Mapping the habitat suitability of *Andira humilis* Mart. ex Benth. (Fabaceae) as a means to detect its associated galling species in Brazil

Valdeir Pereira Lima$^{1,2}$ and Danièla Cristina Calado$^2$

$^1$Departamento de Fitotecnia, Centro de Ciências Agrárias, Universidade Federal de Santa Catarina, Rodovia Admar Gonzaga, 1546, Itacorubi, 88054-000, Florianópolis, Santa Catarina, Brazil. $^2$Laboratório de Entomologia, Centro das Ciências Biológicas e da Saúde, Universidade Federal do Oeste da Bahia, Barreiras, Bahia, Brazil. *Author for correspondence. E-mail: valldeir@gmail.com*

**ABSTRACT.** Host plant species have very specific interconnection with galling species. Here, we estimate the potential distribution of the host plant species *Andira humilis* Mart. ex Benth. (Fabaceae) to consequently locate the potential distribution ranges of its galling species *Lopesia andirae* García, Lima, Calado, and Guimarães (2017) based on ecological requirements. The ecological niche model was built using Maxent v.3.4.1k, an algorithm that estimates species’ distributions. We found suitable habitats for *L. andirae* encompassing areas of the Cerrado, Caatinga and Atlantic Forest. Annual mean temperature (70.2%) and temperature annual range (15.9%) were the most critical factors shaping *A. humilis* and necessarily *L. andirae*. Our results can guide taxonomists and ecologists regarding the delineation of sampling areas as well as conservation strategies for this ecological interaction.

**Keywords:** biodiversity; Cecidomyiidae; computational biology; gall-inducing insects; species distribution modeling.

Received on July 19, 2019. 
Accepted on November 27, 2019.

**Introduction**

Galling species are specialized herbivores able to induce redifferentiation of specialized plant tissues (Arriola, Melo Júnior, Mouga, Isaias, & Costa, 2016; Fernandes, Tameirão Neto, & Martins, 1988; Oliveira & Isaias, 2010; Shorthouse, Wool, & Raman, 2005). An estimation of the overall number of galling species ranges from 21,000 to 211,000 with the highest diversity in warm regions and associated with sclerophyllous vegetation (Lara & Fernandes, 1996). Although the highest galling species diversity is found in the tropical region, most of the taxonomic knowledge about this group is predominantly based on temperate regions (Santo & Fernandes, 2007). The family Cecidomyiidae is the most diverse group of galling insects with 6,590 described species, distributed in 812 genera on the planet (Gagné & Jaschhof, 2017) and approximately 222 species in Brazil (Maia, 2020).

Host plant species have very specific interconnection with galling species (Arriola et al., 2016; Carneiro et al., 2009; Lima & Calado, 2018; Shorthouse et al., 2005). This specificity is demonstrated by the fact that several plants are hosts for a diversity of gall morphotypes, which shows that each plant species presents different stimuli to different galling species (Arriola et al., 2016; Araújo, Scareli-Santos, Guilherme, & Cuevas-Reyes, 2015; Isaias, Oliveira, Carneiro, & Kraus, 2014; Shorthouse et al., 2005). According to Carneiro et al. (2009), approximately 92% Cecidomyiidae species are monophagous and only 5.6% are either oligophagous or have the ability to induce galls in the same plant genus. These authors are convinced that gall morphotypes associated with host plant species may be a reliable indicator of insect-inducing species. To illustrate, in tropical areas where little taxonomic knowledge studies on gall midges are performed, gall morphotypes have been used as a surrogate for insect species (Fernandes & Price, 1988).

The scarceness of information on the distribution of galling insects limits the understanding of population dynamics, dispersion and evolutionary biology of this group (Gagné & Jaschhof, 2017). *Lopesia andirae*, García et al. (2017), was described based on data from three Brazilian states associated with *Andira humilis* Mart. ex Benth (Fabaceae), a shrub endemic to Brazil (Periotto, Perez, & Lima, 2004). The galling species *L. andirae* is merely known in the following three localities: *Parque Nacional Chapada dos Guimarães* (Mato Grosso State), *Universidade Federal do Oeste da Bahia*, Barreiras campus (Bahia State) and Luiz
Antônio (São Paulo State) (Garcia et al., 2017). On the other hand, the host plant *A. humilis* is well documented owning to greater systematic studies incorporated into herbaria.

