

All final species models had a SEDI greater than  $0.87 \pm 0.01$  when evaluated using the block spatial partitioning method. In addition, all selected models have an AUC greater than  $0.86 \pm 0.02$ , except the *T. septentrionalis* model with an AUC of  $0.74 \pm 0.04$ . In comparison, the *T. arizonensis*, *T. carinatus*, *T. desertorum*, and *T. smithi* models had a TSS greater than  $0.81 \pm 0.05$ , while the *T. septentrionalis* and *M. turrifex* models had a TSS of  $0.48 \pm 0.05$  and  $0.67 \pm 0.04$ , respectively. The *T. septentrionalis* and *T. carinatus* models had  $OR_{MTP}$  scores close to 0, while the other models had scores ranging from  $0.11 \pm 0.04$  to  $0.16 \pm 0.07$  (Table 2).

### Areas of High Habitat Suitability

The models show areas of low to high habitat suitability in areas of the United States. The models of *T. arizonensis*, *T. carinatus*, *T. desertorum*, and *T. smithi* show high habitat suitability in the United States southwest: Arizona, New Mexico, and West Texas (Fig. 3). These areas include parts of the North American deserts, Southern semiarid highlands, and the temperate sierras (Omernik and Griffith 2014).

*T. septentrionalis* has high habitat suitability that extends from East Texas along the east coast into Long Island, NY (Fig. 3) throughout Eastern temperate forests, more specifically, the southeastern plains, Texas–Louisiana coastal plain, and the Mississippi alluvial and southeast coastal plains (Omernik and Griffith 2014). The area of high habitat suitability of *M. turrifex* ranges from Texas and slightly north into Oklahoma (Fig. 3) in south central semiarid prairies, Tamaulipas–Texas semiarid plains, and southeastern plains (Omernik and Griffith 2014). This distribution model has areas of high habitat suitability in Western Arizona and Southeastern California, but to our knowledge, this species is not known to occur in these areas.

### Quantifying Importance of Environmental Variables

In all but one model, climate variables have the most predictive value, mean annual temperature (BIO1) had the largest test gain for *M. turrifex*, precipitation seasonality (BIO15) had the largest test gain for *T. arizonensis*, *T. carinatus*, and *T. smithi*, and precipitation of warmest quarter (BIO19) for *T. septentrionalis*. *T. desertorum* was the only species where soil contributed the most to the performance of the model (annual flood frequency). This soil variable

though, arguably, has climatic implications. In most other cases, soil test gains were relatively low in comparison to climatic test gains.

Response curves for the variable with the highest test gain for each species are presented (Figs. 4 and S2 [online only]). Response curves show the relationship between species habitat suitability and a climate or soil variable.

### Comparison of the Individual United States Species Models

Out of 15 possible pair-wise comparisons between each species' distributions, only one was statistically indistinguishable (*T. desertorum* vs *T. septentrionalis*). All other species distribution comparisons show significant differences between species (see Table 4).

