ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE   BOTÁNICA

POTENTIAL DISTRIBUTION OF DOMESTICATED *Sechium edule* (CUCURBITACEAE) IN MEXICO

Distribución potencial de *Sechium edule* (Cucurbitaceae) domesticado en México

Luis Ángel BARRERA-GUZMÁN[1,2] iD; Jorge CADENA-IÑIGUEZ[2,3] iD*; Juan Porfirio LEGARIA-SOLANO[1] iD; Jaime SAHAGÚN-CASTELLANOS[1] iD; Gabriela RAMÍREZ-OJEDA[4] iD

1. Departamento de Fitotecnia, Universidad Autónoma Chapingo, Carr. México-Texcoco km 38.5, Texcoco, México.

2. Grupo Interdisciplinario de Investigación en *Sechium edule* en México (GISeM), 56153, Texcoco, Estado de México, México.

3. Colegio de Postgraduados, Campus San Luis Potosí, Salinas de Hidalgo, 78600 San Luis Potosí, Mexico

4. Campo Experimental Centro Altos de Jalisco, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Tepatitlán de Morelos, Jalisco, México.

* **For correspondence** jocadena@gmail.com

ABSTRACT

Mexico is the centre of origin of the chayote (*Sechium edule* Jacq. Sw), an important plant in human consumption and in pharmaceuticals. The objective of this study was to determine the potential distribution of domesticated *S. edule* in Mexico using seven species distribution algorithms, to efficiently manage S. edule resources and help its conservation by identifying patterns of geographic distribution. Otherwise, areas of high suitability can be used to produce improved seed at a lower cost. 162 GBIF occurrence points and nine layers in raster format were used to evaluate seven algorithms of species distribution models. To evaluate the reliability and performance of the models, the statistics Area Under the Curve (AUC) and true skill statistic was used. Predominant climate types were Cwb (33.3 %) and Aw (17.9 %); predominant soil types were leptosol (33.3 %) and phaozem (16.7 %). The seven models showed areas of high suitability (> 0.75) in Chiapas, Guerrero, Oaxaca, Veracruz, Tabasco, Puebla and Hidalgo states. AUC values for the seven models were > 0.8 and their performance was adequate (0.4 > TSS < 0.7). Classification tree analysis was found to be the best algorithm measured by AUC (0.90); however, the seven models were adequate to explain *S. edule* distribution in Mexico. *S. edule* climatic adaptability also allows to be distributed towards the Yucatan Peninsula and western Mexico. The distribution of *S. edule* in Mexico, according to the studied algorithms, is limited to total annual precipitation and temperature seasonality.
Keywords: algorithms, geographical distribution, performance, temperature, weather.

RESUMEN

México es el centro de origen del chayote (*Sechium edule* Jacq. Sw), una planta importante en la alimentación humana y en la farmacéutica. El objetivo de este estudio fue determinar la distribución potencial de *S. edule* domesticado en México utilizando siete algoritmos de distribución de especies, con el fin de gestionar eficientemente los recursos de *S. edule* y ayudar a su conservación mediante la identificación de patrones de distribución geográfica. Por otra parte, las zonas de alta idoneidad pueden ser utilizadas para la producción de semilla mejorada a un menor costo. Se utilizaron 162 puntos de ocurrencia de la GBIF y nueve capas en formato ráster para evaluar siete algoritmos de modelos de distribución de especies. Para evaluar la confiabilidad y el rendimiento de los modelos se utilizaron los estadísticos Área bajo la curva y el verdadero estadístico de habilidad. Los tipos de clima predominantes fueron Cwb (33,3 %) y Aw (17,9 %); los tipos de suelo predominantes fueron leptosol (33,3 %) y phaozem (16,7 %). Los siete modelos mostraron áreas de alta idoneidad (> 0,75) para los estados de Chiapas, Guerrero, Oaxaca, Veracruz, Tabasco, Puebla e Hidalgo. Los valores de AUC para los siete modelos fueron > 0,8 y su rendimiento fue adecuado (0,4> TSS <0,7). Se encontró que el análisis de árboles de clasificación fue el mejor algoritmo medido por AUC (0,90), sin embargo, los siete modelos fueron adecuados para explicar la distribución de *S. edule* en México. La adaptabilidad climática de *S. edule* también le permite distribuirse hacia la Península de Yucatán y occidente de México. La distribución de *S. edule* en México, según los algoritmos estudiados, se limita a la precipitación total anual y la estacionalidad de la temperatura.
Palabras Clave: algoritmos, clima, distribución geográfica, rendimiento, temperatura.

INTRODUTION

The chayote [*Sechium edule* (Jacq.) Swartz] is a native species to the mountainous regions of Mexico. It is currently found in the wild in the states of Veracruz and Oaxaca (Lira et al., 1999), where its cultivation is also practiced, and to a lesser extent in western and central Mexico. Because wild populations of chayote and the closely related species *Sechium compositum* (Donn. Sm) C. Jeffrey and *Sechium chinantlense* Lira & F. Chiang have only been found in this country, Mexico is considered the centre of origin of *S. edule* (Newstrom, 1990; Lira, 1995; Cross et al., 2006). The cultivated and wild populations of chayote in Mexico shows great phenotypic and phytochemical variability in the characteristics of fruit, leaves and flowers. *S edule* domesticated populations are used for food purposes, while wild populations are used for pharmaceutical research (Cadena-Iñiguez et al., 2008; 2011).

Mexico ranks first in the world production and export of chayote. Around 2972 ha are produced in the country at an average rate of 66 tons per hectare, with Veracruz state contributing more than 90 % of total production (SIACON, 2020). Chayote fruit is used in national gastronomy and its alcoholic extracts are being studied as a potential treatment for various types of cancer (Cadena-Iñiguez et al., 2013b; Aguiñiga-Sánchez et al., 2015; 2017; Salazar-Aguilar et al., 2017).

