CONSERVATION

REVISTA DE Biología Tropical

https://doi.org/10.15517/sy4jvh88

Biogeographical analysis of the Central American clade of *Sechium* (Cucurbitaceae)

Luis Ángel Barrera-Guzmán[1,2]; https://orcid.org/0000-0001-8057-2583
Jorge Cadena-Iñiguez[2,3*]; https://orcid.org/0000-0002-6427-0646
Juan Porfirio Legaria-Solano[4]; https://orcid.org/0000-0002-1371-9482
Víctor Manuel Cisneros-Solano[1,2]; https://orcid.org/0000-0001-8262-9109
Kazuo, N. Watanabe[5]; https://orcid.org/0000-0003-1499-9989
Daniel Alejandro Cadena-Zamudio[2]; https://orcid.org/0000-0002-6972-7414

1. Universidad Autónoma Chapingo, Centro Académico Regional sede Huatusco-Veracruz, México C.P. 94100, Carretera-Federal Huatusco-Xalapa km 6.5; luisangelbg@gmail.com, vcisneross@chapingo.mx
2. Grupo Interdisciplinario de Investigación en *Sechium edule* en México (GISeM), Texcoco, C.P. 56 160, México; jocadena@gmail.com (*Correspondencia), cadenazamudio@gmail.com
3. Colegio de Postgraduados, Campus San Luis Potosí, 78 622, Salinas de Hidalgo, San Luis Potosí, México; jocadena@gmail.com
4. Universidad Autónoma Chapingo, Carretera México-Texcoco, C.P. 56 230, Chapingo, Estado de México; legarias.juan@yahoo.com
5. Tsukuba Plant Innovation Research Center, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki Prefecture, 305-8 571, Japan; nabechanknw@gmail.com

Received 09-IX-2024. Corrected 04-XII-2024. Accepted 02-IX-2025.

ABSTRACT

Introduction: The genus *Sechium* P. Brown (Cucurbitaceae) includes 11 species, of which two are domesticated and nine grow in the wild. The Central American clade of *Sechium* has six species distributed in Panama and Costa Rica. These species have characteristics that can be transferred from wild to domesticated species.

Objective: To use three machine learning stacking algorithms and multivariate tools to describe geographic distribution, diversity degree, and endemism, to identify major conservation areas and to promote research for the improvement of the domesticated species.

Methods: Two hundred and nine occurrence records were retrieved from the Global Biodiversity Information Facility. Raster values extracted from 21 bioclimatic variables were analyzed with descriptive and multivariate statistics. The species distribution algorithms were assembled with the SSDM library from R software.

Results: Most species are distributed in type A and C climates, mainly in volcanic soils, with abundant organic matter. These species can grow at altitudes exceeding 2 000 m and tolerate low temperatures and high humidity levels. K-medoids established two groups and a 0.39 average silhouette coefficient, which indicates a low clustering trend. The stacked distribution models recorded good performance in areas under the curve (AUC) (> 0.75) and true skill statistic (> 0.75).

Conclusions: The main variables that supported the models were elevation, soil types, and precipitation. The main endemism and species diversity areas were in the Cordillera de Talamanca, the Cordillera de Guanacaste, the Cordillera de Tilarán, and the Central Volcanic Range (Costa Rica). These species thrive under similar environmental conditions; however, the diverse areas have significantly different precipitation and soil types.

Key words: domesticated species; machine learning; diversity; endemism; soil types.

RESUMEN

Análisis biogeográfico del clado centroamericano de *Sechium* (Cucurbitaceae)

Introducción: El género *Sechium* P. Brown (Cucurbitaceae) incluye 11 especies, de las cuales dos son domesticadas y el resto son silvestres. El clado centroamericano de *Sechium* incluye seis especies distribuidas en Panamá y Costa Rica. Estas especies tienen características que pueden ser de utilidad y transferibles de especies silvestres a especies modificadas.

Objetivo: Ejecutar tres algoritmos de aprendizaje automático apilados y herramientas multivariadas para describir la distribución geográfica, medir el grado de diversidad y endemismos de las especies centroamericanas de *Sechium* para identificar áreas de conservación y promover la investigación para el mejoramiento de especies cultivadas del género.

Métodos: Doscientos nueve puntos de ocurrencia fueron extraídos de la *Global Biodiversity Information Facility*. Los valores ráster se obtuvieron a partir de 21 variables bioclimáticas, donde se analizaron con estadística descriptiva y multivariada. Los modelos de distribución de especies apiladas fueron ejecutados con la librería SSDM del software R.