## Discussion

### Distributions of *Trachymyrmex* in North America

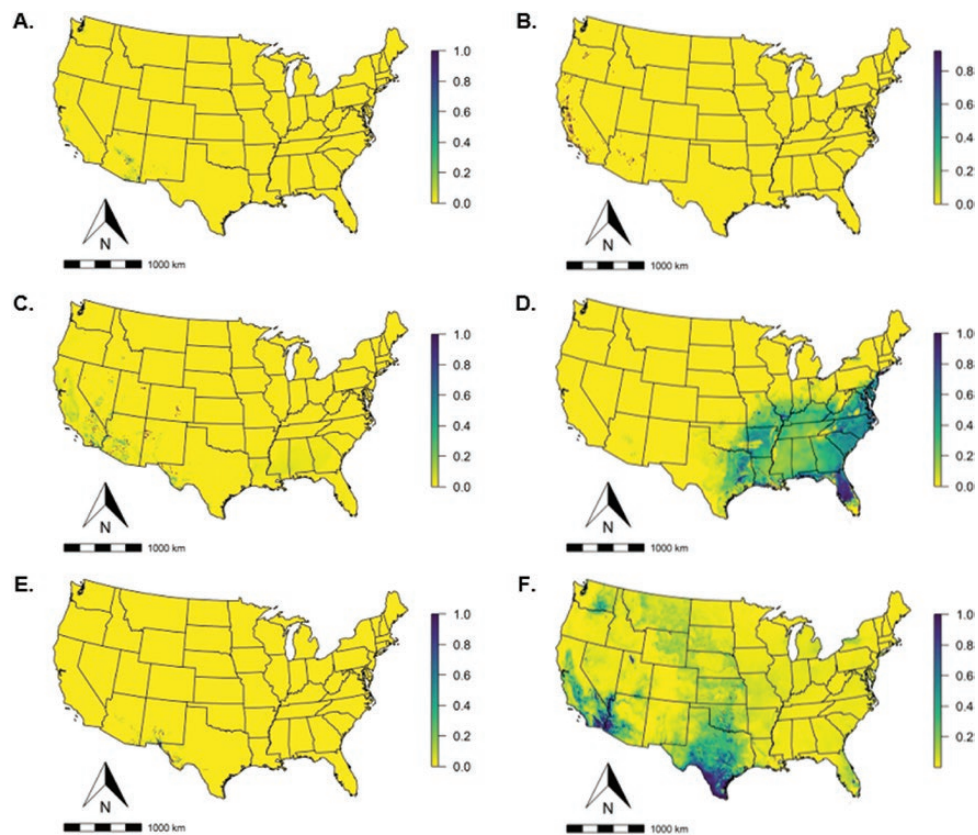
In this study, we sought to create conterminous United States species distribution models for six North American higher-attine nonleaf-cutting ant species from the genera *Trachymyrmex* and *Mycetomoellerius*. By modeling each species' distribution, our main goal was to determine which environmental variable(s) used in creating the model would contribute most to the overall test gain of the model, thus identifying variables that may explain each species' distribution. We also wanted to document whether each species' distribution was significantly different from the other species' distribution. This study provides the first predictive species distribution maps for six North American fungus-farming species and a more detailed look at what environmental factors are ecologically impactful to each species and their predicted range distributions. In all but one of the models (*T. desertorum*), climate variables were the most important to the model's performance. Annual mean temperature (BIO1) had the highest test gain for *M. turrifex*. *T. arizonensis*, *T. carinatus*, and *T. smithi* have precipitation seasonality (BIO15) as the most important variable in their models. These three species are western, desert species, so it seems reasonable that they would share similar climatic variables impacting their distributions. Nevertheless, these species appear to prefer slightly varying levels of precipitation seasonality within their habitat range (Fig. 4). *T. septentrionalis* has precipitation of coldest quarter (BIO19) as being the most important variable in its model's performance. Highlighted in Table 4 are the variables with the highest test gains for each model, however, multiple variables may be important to each species distribution. For example, *T. arizonensis* and *T. carinatus* have four variables that have test gains that are greater than 1: annual mean temperature (BIO1), isothermality (BIO3), precipitation seasonality (BIO15), and precipitation of coldest quarter (BIO19). On the contrary, soil variables contribute relatively little to these two species distributions.

**Table 2.** Selected model validation metrics for each species

Species	Allowed features	Regularization multiplier	AUC	TSS	SEDI	$OR_{MTP}$
<i>T. arizonensis</i>	LQHPT	1.28	$0.95 \pm 0.00$	$0.89 \pm 0.01$	$0.96 \pm 0.00$	$0.16 \pm 0.07$
<i>T. carinatus</i>	LQHP	2.21	$0.97 \pm 0.00$	$0.94 \pm 0.00$	$0.98 \pm 0.00$	$0.05 \pm 0.01$
<i>T. desertorum</i>	L	2.43	$0.89 \pm 0.02$	$0.81 \pm 0.05$	$0.93 \pm 0.01$	$0.14 \pm 0.04$
<i>T. septentrionalis</i>	LQH	2.46	$0.74 \pm 0.04$	$0.48 \pm 0.05$	$0.87 \pm 0.01$	$0.02 \pm 0.00$
<i>T. smithi</i>	LQH	2.30	$0.95 \pm 0.00$	$0.89 \pm 0.00$	$0.99 \pm 0.00$	$0.14 \pm 0.04$
<i>M. turrifex</i>	LQHP	2.17	$0.86 \pm 0.02$	$0.67 \pm 0.04$	$0.90 \pm 0.01$	$0.11 \pm 0.04$

All metrics are calculated from the test set.





**Fig. 3.** MaxEnt species distribution models for six higher-attine nonleaf cutter ant species in the continental United States: (A) *T. arizonensis*, (B) *T. carinatus*, (C) *T. desertorum*, (D) *T. septentrionalis*, (E) *T. smithi*, and (F) *M. turrifex*. USDA soil data and WorldClim climate data were used to create models. Background points were selected using a biased probability density for random sampling across the United States. Areas of dark blue indicate areas of high habitat suitability and light yellow indicate areas of extreme low habitat suitability.