Ecological niche is the relationship between a species with all biotic and abiotic factors that affect it. Ecological niche modeling predicts the probability of suitability of a species in each geographical area, that is, the necessary or ideal conditions for growth and development of species, where values close to 0 indicate little or no suitability, while values close to 1 indicates high suitability (Ashraf et al., 2017). The distribution of a species is a function of the ecological conditions where it develops and reproduces. Species distribution models (SDM) are based on statistical procedures and Geographical Information Systems (GIS). Using presence-absence data, these models allow the inference of ecological niches. Various algorithms for their calculation are grouped into descriptive, discriminating, and mixed techniques (Mateo et al., 2011). Descriptive techniques such as Bioclim include bioclimatic envelop methods to characterise the sites found within environmental space occupied by the species in question. In this group there are also algorithms based on distances (Domain and Mahalanobis) that evaluate sites in terms of environmental similarity with presence data (Elith et al., 2006). Discriminant techniques such as generalized linear models (GLM) require presence-absence data and could model realistic ecological relationships. Within these techniques are those using machine learning, such as maximum entropy (Maxent), random forest (RF), and support vector machines (SVM). These are widely used due to their ability to process large amounts of information and reconstruct groups for classification, or to generate numerical regressions for prediction tasks (Gobeyn et al., 2019). There are also null geographic models such as geographic distance, convex hulls, circles, and presence-absence that use geographic location of presences, but they are not based on predictive variables and are not commonly used since their performance and reliability has yet to be assessed (Hijmans and Elith, 2013). Maxent algorithm is used more frequently (Phillips et al., 2006), however, more algorithms need to be evaluated for a more comprehensive view.

SDMs have various applications in agriculture. In cultivated species such as *Zea mays* and *Panicum virgatum*, Maxent and SVM algorithms have been used to estimate the suitability of geographic regions in which these species can thrive (Evans et al., 2010). Information from SDMs can also provide phylogeographic references and infer historical events that attempt to explain the origin and genetic structure of populations (Carstens and Richards, 2007), although this must also be corroborated with molecular data. SDMs are used to monitor the availability of pollinating insects, as well as to propose strategies to cope with any adverse changes in the environment that affect them (Polce et al., 2013). It is also possible to predict the suitability of noxious species (weeds and insects), forecast their location and distribution, and formulate strategies to mitigate socioeconomic damage in areas vulnerable to invasion (Lantschner et al., 2018). Other applications to be highlighted are the assessment of species richness, endemism, threatened habitats, protection and conservation of threatened species, diversity patterns, conservation studies, response to climate change, reintroduction of threatened species and the study of anthropic effects on the distribution of the species. Although many of these applications are used in animal species, they can also be applied to plant species (Mateo et al., 2011).

Area Under the Curve (AUC) is the most used statistic to measure the reliability of distribution models. This value corresponds to the probability that, by taking presence/absence data at random, the model assigns higher suitability values to presence data (Mateo et al., 2011). AUC has some limitations in its use, especially in species with a high range of adaptability (Lobo et al., 2007). However, through algorithms adjustments such as cross-validation, correlation analysis, Cohen's kappa, true skill statistic (TSS) and raising the number of presences to 20 or over, its consistency and performance can increase considerably (Hijmans and Elith, 2013).

The recalcitrance of *S. edule* seed does not allow it to be protected in *ex situ* germplasm banks, which is why suitable sites must be found to maintain and study *S. edule* accessions. Currently, work is being done on the creation of new germplasm banks to facilitate the access to genetic diversity for chayote producers, as well as to encourage the exchange of genetic material from farmers (Cadena-Iñiguez et al., 2013a).

The objective of this study was to determine the potential distribution of domesticated *S. edule* in Mexico using seven algorithms to calculate species distribution models, to efficiently manage the resources of *S. edule* and aid its conservation, identify geographical distribution patterns.

MATERIALS AND METHODS

OCCURRENCE DATA

S. edule occurrence data were obtained from the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/). Worldwide there are 2327 occurrence data, of which 983 are from Mexico. Data from these records were cleaned to remove entries with repeated coordinates, absence of coordinates, georeferencing errors and exclusion of wild specimens, leaving 162 occurrences data. SSDM package (Schmitt et al., 2017) selects pseudo-absences from randomly selected sites where a species is assumed to be absent. 25 % of the records were randomly selected for use in the test, while the remaining 75 % were used for training.

BIOCLIMATIC VARIABLES

The 19 bioclimatic variables of WorldClim version 2.1 from the period 1970–2000 was used, with a spatial resolution of 1 km^2 (Fick and Hijmans, 2017). Raster models with spatial resolution of 1 km2 was added for elevation data in meters (Fick and Hijmans, 2017) (Table 1). Köppen-Geiger climate classification (Beck et al., 2018) and soil types from Harmonized World Soil Database version 1.2 (Fischer et al., 2008) were used to group occurrence data. Raster values for 22 layers were obtained using Point Sampling Tool of QGIS version 3.16.2 (QGIS Development Team, 2020). Specific locations of occurrence data were found using Physiographic Subprovinces from Digital Map of Mexico, prepared by National Institute of Statistical and Geographic Information (INEGI) (https://www.inegi.org.mx/temas/fisiografia/).

ANALYSIS AND PROCESSING

All packages mentioned in this paper were run using RStudio (R Core Team, 2020). The qualitative variables of climate and soil types were processed using tidyverse package (Wickham et al., 2019) with data presented in bar graphs, showing descriptive information. They were not included in SDM since climate type is correlated with bioclimatic variables related to precipitation such as bio12 (Table 1) (Evans et al., 2010), and soil type generally does not seem to restrict distribution due to the adaptability of the crop to different soil types (Lira et al., 2019). Pearson's correlation was calculated for raster values of 19 WorldClim bioclimatic variables and elevation. Variables with correlation values greater than 0.8 were eliminated to avoid affecting the models (Phillips et al., 2006) (Table 1).

Table 1. List of variables used in the model development for *S. edule* domesticated in Mexico.