Resultados: La mayoría de las especies se distribuyen en climas tipo A y C, principalmente en suelos volcánicos con abundante materia orgánica. Estas especies prosperan en altitudes superiores a los 2 000 m y toleran bajas temperaturas con altos índices de humedad. El análisis k-medoides estableció dos grupos con un coeficiente de la silueta de 0.39, el cual indica una baja tendencia al agrupamiento. Los modelos de distribución apilados tuvieron buenos rendimientos en términos de área bajo la curva y del estadístico de habilidad verdadera (> 0.75).

Conclusiones: Las principales variables que apoyaron los modelos fueron la elevación, los tipos de suelo y la precipitación. Las principales áreas de endemismo y diversidad de especies se ubican en la Cordillera de Talamanca, la Cordillera de Guanacaste, la Cordillera de Tilarán y la Cordillera Volcánica Central (Costa Rica). Estas especies prosperan en condiciones ambientales similares; sin embargo, las diferentes áreas tienen precipitaciones y tipos de suelo significativamente diferentes.

Palabras clave: especies domesticadas; aprendizaje automático; diversidad; endemismo; tipos de suelo.

INTRODUCTION

The genus *Sechium* P. Brown includes 11 species and five of these are distributed in Mexico: *S. edule* Jacq. Swartz, *S. compositum* (Donn. Sm.) C. Jeffrey, *S. chinantlense* Lira & F. Chiang, *S. hintonii* (P.G. Wilson) C. Jeffrey, and *S. mexicanum* Lira & M. Nee. This study focuses on the remaining six species, which are mainly distributed in Costa Rica and Panama: *S. tacaco* (Pittier) C. Jeffrey, *S. venosum* (L.D. Gómez) Lira & F. Chiang, *S. villosum* (Wunderlin) C. Jeffrey, *S. pittieri* (Cogn.) C. Jeffrey, *S. panamense* (Wunderlin) Lira & F. Chiang, and *S. talamancense* (Wunderlin) C. Jeffrey. According to their morphological, geographic, and molecular characteristics, these species are divided into two groups: Mexican and Central American clades (Barrera-Guzmán et al., 2021; Cross et al., 2006; Lira et al., 1997; Lira & Nee, 1994; Monge & Loría, 2017; Sebastian et al., 2012; Wunderlin, 1976). Lira (1995) and other authors have provided valuable morphological information about the Central American clade of *Sechium* and its possible contribution to *S. edule* and *S. tacaco*, the only two species of this genus that have been domesticated. Nevertheless, the biogeographical information about the Central American species is scarce. Geographic Information Systems (GIS) currently provide data on the climate and ecological variables of a given territory. These tools, along with multivariate analyses, are essential for such studies about conservation, localization, potential distribution models, and endemism (Mateo et al., 2011).

The species distribution models (SDM) are based on different machine learning algorithms (Schmitt et al., 2017). One of the most used algorithms is MaxEnt or maximum entropy (Phillips, 2010); however, MaxEnt has some deficiencies, because the bias in sampling or the spatial resolution can provide arbitrary results. Choosing the appropriate configuration requires advanced knowledge, hindering the use of qualitative variables for evaluation purposes (Baldwin, 2009). Stacked SDMs involve compiling fundamental information from each algorithm and subsequently merging it. Although the output raster layer provides strong results, they must be assessed with the area under the curve (AUC-ROC), Cohen's Kappa calculator, and the True Skill Statistic (TSS).

The stacked SDMs are fundamental to determine the biogeographical and ecological characteristics of a species. For instance, they provide information about the species' prevalence, diversity, and degree of endemism (Gelfand, 2022). They also help to establish environmental variables influencing the taxa distribution (Bedair et al., 2023). Other outstanding applications include development of climate change models for spatial biodiversity predictions that support forest restoration (Zwiener & Alves, 2023); research about the climate change-related phenological variation of plants (Bayliss et al., 2022); and biodiversity patterns and assembly processes of communities (Dubuis et al., 2011). Although reviews prefer individual models (such as MaxEnt), a positive trend towards research and evaluations about stacked machine learning models has arisen (Qazi et al., 2022).