Ecological niche models (ENMs) have become one of the most employed tools to estimate species distribution based on occurrence records and environmental variables (Ashraf et al., 2017; Gomes et al., 2018; Guisan & Thuiller, 2005). These models are pivotal as they allow estimation of diversity patterns, determining potential areas of persistence, extinction and colonization (Assis, Araújo, & Serrão, 2018). ENMs can be integrated with Geographic Information System (GIS) to provide valuable information with regard to the development of conservation strategies, including the determination of priority areas for conservation and the understanding of biodiversity patterns (Balram, Dragićević, & Meredith, 2004). Here, we combine ENMs and GIS approach to estimate the potential distribution of the host plant species *A. humilis* and consequently locate the potential distribution ranges of its gall-inducing species *L. andirae* based on ecological needs.

### Material and methods

The species occurrence data was obtained from the literature and online databases such as SpeciesLink (http://splink.cria.org.br) and Global Biodiversity Information Facility (https://www.gbif.org). Information on the species occurrence range was checked at the Flora do Brasil (http://floradobrasil.jbrj.gov.br/) and those records outside the original geographical distribution of the species were excluded. To reduce spatial autocorrelation and improve the performance of ENMs (Boria, Olson, Goodman, & Anderson, 2014; Fourcade, Besnard, & Secondi, 2018), we spatially filtered the species data at a distance of 20 km (Zwiener et al., 2017), using a function from spThin R package (Aiello-Lammens, Boria, Radosavljevic, Vilela, & Anderson, 2015) in the R statistical programming (R Development Core Team, 2014).

To feature species climatic requirements, we downloaded the 19 bioclimatic variables from the Worldclim version 1.4. (http://worldclim.org) at 30 seconds resolution and selected *a priori* the following predictors, annual mean temperature (Bio1), temperature seasonality (Bio4), max temperature of the warmest month (Bio5), min temperature of the coldest month (Bio6), temperature annual range (Bio7), annual precipitation (Bio12), precipitation of the wettest month (Bio13), precipitation of the driest month (Bio14), and precipitation seasonality (Bio15) based on their biological meaning (Elith, Kearney, & Phillips, 2010). Then, we checked the multicollinearity of these predictors through the variance inflation factor (O’Brien, 2007) in the R environment (R Development Core Team, 2014).

ENM was built using Maxent v.3.4.1k, an algorithm that estimates the probability of species’ distributions (Elith et al., 2011; Phillips, Anderson, & Schapire, 2006; Phillips & Dudík, 2008). Maxent is a presence-only method that presents higher performance, even when few occurrence data are available (Elith et al., 2006). Aiming to build a more parsimonious model, we tested different feature classes, linear (L), quadratic (Q), product (P), hinge (H) and threshold (T), as well as different regularization multiplier values to select the best-fit model. For comparing these models, we adopted the corrected Akaike Information Criterion (AICc) implemented in ENM Tools v 1.3 (Warren, Glor, & Turelli, 2010). Lower values of AICc indicate best-fit models (Warren et al., 2010).

The best-fit model was run with the following changes in the MaxEnt default settings: (i) enable response curves to evaluate species response to each predictor variable, (ii) perform jackknife analysis to measure variable importance, (iii) set 75% of the occurrence records for training and the remaining for testing the model, (iv) set replicated run-type as bootstrap with 100 replicates, and (v) enable write background predictions. The final model was evaluated using True Skill Statistic (TSS) (Allouche, Tsoar, & Kadmon, 2006) and Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) (Fielding & Bell, 1997; Peterson et al., 2011).

### Results

The occurrence records of *A. humilis* and *L. andirae* are plotted in Figure 1. A total of 545 occurrence records were obtained for *A. humilis* from the literature and online databases. The best-fit model included linear, quadratic, hinge, product and threshold functions and the regularization multiplier of three (LQHPT 3) (Table 1). Model for *A. humilis* performed better than random, with average AUC test of 0.85 and TSS of 0.72, indicating that the model performed well and generated excellent evaluations.
Figure 1. Occurrence records of Andira humilis (Host plant species) in red circle and Lopesia andirae (Gall-inducing insect) in blue triangle in Brazil.