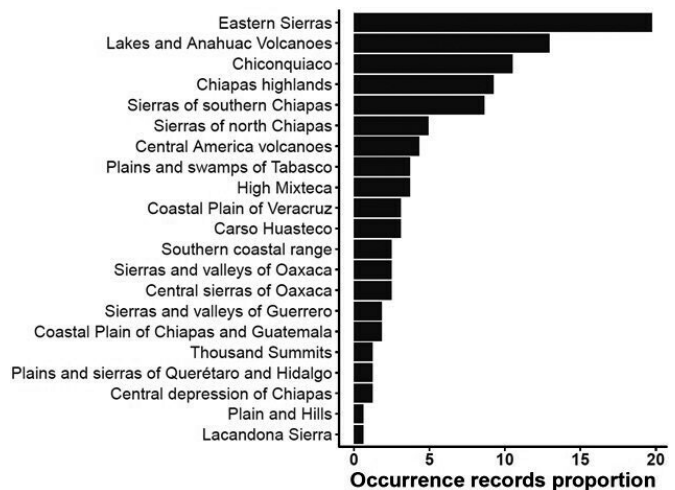| Variable	Description	Unit
Bio1	Annual Mean Temperature	°C
Bio2	Mean Diurnal Range	°C
Bio3	Isothermality	°C
Bio4	Temperature Seasonality	°C
Bio12	Annual Precipitation	Mm
Bio14	Precipitation of Driest Month	Mm
Bio15	Precipitation Seasonality	Mm
Bio18	Precipitation of Warmest Quarter	Mm
Elevation	Digital Elevation Model	M

Figure 1. Occurrence data percentage of *S. edule* for each physiographic subprovince (https://www.inegi.org.mx/temas/fisiografia/) in Mexico.

DISTRIBUTION MODELS

Seven common algorithms from literature were used: Generalized linear model that uses a weighted linear regression to arrive at the estimated maximum likelihood of the parameters (Hijmans and Elith, 2013); Generalized additive model, which is an algorithm that allows non-linear responses to be directly modeled by smoothing the predictive variables (Wood, 2017); Generalized boosted

regression model which is an algorithm that repeatedly fits many decision trees to improve model accuracy (Greenwell et al., 2020); Classification tree analysis is an algorithm that initiates a non-parametric structured approach for the evaluation and comparison of different options, thus helping to choose the most competitive alternative (Schmitt et al., 2017); Random Forest (RF) which is a regression-based algorithm in which bootstrap is used to select subsamples of the data and generate a large number of regression trees (Breiman, 2001); Maxent which is an algorithm that expresses a probability distribution where each cell of the grid has a predicted suitability of the conditions for the species (Phillips et al., 2006) and Vector Support Machines (SVM) (Vapnik, 1998) which uses a functional relationship called the kernel to map data into hyperspace where patterns can be represented more simply. Additionally, an assembly of the seven aforementioned models was executed through a weighted average of the outputs of the seven algorithms used (Schmitt et al., 2017). Likewise, dependencies gam (Hastie, 2020), stats (R Core Team, 2020), maxent (Hijmans et al., 2017), rpart (Therneau and Atkinson, 2019), gbm (Greenwell et al., 2020), randomForest (Liaw and Wiener, 2002) and e1071 (Meyer et al., 2019) were used. Threshold used was 0.5.

Regarding Maxent algorithm, ENMeval package (Muscarella et al., 2014) was used to optimize the model, hinge and threshold functions were deactivated to avoid overfitting response curves, the regularization multiplier was 0.75 and were activated linear, quadratic and product functions. For running the models, the protocol of Hijmans and Elith (2013) was followed. The model's adjustment was measured using AUC, where values > 0.5 indicate adequate models. However, AUC may present bias and variation in its values with the spatial extension to select bottom points (Lobo et al., 2007). To correct this problem, point distance sampling was used. This involved calculating the distance from the presence records for the training and test data sets, as well as the distance from the absence data for the training and test data sets (Hijmans, 2012; Hijmans and Elith, 2013). To evaluate the performance of the models the TSS was used, where TSS > 0.75 indicates excellent performance, 0.4 < TSS < 0.75 optimal performance and TSS < 0.4 poor performance (Eskildsen et al., 2013). The performance of the models and AUC and TSS calculation were carried out using SSDM package (Schmitt et al., 2017). The output raster layers were exported to QGIS Development Team version 3.16.2 (2020) for final editing of distribution model maps.
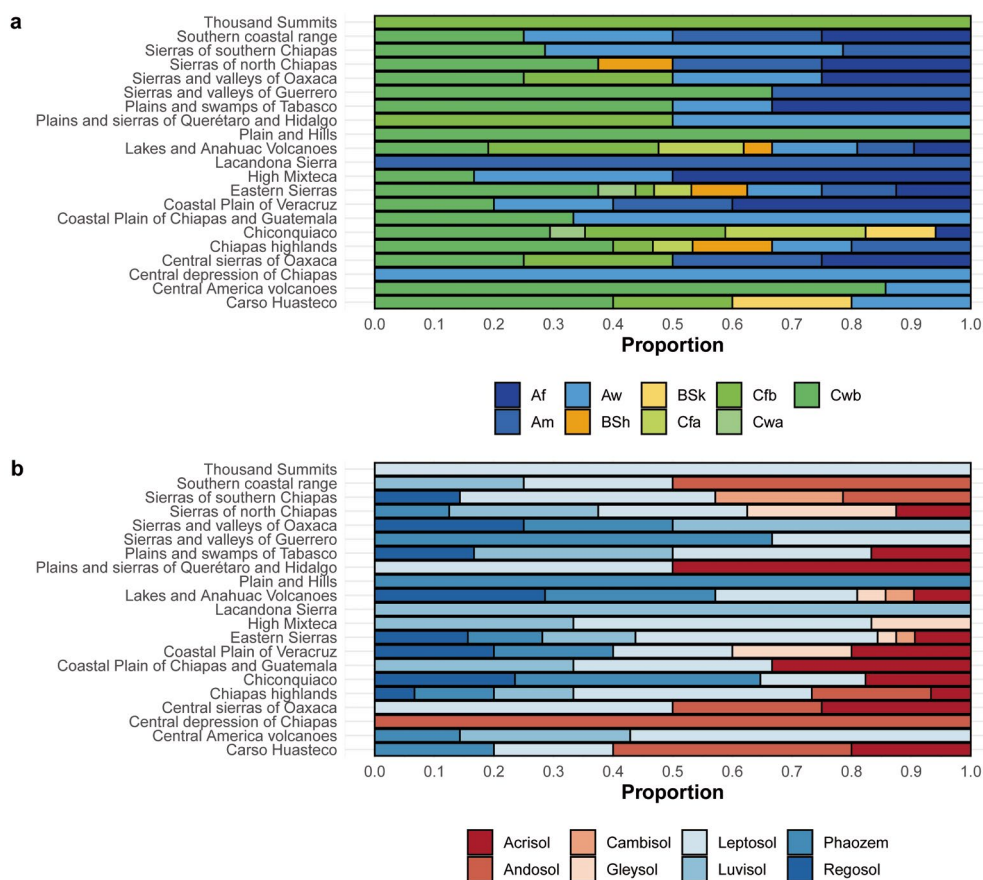
Figure 2. Physiographic subprovinces of Mexico. a. Climate types. Af = Tropical, rainforest; Am = Tropical monsoon; Aw = Tropical savannah; BSh = Arid, step, hot; BSk = Arid, step, cold; Cfa = Temperate, no dry season, hot summer; Cfb = Temperate, no dry season, warm summer; Cwa = Temperate, dry winter, hot summer; Cwb = Temperate, dry winter, warm summer. b. Soil types.