The *Sechium* species of the Central American clade are distributed in the mountains of Panama and Costa Rica (Fig. 1); however, the limited number of samples collected restricts the morphologic, agronomic, and ecological study into the species. The evaluation of SDMs requires a certain number of matches (Schmitt et al., 2017). Given the rarity of some Central American *Sechium* species, the hypothesis is that all the species are distributed under the same environmental conditions and, consequently, independent studies about their distribution are not required. However, the stacked SDMs provide a wider perspective about each species. In addition, their results are more consistent regarding the *Sechium* species communities. Therefore, the objective of this research was to use stacked machine learning algorithms (Random Forest (RF), Support Vector Machines (SVM), and Classification Tree Analysis (CTA) and multivariate tools (clustering and main

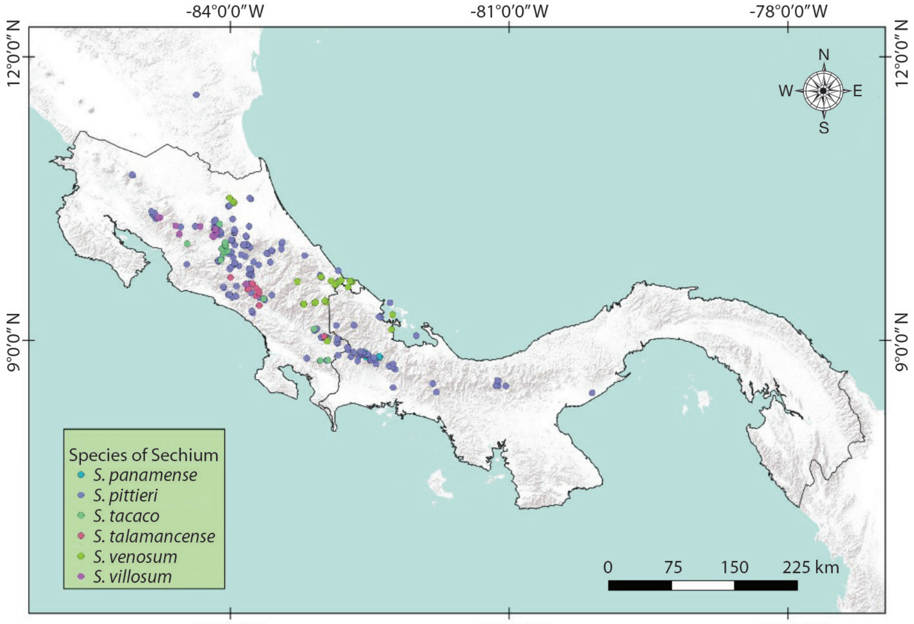

Fig. 1. Geographical distribution of the Central American *Sechium* species in Costa Rica and Panama.

components) to determine the geographical distribution, diversity degree, and endemism of the Central American *Sechium* species in Costa Rica and Panama. This research also aimed to expand databases, identify key conservation areas, and promote the genetic improvement of domesticated species.

MATERIALS AND METHODS

Points of occurrence: Two-hundred and nine points of occurrence were retrieved from the Global Biodiversity Information Facility (GBIF, 2024) for the *Sechium* species of the Central American clade: 147 of *S. pittieri*, 16 of *S. tacaco*, 15 of *S. venosum*, 13 of *S. talamancense*, 9 of *S. panamense*, and 9 of *S. villosum*. Nineteen WorldClim bioclimatic variables (Fick & Hijmans, 2017) were used in raster format, with a 30 arcsecond (~1 km^2) resolution. The raster formats of the Köppen climate classification (Beck, 2012), the Harmonized World Soil Database v 1.2 (Fischer et al., 2002), and the elevation data (~1 km^2 resolution) (Fick & Hijmans, 2017) were added as well. The raster values were obtained with the QGIS Point Sampling Tool version 3.20.1 (QGIS Development Team, 2020) and were exported to a spreadsheet.

Statistical and multivariate analysis: The raster values of the 20 quantitative (19 WorldClim variables and altitude) and qualitative (weather and soil) variables were analyzed with the descriptive statistics of the RStudio software (R Core Team, 2020). The aim was to obtain temperature and humidity referential values, using the package *ggstatsplot* (Patil, 2021) and package *psych* (Revelle, 2020). Given that the quantitative variables did not comply with the assumptions of the analysis of variance, the Kruskal-Wallis test was used to determine the differences between species. If significant differences were detected, Dunn’s post hoc test was applied. In addition, the Support Vector Machine classification method was applied, using the Kernel function, to classify any habitat types found in the area (Meyer et al., 2019). The package *clustertend* (Yilan & Rutong, 2015) was used to calculate the Hopkins (H) statistics, to verify the clustering trend of the raster values of the 20 quantitative variables. The package *clValid* was used to determine the selection of the (hierarchical and non-hierarchical) clustering algorithm (Brock et al., 2008). This procedure was used to determine a non-hierarchical k-medoids clustering, to find similar edaphoclimatic conditions, using package *factoextra* (Kassambara, 2017) and package FactoMineR (Lê et al., 2008). The *NbClust* package (Charrad et al., 2014) enables a consensus of up to 30 clustering indexes and was used to calculate the number of optimal clusters. All the abovementioned packages were executed with the RStudio software (R Core Team, 2020).