In most studies utilizing distribution modeling techniques, species distribution models are used to create distributions of one species at a time. In this study, however, we created models that represent two species simultaneously: each ant species and its obligate symbiont, their fungal garden. Because the fungal gardens are found completely underground, it was hypothesized that the soil data may be important contributors to the ant species distributions. Surprisingly, however, compared to climate variables, soil variables tended to have lower test gain, thus contributing less to the overall models, except in the case of *T. desertorum*. This result was unexpected, as some species are found in habitats with very particular soil qualities; i.e., *T. septentrionalis* is generally found in very sandy soils and *M. turrifex* is found mostly in clay soil but can occur also in sandy soils (J.S., U.M., unpublished data; Rabeling et al. 2007). These findings may be the result of climate variables remaining relatively consistent from pixel to pixel in a given area and having a finer resolution compared to the coarsely grained soil variables used in these analyses. Additionally, soil variables tend to differ drastically in a given area, where climatic variables remain relatively similar in the same size area. Therefore, in order to find very specific soil requirements of each species, localized, smaller scale studies may need to be conducted. Future modeling efforts could focus on another, higher-resolution soil database, SSURGO (USGS 1995b), to focus in on smaller areas, such as a county or state.

The fact that climate was generally more important than soils to the ant distributions suggests that, at the coarse scale of the entire ranges of these species across the United States, broader climatic

factors are more important in determining habitat suitability than fine-scale factors like soil properties, at least at the relatively coarse resolution (i.e., the pixel or grain size) used in this study. This makes sense when considering that, across the entire United States, the most important determinant of whether an ant will be found in a particular region is the favorability of the climate in that region. Climate provides the coarse outline of the species' ranges; soils provide more definition or shading to these outlines. Scale dependency of ecological phenomena like habitat associations are a long-established concept in landscape ecology (Allen and Starr 1982).

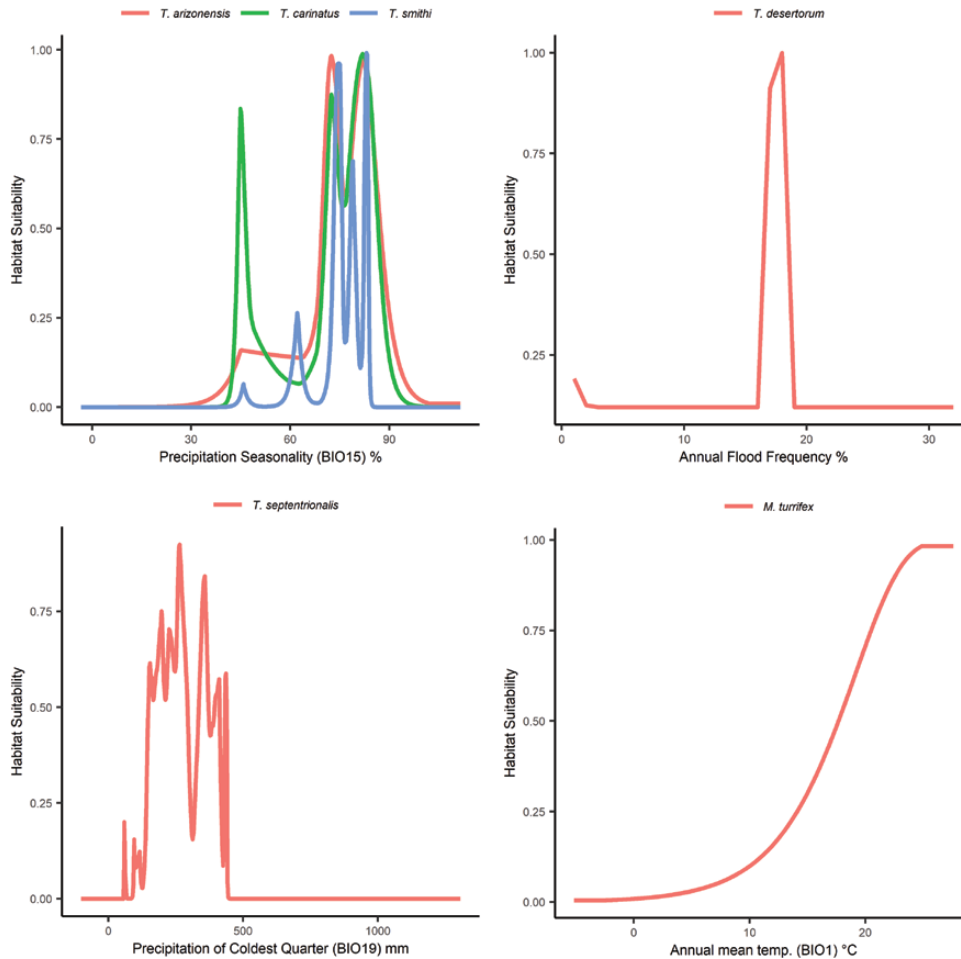
### Species Distribution Model Differences Among Species

Out of 15 possible comparisons between each species' distributions, only one was statistically indistinguishable (*T. desertorum* vs *T. septentrionalis*) ( $P = 0.07$ ) (Table 4). This result is somewhat unexpected, as we thought that *T. arizonensis*, *T. carinatus*, and *T. desertorum* would have greater niche overlap and distribution similarity. Though three out of the six species may occur in Texas, they are found in very different ecoregions (Rabeling et al. 2007). *T. smithi* is found primarily in the Chihuahuan Desert of western Texas, New Mexico, and Northern Mexico (Rabeling et al. 2007). *M. turrifex* is spread throughout most of Texas, southern Oklahoma, Arkansas, and Louisiana. In Fig. 3, however, there is moderate to high suitability found in many states, with high suitability in Texas and in Western Arizona and Southeastern California, though this species is unknown to occur in these localities.