Table 1. Best-fit model for Andira humilis in Brazil. The chosen model is highlighted in bold.

<table>
<thead>
<tr>
<th>Feature classes</th>
<th>Regularization</th>
<th>AICc</th>
<th>AUC</th>
<th>∆AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>LQHPT 5</td>
<td>4002.775275</td>
<td>0.90</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>LQ 1</td>
<td>4015.356285</td>
<td>0.88</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>LQH 5</td>
<td>4016.606819</td>
<td>0.89</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>LQHPT 5</td>
<td>4021.711563</td>
<td>0.90</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>LQ 3</td>
<td>4032.305817</td>
<td>0.88</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td>LQHPT 3</td>
<td>4049.831221</td>
<td>0.89</td>
<td>1.95</td>
<td></td>
</tr>
<tr>
<td>LQHPT 5</td>
<td>4054.186996</td>
<td>0.88</td>
<td>1.97</td>
<td></td>
</tr>
<tr>
<td>LQ 5</td>
<td>4056.504688</td>
<td>0.88</td>
<td>4.05</td>
<td></td>
</tr>
<tr>
<td>LQH 5</td>
<td>4059.250563</td>
<td>0.95</td>
<td>4.72</td>
<td></td>
</tr>
<tr>
<td>L 3</td>
<td>4069.412022</td>
<td>0.89</td>
<td>5.71</td>
<td></td>
</tr>
<tr>
<td>L 5</td>
<td>4070.326033</td>
<td>0.88</td>
<td>8.29</td>
<td></td>
</tr>
<tr>
<td>LQH 1</td>
<td>4071.010457</td>
<td>0.89</td>
<td>12.89</td>
<td></td>
</tr>
<tr>
<td>LQH 5</td>
<td>4080.357651</td>
<td>0.90</td>
<td>18.86</td>
<td></td>
</tr>
<tr>
<td>H 3</td>
<td>4131.585635</td>
<td>0.92</td>
<td>61.44</td>
<td></td>
</tr>
</tbody>
</table>

The analysis of variable contributions executed by the Jackknife test showed that variables contributed with different percentages. We identified that the annual mean temperature (70.2%) and temperature annual range (13.9%) were the most critical factors shaping A. humilis (Table 2). The results of the Jackknife test indicated that the environmental variable with highest gain when used in isolation is annual mean temperature, which therefore appears to have the most useful information by itself. At the same time, the environmental variable that decreases the gain the most when it is omitted is the annual mean temperature, which therefore appears to have the most information that is not present in the other variables.

Table 2. Analysis of variable contributions for A. humilis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent contribution</th>
<th>Permutation importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Mean Temperature</td>
<td>70.2</td>
<td>59</td>
</tr>
<tr>
<td>Temperature Annual Range</td>
<td>13.9</td>
<td>7</td>
</tr>
<tr>
<td>Precipitation of the Driest Month</td>
<td>12.5</td>
<td>29.2</td>
</tr>
<tr>
<td>Precipitation Seasonality</td>
<td>3</td>
<td>3.3</td>
</tr>
<tr>
<td>Precipitation of the Wettest Month</td>
<td>0.5</td>
<td>1.4</td>
</tr>
</tbody>
</table>
The predicted distribution ranges for *A. humilis* in Brazil are illustrated in Figure 2. The suitable areas include Cerrado, Caatinga and Atlantic forest, with higher suitability in São Paulo, Minas Gerais, Rio de Janeiro, Espírito Santo (Southeast Region); west, central-north, central-east and southwest regions of Bahia, Sergipe, Alagoas, Pernambuco, Paraíba, Rio Grande do Norte and Ceará (Northeast Region); southeast, southwest and central-south regions of Mato Grosso, east and south-west regions of Mato Grosso do Sul and northeast, mid-north, east-center, east, south, metropolitan, west and extreme southwest regions of Goiás (Midwest Region); and northwest, central-north and pioneer north regions of Paraná (South Region) (Figure 2).