Algorithms used in the stacked modeling: To develop the distribution models, a Pearson correlation analysis was carried out to eliminate the $R < -0.90$ and $R > 0.90$ variables, to prevent collinearity, and to prevent compromising the effectiveness of the models (Phillips et al., 2010). Meanwhile, qualitative variables, such as climate and soil, were added. The Random Forest (RF) algorithms (Breiman, 2001) were used to model the presence or absence of species in areas where sampling was not carried out, based on the occurrence records. In addition, the classification tree analysis algorithm (CTA) was used to identify priority areas, given its capacity to predict the presence of endangered species or species with limited distribution (Schmitt et al., 2017). Finally, the support vector machine (SVM) algorithm was chosen because of its high accuracy and capacity to manage non-linear data (Vapnik, 1998). When outliners are in place, they have a high prediction level and provide strong results. The three algorithms were executed and assembled with the stacked species distribution models (SSDM) (Schmitt et al., 2017), which can generate random data about pseudo-absences in areas where species are absent. In addition, they can calculate the diversity of the species and the weighted endemism index (WEI). Twenty-five percent of the data were used in the test, while the remaining 75 % were used for training objectives (Hijmans & Elith, 2013). The aim was to eliminate the spatial classification bias studied by Lobo et al. (2007).

To evaluate the effectiveness of the models, some indicators such as the rate of change indicator (ROC) and the total severity score (TSS) were used. ROC values higher than 0.75 indicate a good performance, while TSS values higher than 0.75 indicates excellent performance. Meanwhile, 0.4 < TSS < 0.75 indicates an optimum performance and TSS < 0.4 indicates a low performance. Cohen's Kappa coefficient accurately quantifies the predicted points after the elimination of the random match probability (RMP). In this study, its value fluctuated between -1 and 1: values closer to 1 indicate an excellent performance of the model, while values closer to -1 show a poor performance. The TSS was calculated (specificity + sensitivity) -1) to adjust for the dependence of Cohen's Kappa on prevalence. TSS optimizes and corrects the problems of Kappa, considering the performance criteria. In addition, the TSS and Kappa share the same performance criteria (Allouche et al., 2006).

Meanwhile, the Jackknife test was used to observe the contribution of each environmental variable to the distribution models, along with ROC, Cohen's Kappa, TSS, and the percentage of correct predictions (PCP). The assembly of the three algorithms was calculated with the package SSDM (Schmitt et al., 2017). The raster format resulting from the assembly of the three models (diversity and endemism maps) were exported with the raster package (Hijmans, 2020), to appropriately edit and visualize them in the QGIS Point Sampling Tool version 3.16.2 (QGIS Development Team, 2020).

RESULTS

Climate and soil diversity: The predominant climate was tropical wet (Af). This type of climate can be found in 29.2 % of the 209 points of occurrence, followed by Cfb (27.3 %), Am (20.6 %), Cwb (18.7 %), Aw (3.83 %), and Cwc (0.47 %). The last climate type was only reported for a *S. pittieri* ascension. Most Central American *Sechium* species can be found in tropical (A) and temperate (C) climates, with their respective variations (Fig. 2). *S. talamancense, S. tacaco*, and *S. pittieri* were found in most occurrences' records and, consequently, covered different climates. The *S. villosum* and *S. panamense* ascensions can only be found in andosols. The *S. pittieri* ascensions can be found in cambisols (38.1 %), andosols (25.9 %), arenosols (18.4 %), nitisols (7.48 %), kastanozems (5.44 %), regosols (4.08 %), and leptosols (0.68 %) (Fig. 3).

***S. panamense*:** This species is native to Chiriquí, Panama, where it grows in altitudes that fluctuate between 1 500 and 3 000 m.a.s.l. (Lira, 1995). Its botanical description is based merely on six collected samples. Consequently, registering more individuals and populations is fundamental to find variations and possible hybridizations that could drive their study and exploitation. As a result of the altitude in which it thrives, this species could be used in genetic improvement programs. Regarding the points of occurrence of this study, *S. panamense* grows at a mean annual temperature of 19 °C; however, it can thrive in lower (10 °C minimum) and higher (31 °C maximum) temperatures and tolerates precipitations of 2 600-3 000 mm. It is typically found in Am, Af, and Cfb climate types (Fig. 2). *S. panamense* can perfectly adapt to volcanic and permeable andosols, with a light layer of organic matter (Fig. 3). The altitude interval of this species is 1 300-2 000 m.a.s.l.; however, Lira (1995) reported that it can expand its limits to 3 000 m.a.s.l.