Table 2. Jacknife test for contribution (%) of variables to algorithms and statistics AUC and TSS.

Model	Bio1	Bio2	Bio3	Bio4	Bio12	Bio14	Bio15	Bio18	Elevation	AUC	TSS
GLM	4.35	13.99	12.38	4.19	47.34	0.04	0.17	16.34	1.16	0.82	0.63
GAM	15.30	1.64	9.78	26.77	2.73	16.03	14.03	10.14	3.35	0.83	0.64
GBM	1.15	9.89	0.08	85.45	9.89	0.12	3.03	0.05	0.07	0.87	0.74
CTA	0.34	0.06	4.27	46.54	0.07	0.07	13.99	21.16	13.66	0.90	0.89
RF	8.34	6.09	7.64	29.65	10.91	5.64	17.56	5.16	8.98	0.87	0.74
Maxent	11.18	11.22	11.25	10.89	11.21	11.19	10.63	11.24	11.15	0.85	0.70
SVM	5.01	5.17	10.41	10.75	38.20	3.23	16.14	4.36	6.68	0.87	0.74
Ensemble	7.49	7.67	7.03	25.17	14.20	5.36	16.21	8.19	8.62	0.84	0.68

GLM: Generalized lineal model; GAM: Generalized additive model; GBM: Generalized boosted regression model; CTA: Classification tree analysis; RF: Random Forest, SVM: Support vector machines. bio1: annual mean temperature; bio2: mean diurnal range (mean of monthly (max temp - min temp)); bio3: isothermality; bio4: temperature seasonality; bio12: annual precipitation, bio14: precipitation of driest month; bio15: precipitation seasonality; bio18: precipitation of warmest quarter; AUC: Area under the curve, TSS: True skill statistic.

RESULTS

PHYSIOGRAPHY, CLIMATE, AND SOIL TYPES

Twenty-one physiographic subprovinces were registered for the 162 occurrence records, most of which correspond to the physiographic subprovinces of Eastern Sierra (20 %), Lakes and Anahuac volcanoes (13 %), Chiconquiaco (10.5 %), Chiapas highlands (9.5 %) and Sierras of southern Chiapas (8.5 %) (Fig. 1). Three climatic groups were found to be present in the occurrence records of *S. edule* according to the Köppen classification: tropical (A), dry (B) and temperate (C). The predominant type of climate was Cwb (temperate, dry winter, warm summer, 33.3 %), also known as temperate mountain with dry winter, followed by Aw (tropical monsoon, 17.9 %) and Af (tropical, rainforest, 11.7 %). In the physiographic subprovinces of Lakes and Anahuac volcanoes and Eastern Sierras, up to seven and eight different climatic types were seen, respectively. In Plains and Hills and the Sierra Lacandona, with few points of occurrence, the climates Cwb and Am predominated, respectively (Fig. 2a). Eight predominant soil groups are present in the occurrence records of *S. edule*, according to raster information from Fischer et al. (2008). Leptosol contributes 33.3 % of the total, phaozem 16.7 %, followed by luvisol and regosol, both contributing 13 % y 12.3 %, respectively. In some physiographic subprovinces, only one type of soil and climate predominated (Fig. 2).
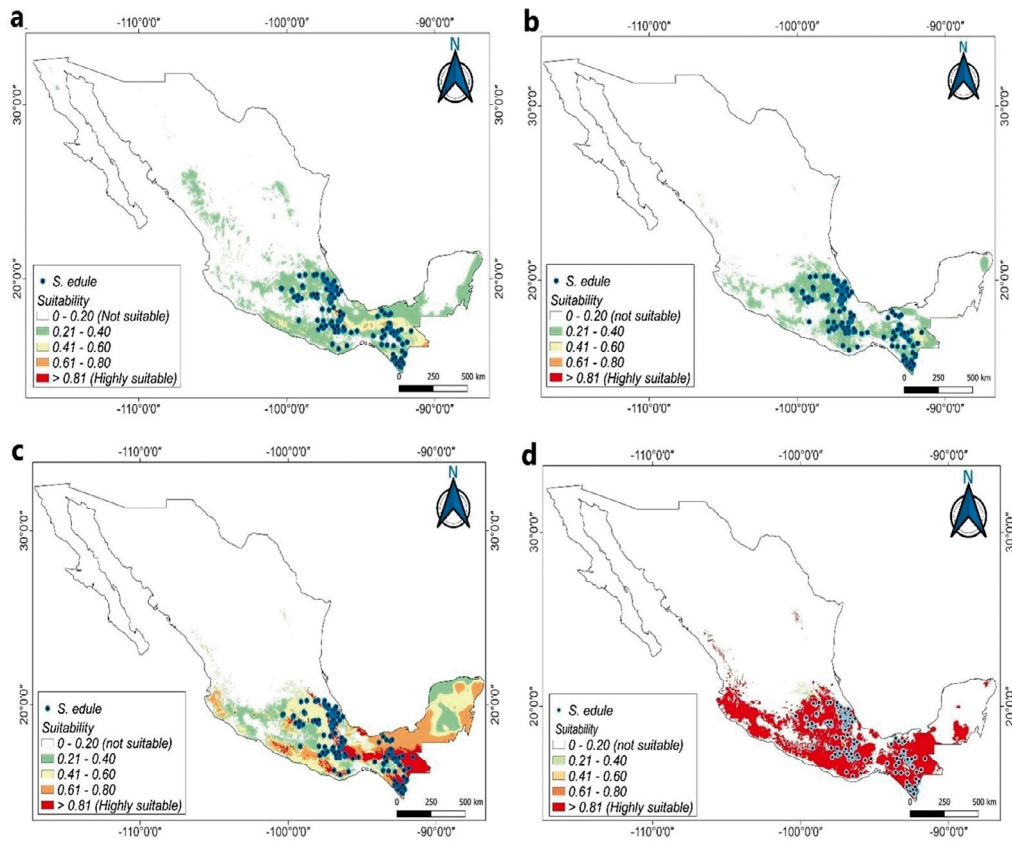
DISTRIBUTION MODELS