**Table 3.** Summary information for each individual species' distribution models for the United States using biased random background point selection

	Test gain					
	<i>T. arizonensis</i>	<i>T. carinatus</i>	<i>T. desertorum</i>	<i>T. septentrionalis</i>	<i>T. smithi</i>	<i>M. turrifex</i>
Full model	5.82	$1.49 \times 10^6$	4.04	1.18	-1.79	1.13
Annual mean temp. (BIO1)	1.09	1.17	0.27	$-1.65 \times 10^7$	0.68	<b>0.38</b>
Isothermality (BIO3)	1.50	1.72	0.39	1.15	1.67	0.00
Min. temp. of coldest month (BIO6)	0.97	0.96	0.05	$-1.91 \times 10^4$	1.12	0.25
Temp. annual range (BIO7)	0.67	0.94	0.12	0.48	0.53	0.07
Mean temp. of wettest quarter (BIO8)	0.80	0.41	-0.01	0.30	0.97	0.18
Precipitation seasonality (BIO15)	<b>1.82</b>	<b>1.72</b>	0.30	-0.94	<b>2.10</b>	0.03
Precipitation of warmest quarter (BIO18)	0.78	0.02	0.65	$-2.05 \times 10^9$	1.59	0.13
Precipitation of coldest quarter (BIO19)	1.37	1.54	0.18	<b>0.49</b>	1.17	0.16
Annual flood frequency	0.24	0.15	<b>1.34</b>	0.24	0.18	0.21
Available water capacity	0.10	0.10	0.09	-0.01	-0.11	0.02
Calcium carbonate in soil layer	0.33	1.11	0.00	-0.62	1.08	0.16
Cation exchange capacity	0.26	0.51	-0.02	-0.54	1.36	0.10
Share of map unit with hydric soils	0.11	0.11	0.09	-0.02	-0.10	0.02
Erodibility	0.23	0.08	0.02	0.14	0.90	0.16
Average depth of bedrock	0.27	0.29	0.00	-9.64	1.10	0.01
Slope of map unit	0.24	0.46	0.59	0.14	1.18	0.00
Depth of soil	0.35	0.90	0.07	0.16	0.14	0.14

The test gains for the full models are presented, as well as test gains for model fit with only one single variable. The importance of a variable to the full model can be estimated by how much of the gain of the full model is accounted for by the gain of the model built with only that single variable. Bold values indicate the environmental variable with the highest test gain.

**Fig. 4.** Response curves for the most important layer for each species.



**Table 4.** Observed *I*-values and critical *I* values from the permutation tests

Species comparison	Observed <i>I</i>	5% critical <i>I</i>	Estimated <i>P</i> -value
<i>T. arizonensis</i> vs <i>T. carinatus</i>	0.152	0.270	0.01
<i>T. arizonensis</i> vs <i>T. desertorum</i>	0.072	0.152	0.01
<i>T. arizonensis</i> vs <i>T. septentrionalis</i>	0.039	0.168	0.02
<i>T. arizonensis</i> vs <i>T. smithi</i>	0.018	0.242	0.00
<i>T. arizonensis</i> vs <i>M. turrifex</i>	0.055	0.266	0.00
<i>T. carinatus</i> vs <i>T. desertorum</i>	0.070	0.306	0.00
<i>T. carinatus</i> vs <i>T. septentrionalis</i>	0.017	0.020	0.04
<i>T. carinatus</i> vs <i>T. smithi</i>	0.024	0.329	0.00
<i>T. carinatus</i> vs <i>M. turrifex</i>	0.033	0.392	0.00
<i>T. desertorum</i> vs <i>T. septentrionalis</i>	0.042	0.025	0.07
<i>T. desertorum</i> vs <i>T. smithi</i>	0.049	0.205	0.00
<i>T. desertorum</i> vs <i>M. turrifex</i>	0.093	0.111	0.03
<i>T. septentrionalis</i> vs <i>T. smithi</i>	0.037	0.528	0.03
<i>T. septentrionalis</i> vs <i>M. turrifex</i>	0.505	0.859	0.00
<i>T. smithi</i> vs <i>M. turrifex</i>	0.140	0.362	0.00

Significant results (nonidentical niches) occur when the observed value is below the 5% critical value from the permutation analysis. Pairs of niches that are significantly or marginally significantly different are highlighted. Models being compared were generated with biased random background selection. Significant *P*-values are highlighted in bold.