***S. pittieri*:** This species is rarely found in Nicaragua; however, it prevails in Costa Rica, where it is consumed as a vegetable. As a result of its phenotypic and genetic plasticity, it can adapt to different environments, with a wide range of altitudes (up to 2 500 m.a.s.l.). This phenomenon could indicate that *S. pittieri* has a wide genetic diversity and could be used to improve domesticated species, just like *S. talamancense* and *S. panamense* (Lira, 1995). According to this research, this species grows

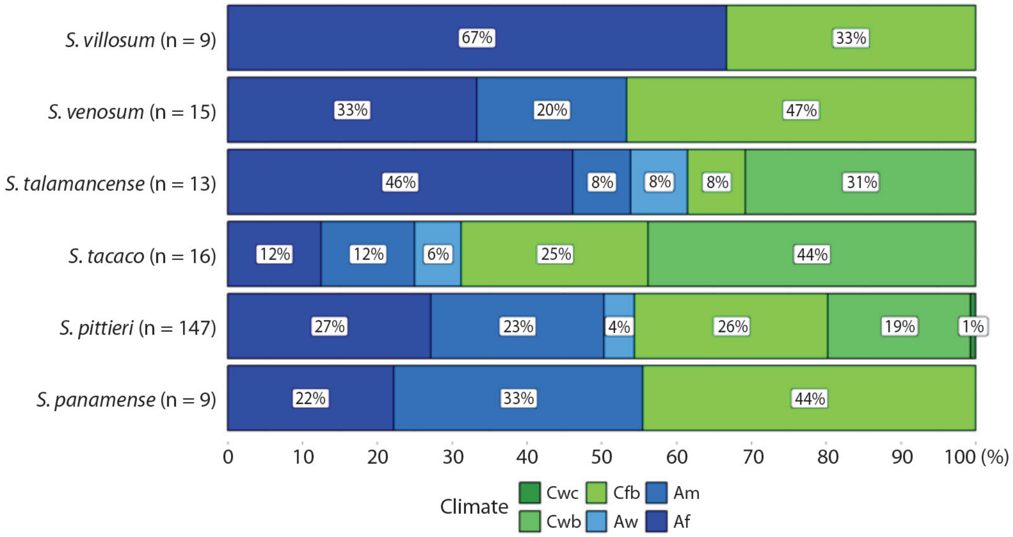

Fig. 2. Climate type percentages of the points of occurrence of *Sechium* species. Cwc = Temperate, dry winter, and cold summer; Cwb = Temperate, dry winter, and warm summer; Cfb = Temperate, without dry season, and warm summer; Aw = Tropical savannah; Am = Tropical monsoon; Af = Tropical rainforest. Climate Group A: tropical, no month with average temperatures below 18 °C, precipitation is still higher than evaporation. Climate Group C: the average temperature of the coldest month is between -3 °C and 18 °C and that of the warmest month exceeds 10 °C, these climates are in temperate forests.

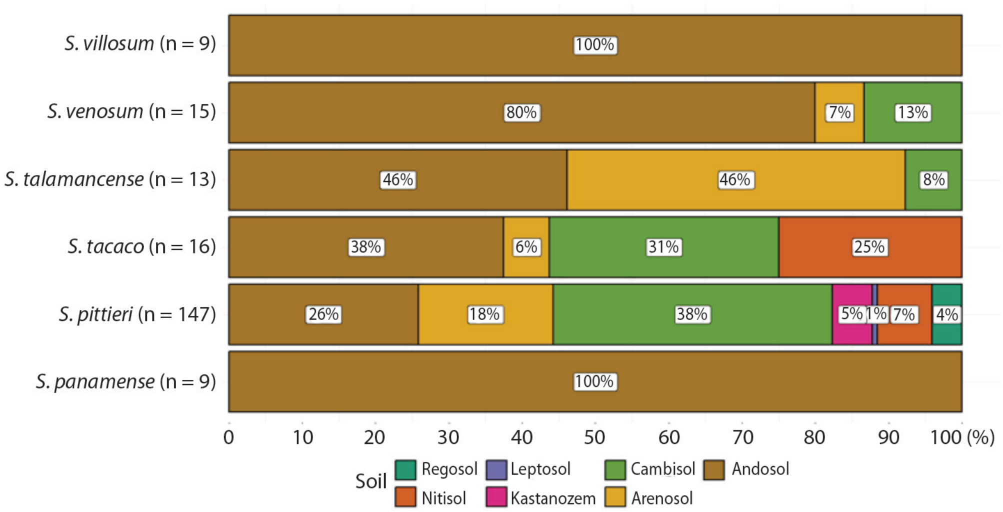

Fig. 3. Soil type ratio in the points of occurrence of *Sechium* species.