The distribution models of *S. edule* in Mexico were inferred by layers bio1, bio2, bio3, bio4, bio12, bio14, bio15, bio18, and elevation, which were derived from correlation analysis. The AUC and TSS values for training and testing

were averaged because they are very similar values. The CTA algorithm obtained the highest AUC score (0.90) In general, the AUC for the seven models was > 0.5 and they showed adequate performance (0.4 < TSS > 0.7) (Table 2).

The GLM algorithm (Fig. 3a) showed low suitability (0.21-0.40) for the states of Chiapas, Oaxaca, Guerrero, Puebla, Hidalgo y Veracruz. GAM model (Fig. 3b) showed results very similar to GLM algorithm. GBM algorithm showed areas of high suitability for regions of Chiapas, Oaxaca, Veracruz, the Yucatan peninsula and the western coastal zone of Mexico. (Fig. 3c). CTA algorithm showed areas of high suitability for the Mesoamerican region of Mexico (Fig. 3d). Surprisingly, moderate suitability was found in small regions in the north of the country, where the weather is generally drier with little precipitation. This can be explained by the coincidence of the clime arid, steppe and hot (BSh) and clime arid, steppe and cold (BSk) weather in certain occurrence records.

RF, Maxent and SVM algorithms showed very similar results in terms of suitable areas for *S. edule* (Fig. 4 a-c). The RF and MVS algorithms (Fig. 4a–c) showed similar maps, with areas of high suitability like those seen in the Maxent model but with smaller regions. The Maxent algorithm (Fig. 4b) has high suitability (> 0.81) zones similar to the RF and SVM models, however the high suitability was emphasised in large regions of Veracruz, Puebla, Hidalgo, Guerrero, Oaxaca and Chiapas. The assembly of the seven algorithms (Fig. 4d) showed zones of high suitability very similar to Maxent algorithm (Fig. 4b).

According to Jacknife test, the generalized algorithms (GLM and GAM) and SVM gave greater weight to variables related to precipitation (bio12 and bio14); while the GBM, CTA, RF algorithms had more influence of variables related to temperature (bio4). Maxent algorithm gave a fair percentage to all bioclimatic variables. The assembled model gave more importance to bio4 (Table 2).

Figure 3. Distribution models for *S. edule* in Mexico. a. Generalized lineal model. b. Generalized additive model. c. Generalized boosted regressions model. d. Classification tree analysis.
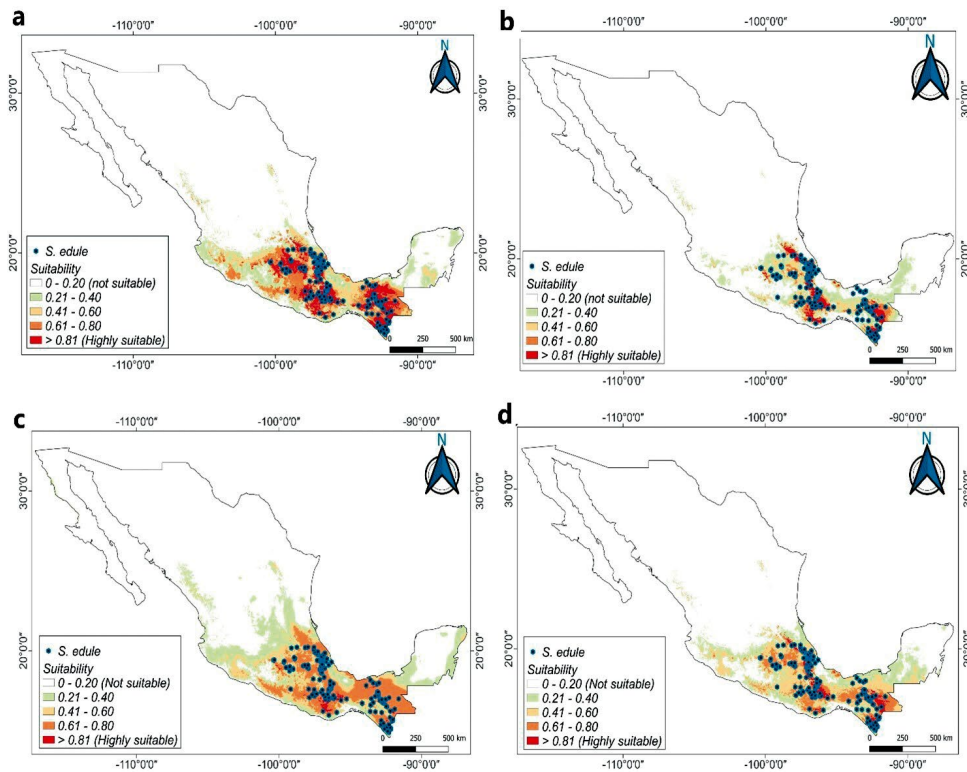
Figure 4. Distribution models for *S. edule* in Mexico. a. Random Forest. b. Maxent. c. Support vector machines. d. Ensemble.