![Figure 2](image.png)

**Figure 2.** Habitat suitability for *Andira humilis* in Brazil. The binary map was built using maximum training sensitivity plus specificity logistic threshold.

**Discussion**

Model evaluation is a pivotal step to assess the accuracy of ecological niche models and consequently its resulting predictions (Peterson et al., 2011). MaxEnt often present high performance when compared to other algorithms (Elith et al., 2006; Guo, Li, Zhao, & Nawaz, 2019). Models run with different settings tend to perform better when compared to those with default settings. Thus, it is important to test different configurations of Maxent in order to obtain a better performance in the construction of ecological models (Warren, Wright, Seifert, & Shaffer, 2014). Our model showed high accuracy, performing better than random.

The analysis, as well as the selection of the environmental predictors, is a very important step for the construction of more parsimonious models (West et al., 2015). The annual mean temperature seems to be the main biologically important variable, shaping the distribution of several species in the Neotropical region. Similar to our study, annual mean temperature was the variable that contributed most to the ecological model of *Passiflora actinia* Hook (Passifloraceae) from the southern Atlantic Forest (Teixeira, Mäder, Arias, Bonatto, & Freitas, 2016).

Owing to the specialized relationship between galling insects and their host plants, Arriola et al. (2016) hypothesized that the distribution of the host plant *Calopodium brasiliense* Cambess. (Calophyllaceae) matches the distribution of its galling insects. For that, they estimated the geographic distribution of galling insects based on virtual collections of plants, extending the galling insect occurrence to 13 Brazilian states and 11 countries of the Neotropical region. Although we acknowledge authors’ efforts in estimating the distribution of this taxon, this assumption might not be realistic as species require very specific ecological
conditions for their survival (Slater & Michael, 2012). To illustrate, the Bahia State is characterized by three different biomes (Cerrado, Caatinga and Atlantic forest), which can provide different habitat suitability for particular species. We likewise agree that the distribution of galling species and their host plants may overlap each other. However, we argue that the best approach to estimate the potential distribution of galling insects are those based on ecological requirements as demonstrated in this study.

A straightforward consequence of galling insect dependence on host plants is the fact that their potential distribution is conditioned by environmental disturbances that plants may suffer. A. humilis has a wide distribution in Brazil since it finds suitable areas in the Cerrado, Caatinga and Atlantic Forest. Considered hotspots of biodiversity, Cerrado and Atlantic Forest are some of Earth’s most species-rich terrestrial regions, though they are threatened with destruction (Myers, Mittermeier, Mittermeier, Fonseca, & Kent, 2000) predominantly because of the conversion of their natural areas into large monocultures (Batistella & Valladares, 2010) and habitat fragmentation (Ribeiro et al., 2011). In this scenario of insect dependence on host plants and environmental degradation, we clearly observed that human activities may impair and drive insect gall species to extinction without the opportunity to know them.

Research on galling insects have significant geographic bias (Araújo, 2018; Maia, 2013). Analyzing the last 30 years of research on insect galls in Brazil, Araújo (2018) noticed that more than 60% of the studies of gall inventories are carried out in the southeastern region, which coincides with researchers living in this region and that more than 50% of Brazilian states do not have studies on the occurrence of galling insects. Although our study here displays the habitat suitability for L. andirae, a poorly known and perhaps threatened species, to guide taxonomists during data collections, the lack of researchers interested in taxonomy might be a major limitation to record and map gall insect diversity in Brazil.

**Conclusion**

Here, we applied ENM to estimate the habitat suitability for the host plant species A. humilis and consequently identify suitable areas for its galling species L. andirae in Brazil. The habitat suitability for L. andirae encompasses areas of Cerrado, Caatinga and Atlantic Forest. Our study provides a valuable contribution to knowledge of the distribution of insect galls and the results obtained here can guide taxonomists and ecologists regarding the delineation of sampling areas as well as conservation strategies for this ecological interaction.

**Acknowledgements**

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

**References**