and develops at a mean annual temperature of 19 °C; however, it can thrive at lower (2 °C minimum) and higher (32 °C maximum) temperatures. In addition, it can adapt to areas with average precipitations of 2 600-3 200 mm, although one specimen was found in an area with 4 000 mm annual precipitation. Its climatic adaptability range allows its establishment in the six climate types (Af, Am, Aw, Cfb, Cwb, and Cwc) (Fig. 2) and seven soil types (Fig. 3) registered in this study. Its altitude interval is 900-2 900 m.a.s.l., which matches the findings of Lira (1995). However, *S. pittieri* was also found at an altitude of 16 m.a.s.l. The adaptability of this species suggests a wide genetic variability that has not been explored and that should be studied with morphologic and molecular markers.

***S. tacaco*:** This species can be found in the Cordillera de Talamanca, Costa Rica, in altitudes of 1 000-1 700 m.a.s.l. It is locally known as *tacaco* (Wunderlin, 1976). Domesticated and semi-domesticated populations of this species can be found in San José, Costa Rica, where its fruits are used in regional dishes (Lira, 1995). Morales (1994) studied the plant and reproductive morphology of *S. tacaco* and found similarities with the anatomical structure of the organs, multicellular trichomes, floral nectary structures, and anomocytic stomata in the plant and reproductive organs of *S. edule*. Lira (1995) emphasized the low morphological diversity of the *S. tacaco* fruits, except for the presence/absence of thorns and the amount of fiber. However, Monge and Loría (2017) described the morphology of five *S. tacaco* populations from distant localities with a similar average altitude (1 100 m.a.s.l.) and observed diversity and significant differences regarding the weight, length, and width of its fruit. The most important findings of Monge and Loría (2017) were fruits with 6-7 complete longitudinal sutures and 2-5 incomplete longitudinal sutures; this information might help to clarify the evolution processes of the species and the genus *Sechium.* According to the ecological data of this study, the mean annual temperature requirement of this species is 18 ºC; however, it can tolerate lower (9 ºC minimum) and higher (30 ºC maximum) temperatures. It grows in areas with 2 200-2 800 mm precipitations, and its altitude interval is 1 400-2 000 m.a.s.l. The occurrence records match tropical (Aw, Af, and Am) and warm (Cwb and Cfb) areas (Fig. 2). Although it can adapt to a wide range of soils, it thrives under crop conditions in andosols and nitisols, because of its high structural stability, depth, and low base saturation (Fig. 3).

***S. talamancense*:** Also known as *chayotillo* and *tacaquillo*, this species (just like *S. tacaco*) is endemic to the Cordillera de Talamanca, Costa Rica. It can be found in the cloud forest, at an altitude of 2 400-3 200 m.a.s.l. Because of the low temperatures of the area where it thrives, *S. talamancense* could be used for gene transfer and to improve the resistance to frost of the domesticated species (Lira, 1995). Few specimens have been kept in herbaria. The species requires a mean annual temperature of 20 ºC; nevertheless, it can tolerate lower (9 ºC minimum) and higher (30 ºC maximum) temperatures. This species grows under humid conditions, with a precipitation of 3 300-3 800 mm and at an altitude interval of 500-2 200 m.a.s.l. However, it has been found at an altitude of up to 3 200 m.a.s.l. (Lira, 1995). *S. talamancense* can be found in Af, Am, Aw, Cfb, and Cwb climates. As a result of its water requirements, it prefers arenosols, but it can also develop in cambisols and andosols (Fig. 2, Fig. 3).

***S. venosum*:** This species is endemic to the Caribbean Coast of Costa Rica and its pendular inflorescence is very similar to *S. hintonii*. It adapts to high-humidity conditions (Lira, 1995), suggesting that its genetic structure is resistant to phytopathogens, which represent a significant issue for domesticated *Sechium* species (Olguín-Hernández et al., 2013). Consequently, it is one of the main foci of the genetic improvement of the agriculturally important crops (Newstrom, 1990). Like the rest of the Central American species, there are few *S. venosum* specimens in herbaria. It requires a mean annual temperature of 17 ºC but can tolerate lower (6 ºC minimum) and higher (30 ºC maximum) temperatures. Like *S. talamancense*, it can thrive in high-humidity environments, with annual precipitations of 2 600-3 700 mm. Its altitude interval fluctuates between 1 100 and 2 500 m.a.s.l. This species can adapt to Cfb, Am, and Af climates and prefers andosols, arenosols, and cambisols (Fig. 2, Fig. 3).