DISCUSION

Annual mean temperature plays an important role in *S. edule* distribution, with an average of 20 °C across the 162 occurrences data. The optimum temperature range is 13–21 °C, as temperatures below this cause damage to the fruits, while those above 28 °C cause problems in flower and fruit retention (Lira, 1996). Wild populations of *S. edule* generally tend to be distributed in Am and Af climates, however, under growth conditions, cultivated *S. edule* can thrive in a diverse range of dry and temperate climates (Aung et al., 1990). The calcareous components of leptosols can immobilise the mineral nutrients in the soil and, in the case of chayote, can reduce the productivity and quality of the fruits (Cadena-Iñiguez, 2005). However, some leptosols may have small layers rich in organic matter, which is an edaphic requirement for chayote plantations (Lira, 1996). Phaozem, present in 50 % of physiographic subprovinces, are soils considered ideal for agriculture due to their high organic matter content. Ideal soils for chayote production must be well drained, rich in organic matter and have a pH of 4.5–6.5 (Lira, 1996), however, the different phenotypic variants of chayote enable a high level of adaptability to different types of soil (Lira et al., 2019).

High suitability zones (> 0.76) shown in the seven algorithms coincided with the presence of two species that are closely related to *S. edule*: *S. chinantlense* and *S. compositum* (Cross et al., 2006, Barrera et al., 2021). The varietals *virens levis*, *nigrum spinosum*, *nigrum xalapensis*, *albus dulcis* and *albus minor* are also distributed in these regions (González-Santos et al., 2017). When there are different ecotypes or populations of the same species, the species distribution algorithms can be considered representative of the entire species (Mateo et al., 2011). Although distribution models are applicable for wild species, *S. edule* populations are domesticated and very few of them are cultivated, for example, the case of varietal complexes *virens levis* and *nigrum xalapensis*. However, the before mentioned varietal complexes of *S. edule* are vulnerable in these same areas. The risk of genetic erosion in these places is mainly due to the presence of new improved varieties that are displacing semi-domesticated populations, which generally have fruit with morphological characteristics not suitable for the market, such as the presence of thorns, small size, and bitter flavour. Also, the introduction of crops such as coffee causes a decrease in *S. edule* populations (Cadena-Iñiguez, 2005). In 2007, to counteract the effects of genetic erosion, the *Sechium* spp. germplasm bank was created in Huatusco, Veracruz (a geographic site in an area of high suitability according to this study), where domesticated and wild variants from different parts of Mexico and Central America are protected *in situ* (Cadena-Iñiguez and Arévalo-Galarza, 2011).

In this study, only occurrence data pertaining to cultivated *S. edule* were considered. However, in a study by Lira et al. (2018) using Maxent algorithm, only occurrence data for wild *S. edule* were used and their results are very similar to those from this study using Maxent model. This is because ecological requirements of cultivated and wild *S. edule* are similar. The evolutionary history of *S. edule* can complement the analysis of distribution models. The first wild populations of *S. edule* originated in the states of Guerrero, Chiapas, and Veracruz. Hybridisation between populations and varietals of *S. edule*, as well as with *S. chinantlense* and *S. compositum* species, together with adaptive specialisation, have led to new phenotypic variants adapted to different environmental conditions and thus have dispersed to new regions with human help.

Regarding *S. edule*, González-Santos et al. (2017) evaluated 106 occurrence data belonging to five chayote varietals, which were modelled using the 19 bioclimatic layers of World Clim and the Maxent algorithm to forecast the effect of climate change on the distribution of these varietals in the year 2050. Some varieties such as *nigrum spinosum* will lose around 50% of their current distribution, while others such as *virens levis* will maintain and slightly increase their distribution, since this varietal is the very valuable and is commonly used in breeding (Cadena-Iñiguez and Arévalo-Galarza, 2011; Cadena-Iñiguez et al., 2013a). Otherwise, most of the chayote varietals are found semi-domesticated in domestic gardens and cloud forest areas, the latter being highly disturbed by human activities (Lira, 1995).

Sechium edule is a species that can thrive in different climatic conditions. Because of this, the AUC values can be close to one, but are unreliable, as the range of values for each variable obtained from large databases is very wide. Although the chayote varietals differ morphologically, the differences are not very noticeable when the values of the bioclimatic variables are compared (González-Santos et al., 2017). Therefore, the models used may be representative of all *S. edule populations* (Mateo et al., 2011). Species distribution can also be limited by other factors that must be considered, but it can be difficult to incorporate into distribution models. For example, competition with other species, pollination and human influence can all affect distribution. It is necessary to evaluate different models to gain a broad perspective on the distribution of species.

CONCLUSIONS

Cultivated populations of *S. edule* thrive in a wide range of climate and soil types. CTA was found to produce the best distribution model according to the AUC statistics. However, all seven models were relatively adequate to explain the distribution of *S. edule* in Mexico. The areas of high suitability were found in regions of the states of Chiapas, Guerrero, Oaxaca, Veracruz, Tabasco, Puebla, and Hidalgo, which should be considered as high priority areas for conservation and breeding. The distribution of domesticated *S. edule* in Mexico, according to the algorithms studied, is limited to total annual precipitation and temperature seasonality.

Climatic adaptability of *S. edule* also allows it to be distributed towards the Yucatan Peninsula and western Mexico.