The distributions of *T. smithi* and *M. turrifex* may have some overlap in far-west Texas, but otherwise have different distributions, which is documented in collection records and is supported by the distribution models in this study. In west Texas, the north-to-south mountain chain of the Guadalupe mountains, Delaware mountains, Apache mountains, and Davis mountains form a higher-elevation barrier where *Trachymyrmex* do not seem to occur, and where *T. smithi* occurs in sandy areas mostly west of that barrier and *M. turrifex* mostly east of that barrier, but both *T. smithi* and *M. turrifex* cooccur in the Big Bend area of south-west Texas (Rabeling et al. 2007; U.M., unpublished data). The distributions of *T. smithi* and *M. turrifex* both differ from the remaining fungus-farming species found in Texas, *T. septentrionalis*. *T. septentrionalis* is found primarily in the Post Oak Savannahs and Piney Woods of central and east Texas. This species shares some overlap with *M. turrifex* in these areas, but because their soil requirements may differ, this overlap tends to be locally patchy (J.S., U.M., unpublished data). The distributions of the three most commonly occurring species found in Texas were significantly different than the desert occurring species, which is not extraordinary, as their habitat requirements are, in general, drastically different.

An additional two species of higher-attine nonleaf-cutting species that are found in the conterminous United States, *T. nogalensis* and *T. pomonae*, did not have habitat suitability maps created, as there are not enough known locations of these species to create dependable models. A ninth higher-attine nonleaf-cutting species, *Mycetomoellerius jamaicensis*, was also not included in this study because its distribution is limited to a few locations in coastal southeast Florida, the Florida Keys, and the Caribbean (Rabeling et al. 2007). Future work could readily model *M. jamaicensis* using the same approach and the same soil and climate layers described here, but at a smaller spatial extent more appropriate to forecasting the distribution of a localized species. More occurrence data is needed to create models for *T. nogalensis* and *T. pomonae*.

## Conclusion

The methodology used in this study yielded predictive models that performed well on a range of metrics. However, these findings do not guarantee a realized niche prediction, or where the species is

found, but instead areas of high habitat suitability for each species, or the fundamental niche. The relative high predictive ability, based on model validation metrics, of these models may help focus surveying efforts to pinpoint areas of high habitat suitability resulting in the discovery of more populations of these ecologically important species. In this way, the species distribution models presented here serve as a first rough draft of the range maps for these species. This study supports the hypothesis that temperature and precipitation, rather than soil in most cases, is a range-limiting factor in most modeled species distributions.

## Supplementary Data

Supplementary data are available at *Journal of Insect Science* online.

## Acknowledgments

The work was supported by National Science Foundation award DEB-1354629 to K.K. and by National Science Foundation CAREER award IOS 1552822 to J.N.S. Additional funding and access to private lands was obtained from Texas Ecolab. We would like to thank the three anonymous reviewers and the editor for insightful comments which were most helpful and have substantially improved our manuscript.

## References

- Akman Gündüz, E., and A. E. Douglas. 2008. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proc. R. Soc. B.* 276: 987–991.
- Allen, T. F. H., and T. B. Starr. 1982. *Hierarchy: perspectives for ecological complexity*. University of Chicago Press, Chicago, IL.
- Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43: 1223–1232.
- Aylward, F., C. Currie, and G. Suen. 2012. The evolutionary innovation of nutritional symbioses in leaf-cutter ants. *Insects* 3: 41–61.
- Bolton, B. 2016. An online catalog of the ants of the world. Available from <https://www.antweb.org/>
- Branstetter, M. G., A. Ješovnik, J. Sosa-Calvo, M. W. Lloyd, B. C. Faircloth, S. G. Brady, and T. R. Schultz. 2017. Dry habitats were crucibles of domestication in the evolution of agriculture in ants. *Proc. R. Soc.* 284: 20170095.