***S. villosum*:** This species in endemic to Costa Rica and thrives in disturbed tropical or cloud forest environments, at an altitude of 1 500-2 000 m.a.s.l. Like *S. venosum*, it develops in humid environments, which could be a source of gene resistance against fungal diseases (Lira, 1995). *S. villosum* requires a temperature of 19 ºC; however, it tolerates lower (6 ºC minimum) and higher (29 ºC maximum) temperatures. Like all the other species of the clade, it develops in humid environments, with precipitations of 2 900-3 500 mm, and its altitude interval fluctuates between 800 and 900 m.a.s.l.; however, it can be found above 2 700 m.a.s.l. Lira (1995) reported altitude intervals of 1 500-2 000 m.a.s.l. for this species. The occurrence records match the Cfb and Af climates, with a predominance of andosols (Fig. 2, Fig. 3).

Multivariate analysis: Two clusters (k = 2) were created using the non-hierarchical k-medoids and the 209 points of occurrence. The NbClust package was used to calculate the optimum number. The H coefficient was 0.04, which is lower than the threshold (0.5); consequently, the data can be subjected to a clustering analysis, despite their low trend. The two first dimensions accounted for 59.6 % of the total variation (Fig. 4A). Dimension 1 provided 35.1 % of the total variation and was integrated by the elevation variables bio1, bio15, bio4, and bio2 (Fig. 4C).

Stacked modeling: The RF, CTA, and SVM algorithms recorded good performances: AUC > 0.85, Kappa > 0.75, and TSS > 0.75 (Fig. 5A). The Jackknife test showed that the soil variable made the greatest contribution to the *Sechium* species and is responsible for its distribution (Fig. 5B). The three algorithms showed a significant Pearson correlation (average: 0.85). This result indicated that the individual models had similar percentages (86 %) of correct predictions.

The assembly of the RC, CTA, and SVM algorithms —used to quantify the diversity of the species— indicated that the greatest diversity of *Sechium* can mainly be found in the

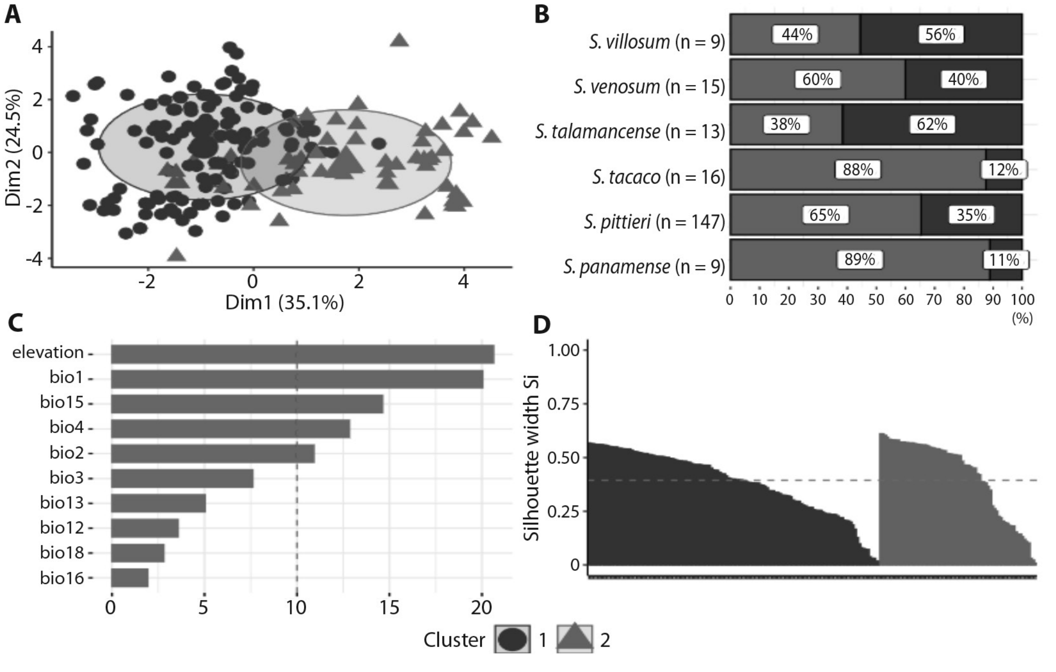

Fig. 4. Multivariate analysis for *Sechium* species: **A.** K-medoids cluster plot. **B.** Percentages of the points of occurrences of *Sechium* species in the clusters. **C.** Contribution of the variables to the components 1 and 2. **D.** Cluster silhouette plot (average silhouette width: 0.39).