DISCLOSURE OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Aguiñiga-Sánchez, I., Cadena-Íñiguez, J., Santiago-Osorio, E., Gómez-García, G., Mendoza-Núñez, VM., Rosado-Pérez, J., Ruíz-Ramos, M., Cisneros-Solano, V. M., Ledesma-Martínez, E., Delgado-Bordonave, A. J. and Soto-Hernández, R. M. (2017). Chemical analyses and in vitro and in vivo toxicity of fruit methanol extract of *Sechium edule* var. *nigrum spinosum*. *Pharm Biol*, *55*(1), 1638-1645. https://doi.org/10.1080/13880209.2017.1316746

Aguiñiga-Sánchez, I., Soto-Hernández, M., Cadena-Iñiguez, J., Ruíz-Posadas, L. M., Cadena-Zamudio, J. D., González-Ugarte, A. K., Steider, B. W. and Santiago-Osorio, E. (2015). Fruit extract from a *Sechium edule* hybrid induce apoptosis in leukaemic cell lines but not in normal cells. *Nutr Cancer, 67*(2), 250-257. https://doi.org/10.1080/01635581.2015.989370

Ashraf, U., Peterson, A. T., Chaudhry, M. N., Ashraf, I., Saqib, Z., Ahmad, S. R., and Ali, H. (2017). Ecological niche model comparison under different climate scenarios: a case study of *Olea* spp. in Asia. *Ecosphere*, *8*(5), e01825. https://doi.org/10.1002/ecs2.1825

Aung, L. H., Ball, A. and Kushad, M. (1990). Developmental and nutritional aspects of chayote (*Sechium edule*, Cucurbitaceae). *Economic Botany, 44*(2), 157-164. https://doi.org/10.1007/BF02860483

Barrera-Guzmán, L. A., Legaria-Solano, J. P., Cadena-Iñiguez, J. and Sahagún-Castellanos, J. (2021). Phylogenetic relationships among Mexican species of the genus *Sechium* (Cucurbitaceae). *Turkish Journal of Botany, 45*, 302-314. https://doi:10.3906/bot-2007-18

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data, 5*(1), 1-12. https://doi.org/10.1038/sdata.2018.214

Breiman, L. (2011). Random Forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Cadena-Iñiguez, J., Soto-Hernández, M., Arévalo-Galarza, M. L., Avendaño-Arrazate, C. H., Aguirre-Medina, J. F. and Ruíz-Posadas, L. M. (2011). Biochemical characterization of domesticated varieties of chayote *Sechium edule* (Jacq.) Sw. fruits compared to wild relatives. *Revista Chapingo. Serie Horticultura, 17*(2), 44-55.

Cadena-Iñiguez, J. (2005). Caracterización morfoestructural, fisiológica, química y genética de diferentes tipos de chayote *(Sechium edule)*. [PhD thesis]. Colegio de Postgraduados, Campus Montecillo, México.

Cadena-Iñiguez, J. and Arévalo-Galarza, M. L. C. (2011). *Las variedades de chayote (Sechium edule (Jacq.) Sw.) y su comercio mundial*. bba.

Cadena-Iñiguez, J., Avendaño-Arrazate, C. H., Soto-Hernández, M., Ruíz-Posadas, M. L., Aguirre-Medina, J. F. and Arévalo-Galarza, M. L. C. (2008). Infraspecific variation of *Sechium edule* (Jacq.) Sw. in the state of Veracruz, Mexico. *Genetic Resources and Crop Evolution, 55*(6), 835-847. https://doi.org/10.1007/s10722-007-9288-4

Cadena-Iñiguez, J., Soto-Hernández, M., Arévalo-Galarza, M. L. C., Avendaño-Arrazate, C. H. and Aguirre-Medina, J. F. (2013a). *Modelos de mejoramiento genético participativo en chayote (*Sechium *spp)*. Editorial del Colegio de Postgraduados.

Cadena-Iñiguez, J., Soto-Hernández, M., Torres-Salas, A., Aguiñiga-Sánchez, I., Ruíz-Posadas, L. M., Rivera-Martínez, A. R., Avendaño-Arrazate, C. H. and Santiago-Osorio, E. (2013b). The antiproliferative effect of chayote varieties (*Sechium edule* Jacq.) Sw.) on tumour cell lines. *Journal of medicinal plants research*, *7*(8), 455-460. https://doi.org/10.5897/JMPR12.866

Carstens, B. C., and Richards C. L. (2007). Integrating coalescent and ecological niche modelling in comparative phylogeography. *Evolution, 61*(6), 1439-1454. https://doi.org/10.1111/j.1558-5646.2007.00117.x

Cross, H., Lira, S. R., and Motley, T. J. (2006). Origin and diversification of chayote. In T. J. Motley, N. Zerega, and H. Cross, (Eds.). *Darwin's harvest: new approaches to the origins, evolution, and conservation of crops* (pp. 171-194). Columbia University Press. https://doi.org/10.7312/motl13316

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M. and Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A... Zimmermann, N. E. (2006). Novel methods improve prediction of species distributions from occurrence data. *Ecography*, *29*(2), 129-151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Eskildsen, A., Roux, P. C., Heikkinen, R. K., Høye, T. T., Kissling, W. D., Pöyry, J., Wisz, M. S., and Luoto, M. (2013). Testing species distribution models across space and time: high latitude butterflies and recent warming. *global ecology and biogeography, 22*(12), 1293-1303. https://doi.org/10.1111/geb.12078

Evans, J. M., Fletcher Jr., R. J., and Alavalapati, J. (2010). Using species distribution models to identify suitable areas for biofuel feedstock production. *GCB Bioenergy, 2*(2), 63-78. https://doi.org/10.1111/j.1757-1707.2010.01040.x

Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology, 37*(12), 4302-4315. https://doi.org/10.1002/joc.5086

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H. T., Verelst, D., and Wiberg, D. (2008). *Global agroecological zones assessment for agriculture.* IIASA, Luxemburg, Austria, and FAO.