- Brune, A. 2014. Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12: 168–180.
- Cardoso, D. C., and M. P. Cristiano. 2010. Myrmecofauna of the Southern Catarinense Restinga Sandy Coastal Plain: new records of species occurrence for the state of Santa Catarina and Brazil. *Sociobiology* 55: 229–240.
- Cardoso, D. C., T. G. Sobrinho, and J. H. Schoeder. 2010. Ant community composition and its relationship with phytophysiognomies in a Brazilian Restinga. *Insectes Sociaux* 57(3): 293–301.
- Chapman, B. R., and E. G. Bolen. 2018. The natural history of Texas. Texas A&M University Press.
- DeMillo, A.M., M. Rouquette Jr., U. G. Mueller, K. Kellner, and J. N. Seal. 2017. Effects of substrate, ant and fungal species on plant fiber degradation in a fungus-gardening ant symbiosis. *J. Insect Physiol.* 98: 301–308.
- Diehl-Fleig, E., and E. S. Da Rocha. 1998. Soil selection by *Acromyrmex striatus* (Roger) females (Hymenoptera: Formicidae) for nest building. *An. Soc. Entomol. Bras.* 27(1): 41–45.
- Diniz-Filho, J. A. F., P. De Marco Jr., and B. A. Hawkins. 2010. Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. *Insect Conserv. Divers.* 3(3): 172–179.
- Douglas, A. E. 2009. The microbial dimension in insect nutritional ecology. *Funct. Ecol.* 23: 38–47.
- Douglas, A. E. 2010. The symbiotic habit. Princeton University Press, Princeton, NJ, 232 pp.
- Douglas, A. E. 2015. Multiorganismal insects: diversity and function of resident microorganisms. *Annu. Rev. Entomol.* 60: 17–34.
- Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* 40: 677–697.
- Elith, J., S. Phillips, T. Hastie, M. Dudík, Y. Chee, and C. Yates. 2010. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17(1): 43–57.
- Engel, P., and N. A. Moran. 2013. The gut microbiota of insects—diversity in structure and function. *FEMS Microbiol. Rev.* 37(5): 699–735
- Fernandes, R. F., D. Scherrer, and A. Guisan. 2018. Effects of simulated observation errors on the performance of species distribution models. *Divers. Distrib.* 25: 400–413.
- Fourcade, Y., J. O. Engler, D. Rödder, and J. Secondi. 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLOS ONE* 9(5): e97122.
- Fourcade, Y., A. G. Besnard, and J. Secondi. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Glob. Ecol. Biogeogr.* 27: 245–256.
- van Gils, H., and C. Vanderwoude. 2012. Leafcutter ant (*Atta sexdens*) (Hymenoptera: Formicidae) nest distribution responds to canopy removal and changes in micro-climate in the Southern Colombian Amazon. *Fla. Entomol.* 95(4): 914–921.
- GRASS Development Team. 2017. Geographic resources analysis support system (GRASS) software, version 7.2. Open Source Geospatial Foundation. <http://grass.osgeo.org>
- Hijmans, R. J. 2019. raster: geographic data analysis and modeling. R package version 3.0-2. <https://CRAN.R-project.org/package=raster>
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25: 1965–1978.
- Hijmans, R. J., S. Phillips, Leathwick, J. and J. Elith. 2017. dismo: species distribution modeling. R package version 1.1-4. <https://CRAN.R-project.org/package=dismo>
- Hölldobler, B., and E. O. Wilson. 1990. The ants. Harvard University Press, Cambridge, MA.
- Hu, Y., J. G. Sanders, P. Łukasik, J. S. Millar, D. R. Vann, J. A. Newton, M. Schotanus, D. J. C. Kronauer, N. E. Pierce, C. S. Morreau, et al. 2018. Herbivorous turtle ants obtain essential nutrients from a conserved nitrogen-recycling gut microbiome. *Nat. Commun.* 9(1): 1–14.
- Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Glob. Ecol. Biogeogr.* 21: 498–507.