Gobeyn, S., Mouton, A. M., Cord, A. F., Kaim, A., Volk, M., and Goethals, P. L. M. (2019). Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. *Ecological Modelling, 392*, 179-195. https://doi.org/10.1016/j.ecolmodel.2018.11.013

González-Santos, R., Cadena-Íñiguez, J., Morales-Flores, F. J., Ruiz-Vera, V. M., and Pimentel-López, J. (2017). Prediction of the effects of climate change on *Sechium edule* (Jacq.) Swartz varietal groups in Mexico. *Genetic Resources and Crop Evolution*, 64(4), 791-804. https://doi.org/10.1007/s10722-016-0401-4

Greenwell, B., Boehmke, B., and Cunningham, J. (01 de enero de 2020). Gbm: Generalized boosted regression models. R package version 2.1.8. https://CRAN.R-proyect.org/package=gbm

Hastie, T. (01 de enero de 2020). gam: Generalized additive models. R package version 1.20. https://CRAN.R-project.org/package=gam

Hijmans, R. J. (2012). Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null-model. *Ecology*, 93(3), 679-688. https://doi.org/10.1890/11-0826.1

Hijmans, R. J., and Elith, J. (01 de enero de 2013). Species distribution modelling with R. https://rspatial.org/raster/sdm/

Hijmans, R. J., Phillips, S. J., Leatwick, J., and Elith, J. (01 de enero de 2017). Dismo: species distribution modelling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M. M., Peterson, A. T., ... Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Lantschner, V., Vega, G., and Corley, J. (2018). Predicting the distribution of harmful species and their natural enemies in agricultural, livestock and forestry systems: An overview. *International Journal of Pest Management*, 65, 1-17. https://doi.org/10.1080/09670874.2018.1533664

Liaw, M., and Wiener, M. (2002). Classification and regression by RandomForest. R News, 2(3), 18-22.

Lira, R. (1996). Chayote. *Sechium edule* (Jacq.) Sw., *promoting the conservation and use of underutilized and neglected crops. Institute of Plant Genetics and Crop Plant* Research. Gatersleben/International Plant Genetic Resources Institute.

Lira, R., Eguiarte, L. E., and Montes-Hernández, S. (2019). *Proyecto recopilación y análisis de la información existente de las especies de los géneros Cucurbita y Sechium que crecen y/o se cultivan en México.* Universidad Nacional Autónoma de México

Lira, R., Castrejón, J., Zamudio, S., and Roja-Zenteno, C. R. (1999). Propuesta de ubicación taxonómica para los chayotes silvestres (*Sechium edule*, Cucurbitaceae) de México.*Acta Botánica Mexicana, 49*, 47-61.

Lira, S. R. (1995). Estudios taxonómicos en el género Sechium P. Br. Cucurbitaceae [PhD Tesis]. Universidad Nacional Autónoma de México.

Lira, S. R., Sosa-Guzmán, L. J., Cabral-Soto, I. V., and Téllez, O. V. (2018). *Distribución potencial de las especies silvestres de la familia Cucurbitaceae de México* (Informe final SNIB-CONABIO Proyecto No. JM003). Universidad Nacional Autónoma de México. Facultad de Estudios Superiores Iztacala.

Lobo, J. M., Jiménez-Valverde, A., and Real, R. (2007). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17(2),145-151.
https://doi.org/10.1111/j.1466-8238.2007.00358.x

Mateo, R. G., Felicísimo, A. M., and Muñoz, J. (2011). Species distributions models: A synthetic revision. *Revista Chilena de Historia Natural, 84*(2), 217-240.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (01 de enero de 2020). e1071: Misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien. R package version 1.7-4. https://CRAN.R-proyect.org/package=e1071

Muscarella, R., Galante, P. J., Soley-Guardia, L., Boria, R. A., Kass, J., Uriarte, M., and Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for ecological niche models. *Methods in Ecology and Evolution, 5*(11), 1198-1205. https://doi.org/10.1111/2041-210X.12261

Newstrom, L. E. (1990). Origin and evolution of chayote, *Sechium edule*. In D. M. Bates, R. W. Robinson, C. Jeffrey, (Eds). *Biology and Utilization of the Cucurbitaceae* (pp. 141-149). Cornell University Press.

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling, 190*(3-4), 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Polce, C., Termansen, M., Aguirre-Gutiérrez, J., Boatman, G., Budge, G., Pietravalle, S., Potts, S., Ramírez, J., Somerwill, K., and Biesmeijer, J. (2013). Species distribution models for crop pollination: A modelling framework applied to Great Britain. *PLoS One, 8*, e76308. https://doi.org/10.1371/journal.pone.0076308

QGIS Development Team. (01 de enero de 2020). *QGIS Geographic Information System. Version 3.16.2 Open-Source Geospatial Foundation Project*. https://www.qgis.org/en/site/

R Core Team. (01 de enero de 2020). *R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna*. https://www.R-project.org

Salazar-Aguilar, S., Ruiz-Posadas, L. D. M., Cadena-Iñiguez, J., Soto-Hernández, M., Santiago-Osorio, E., Aguiñiga-Sánchez, I., Rivera-Martínez, AR. and Aguirre-Medina J. F. (2017). *Sechium edule* (Jacq.) Swartz, a new cultivar with antiproliferative potential in a human cervical cancer HeLa cell line. *Nutrients, 9*(8), E798. https://doi.org/10.3390/nu9080798

Schmitt, S., Pouteau, R., Justeau, D., de Boissieu, F., and Birnbaum, P. (2017). SSDM: an R package to predict distribution of species richness and endemism based on stacked species distribution models. *Methods in Ecology and Evolution, 8*(12), 1795-1803. https://doi.org/10.1111/2041-210X.12841

SIACON. (01 de enero de 2020). Sistema de Información Agroalimentaria de Consulta Nueva Generación. SIAP. https://www.gob.mx/siap

Therneau, T., and Atkinson, B. (01 de enero de 2019). rpart: Recursive partitioning and regression trees. R package version 4.1-15. https://CRAN.R-project.org/package=rpart

Vapnik, V. N. (1998). *Statistical Learning Theory*. Jhon Wiley & Sons.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., and Takahashi, K. (2019). Welcome to the Tidyverse. *J Open Source Softw, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.