- King, J. R., R. J. Warren, and M. A. Bradford. 2013. Social insects dominate Eastern US temperate hardwood forest Macroinvertebrate communities in warmer regions. *PLOS ONE* 8(10): e75843.
- Kramer-Schadt, S., J. Niedballa, J. D. Pilgrim, B. Schröder, J. Lindenborn, V. Reinfelder, M. Stillfried, I. Heckmann, A. K. Scharf, D. M. Augeri, et al. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19: 1366–1379.
- Lobo, J. 2016. The use of occurrence data to predict the effects of climate change on insects. *Curr. Opin. Insect Sci.* 17: 62–68.
- Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17: 145–151.
- Malherbe, C., and N. Vayatis. 2017. Global optimization of lipschitz functions. *In Proceedings of the 34th International Conference on Machine Learning*, vol. 70. pp. 2314–2323. [JMLR.org](http://jmlr.org).
- Marcen, A., L. Sáez, R. Molowny-Horas, X. Pons, and J. Pino. 2013. Using species distribution modelling to disentangle realised versus potential distributions for rare species conservation. *Biol. Conserv.* 166: 221–230.
- Mednick, A. C. 2010. Does soil data resolution matter? State soil geographic database versus soil survey geographic database in rainfall-runoff modeling across Wisconsin. *J. Soil Water Conserv.* 65(3): 190–199.
- Meyer, S. T., I. R. Leal, M. Tabarelli, and R. Wirth. 2011. Ecosystem engineering by leaf-cutting ants: nests of *Atta cephalotes* drastically alter forest structure and microclimate. *Ecol. Entomol.* 36(1): 14–24.
- Mueller, U. G., N. M. Gerardo, D. K. Aanen, D. L. Six, and T. R. Schultz. 2005. The evolution of agriculture in insects. *Annu. Rev. Ecol. Evol. Syst.* 36: 563–595.
- Mueller, U. G., H. D. Ishak, S. M. Bruschi, C. C. Smith, J. J. Herman, S. E. Solomon, A. S. Mikheyev, C. Rabeling, J. J. Scott, M. Cooper, et al. 2017. Biogeography of mutualistic fungi cultivated by leafcutter ants. *Mol. Ecol.* 26(24): 6921–6937.
- Muhamad, A., Y. Fang, Y. Hou, and Z. Shi. 2017. The gut entomotype of red palm weevil *Rhynchophorus ferrugineus* Olivier (Coleoptera: Dryophthoridae) and their effect on host nutrition metabolism. *Front. Microbiol.* 8: 2291.
- Muscarella, R., P. J. Galante, M. Soley-Guardia, R. A. Boria, J. M. Kass, M. Uriarte, and R. P. Anderson. 2014. ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. *Methods Ecol. Evol.* 5(11): 1198–1205.
- Noss, R. F., W. J. Platt, B. A. Sorrie, A. S. Weakley, D. B. Means, J. Costanza, and R. K. Peet. 2015. How global biodiversity hotspots may go unrecognized: lessons from the North American coastal plain. *Divers. Distrib.* 21(2): 236–244.
- O'Donnell, M. S., and D. A. Ignizio. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States. U.S. Geological Survey Data Series, vol. 691, 10 pp.
- Oliver, K. M., and A. J. Martinez. 2014. How resident microbes modulate ecologically-important traits of insects. *Curr. Opin. Insect Sci.* 4: 1–7.
- Omernik, J. M., and G. E. Griffith. 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. *Environ. Manage.* 54(6): 1249–1266.
- Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J. Biogeogr.* 34(1): 102–117.
- Peterson, A. T., J. Soberon, R. G. Pearson, R. P. Anderson, E. Martinez-Meyer, M. Nakamura, and M. S. Araújo. 2011. Ecological niches and geographic distributions. Princeton University Press, Princeton, NJ.
- Phillips, S. J. 2006. A brief tutorial on Maxent. AT&T Research. Available from [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/)
- Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31(2): 161–175.
- Phillips, S. J., M. Dudík, and R. E. Schapire. [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/)
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190(3–4): 231–259.

- van Proosdij, A. S. J., M. S. M. Sosef, J. J. Wieringa, and N. Raes. 2016. Minimum required number of specimen records to develop accurate species distribution models. *Ecography* 39(6): 542–552.
- Rabeling, C., S. P. Cover, R. A. Johnson, and U. G. Mueller. 2007. A review of the North American species of the fungus gardening ant Genus *Trachymyrmex* (Hymenoptera: Formicidae). *Zootaxa* 1664(1): 1–53.
- Russel, J. A., J. G. Sanders, and C. S. Moreau. 2016. Hotspots for symbiosis: function, evolution, and specificity of ant-microbe associations from trunk to tips of the ant phylogeny (Hymenoptera: Formicidae). *Myrmecol. News* 24: 43–69.
- Russell, J. A., C. S. Moreau, B. Goldman-Huertas, M. Fujiwara, D. J. Lohman, and N. E. Pierce. 2009. Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proc. Natl. Acad. Sci. U. S. A.* 106(50): 21236–21241.
- Savopoulou-Soultani, M., N. T. Papadopoulos, P. Milonas, and P. Moyal. 2012. Abiotic factors and insect abundance. *Psyche* (Camb Mass) 2012: 1–2.
- Scavetta, J. 2019. maxentools (Version v1.1.0). Zenodo. <https://github.com/jscavetta95/maxentools/http://doi.org/10.5281/zenodo.3537679>
- Schultz, T. R., and S. G. Brady. 2008. Major evolutionary transitions in ant agriculture. *Proc. Natl. Acad. Sci. U. S. A.* 105(14): 5435–5440.
- Seal, J. N., and W. R. Tschinkel. 2006. Colony productivity of the fungus-gardening ant *Trachymyrmex septentrionalis* (Hymenoptera: Formicidae) in a Florida pine forest. *Ann. Entomol. Soc. Am.* 99(4): 673–682.
- Seal, J. N. and W. R. Tschinkel. 2010. Distribution of the fungus-gardening ant (*Trachymyrmex septentrionalis*) during and after a record drought. *Insect Conserv. Divers.* 3(2): 134–142.
- Seal, J. N., L. Brown, C. Ontiveros, J. Thiebaud, and U. G. Mueller. 2015. Gone to Texas: phylogeography of two *Trachymyrmex* (Hymenoptera: Formicidae) species along the southeastern coastal plain of North America. *Biol. J. Linn. Soc. Lond.* 114: 689–698.
- Simões-Gomes, F. C., D. C. Cardoso, and M. P. Cristiano. 2017. Geographical distribution patterns and niche modeling of the iconic leafcutter ant *Acromyrmex striatus* (Hymenoptera: Formicidae). *J. Insect Sci.* 17(2): 32, 1–5.
- Sokal, R. R., and F. J. Rohlf. 1995. *Biometry: the principles and practice of statistics in biological research*. 2nd ed. W.H. Freeman and Company, New York, NY.
- Solomon, S. E., C. Rabeling, J. Sosa-Calvo, C. T. Lopes, A. Rodrigues, H. L. Vasconcelos, M. Bacci Jr., U. G. Mueller, and T. R. Schultz. 2019. The molecular phylogenetics of *Trachymyrmex* ants and their fungal cultivars provide insights into the origin and co-evolutionary history of “higher-attine” ant agriculture. *Syst. Entomol.* 44(4): o1–o3, 667–1003.
- Soltis, D. E., A. B. Morris, J. S. Mclachlan, P. S. Manos, and P. S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. *Mol. Ecol.* 15: 4261–4293.
- Swanson, A. C., L. Schwendenmann, M. F. Allen, E. L. Aronson, A. Artavia-León, D. Dierick, A. S. Fernandez-Bou, T. C. Harmon, C. Murillo-Cruz, S. F. Oberbauer, et al. 2019. Welcome to the Atta world: a framework for understanding the effects of leaf-cutter ants on ecosystem functions. *Funct. Ecol.* doi:10.1111/1365-2435.13319
- Tschinkel, W. R., and J. N. Seal. 2016. Bioturbation by the fungus-gardening ant, *Trachymyrmex septentrionalis*. *PLOS ONE* 11: e0158920.
- Tschinkel, W. R., T. Murdock, J. R. King, and C. Kwapich. 2012. Ant distribution in relation to ground water in north Florida pine flatwoods. *J. Insect Sci.* 12: 114.
- USDA NRCS (Natural Resources Conservation Service). 1994. State Soil Geographic (STATSGO) Data Base: data use information. Miscellaneous Publication Number 1492. USDA Natural Resources Conservation Service, Washington, DC.
- United States Department of Agriculture, National Resources Conservation Service, National Soil Survey Center. 1995a. State Soil Geographic (STATSGO) Data Base: data use information. Miscellaneous Publication Number 1492. Downloaded from <http://www.fsl.orst.edu/pnwer/wrb/metadata/soils/statsgo.pdf> on 13 September 2018.
- United States Department of Agriculture, National Resources Conservation Service, National Soil Survey Center. 1995b. Soil Survey Geographic (SSURGO) Data Base: data use information. Miscellaneous Publication Number 1527. Downloaded from <http://www.fsl.orst.edu/pnwer/wrb/metadata/soils/ssurgo.pdf> on 13 September 2018.
- Walters, A. D., D. Ford, E. T. Chong, M. G. Williams, N. B. Ford, L. R. Williams, and J. A. Banta. 2017. High resolution ecological niche modelling of threatened freshwater mussels in east Texas, USA. *Aquat. Conserv.* 27(6): 1251–1260.
- Ward, P. S. 2005. A synoptic review of the ants of California (Hymenoptera: Formicidae). *Zootaxa* 936: 3–68.
- Ward, P. S. 2014. The phylogeny and evolution of ants. *Annu. Rev. Ecol. Evol. Syst.* 45: 23–43.
- Ward, P. S., S. G. Brady, B. L. Fisher, and T. R. Schultz. 2015. The evolution of myrmicine ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). *Syst. Entomol.* 40(1): 61–81.
- Warren, D. L. 2012. In defense of ‘niche modeling’. *Trends Ecol. Evol.* 27(9): 497–500.
- Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecol. Appl.* 21: 335–342.
- Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62(11): 2868–2883.
- Warren, D. L., R. E. Glor, and M. Turelli. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33(3): 607–611.
- Weber, N. A. 1956. Fungus growing ants and their fungi: *Trachymyrmex septentrionalis*. *Ecology* 37: 150–161
- Weber, N. A. 1972. *Gardening ants: the attines*. American Philosophical Society 92, Philadelphia, PA.
- Wolock, D. M. 1997. STATSGO soil characteristics for the conterminous Unites States. Available from <https://water.usgs.gov/GIS/metadata/usgswrd/XML/muid.xml>
- Wunderlich, R. F., Y-P. Lin, J. Anthony, and J. R. Petway. 2019. Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. *Nat. Conserv.* 35: 97–116.