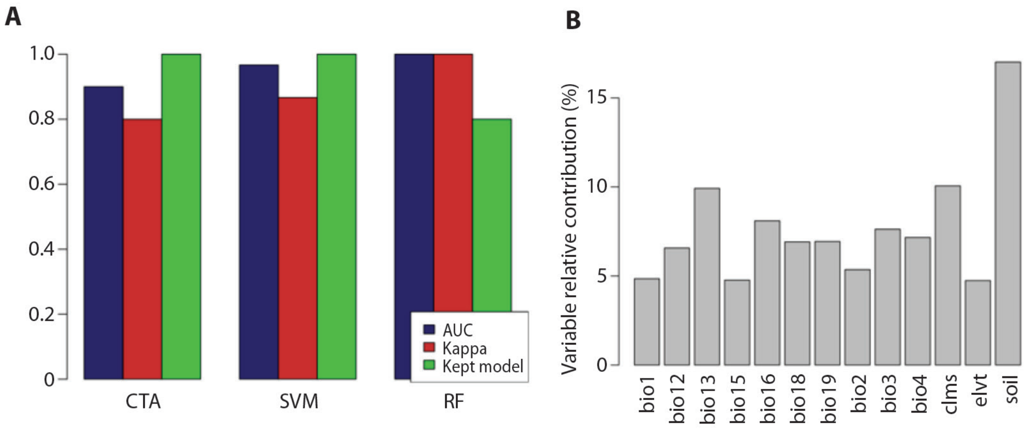

Fig. 5. A. Algorithms evaluated with AUC (> 0.75) and subsequently weighed with the previous metric means and **B.** Jackknife test for variable relative contribution to the algorithms. bio1 = mean annual temperature; bio2 = mean diurnal range (monthly mean (max temperature - min temperature)); bio3 = isothermality (bio2/bio7) (×100); bio4 = temperature seasonality (standard deviation ×100); bio12 = annual precipitation; bio13 = wettest month precipitation rate; bio15 = precipitation seasonality (coefficient of variation); bio16 = wettest quarter precipitation rate; bio18 = warmest quarter precipitation rate; bio19 = coldest quarter precipitation rate; clms = climates; elvt= elevation.

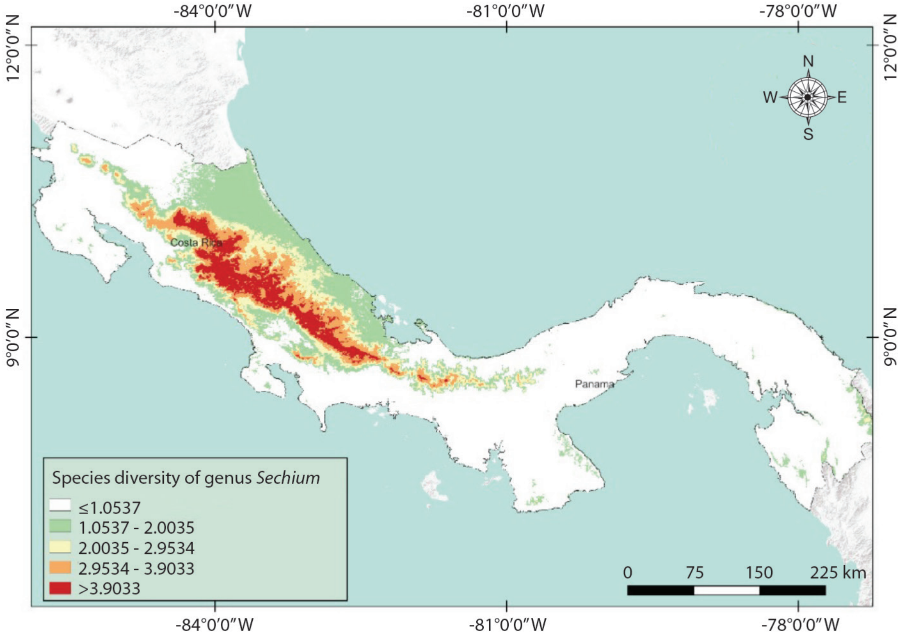

Fig. 6. Map of *Sechium* species diversity of Central America in Costa Rica and Panama.

central mountain region of Costa Rica (Limón, Cartago, Puntarenas, San José Heredia, and Alajuela) and Panama (Bocas del Río, Chiriquí, Bocas de Toro). This phenomenon could be associated with factors including altitude, temperature, and humidity (Fig. 6). Although the red areas on the maps show the greatest diversity, the number of species is closer to four,

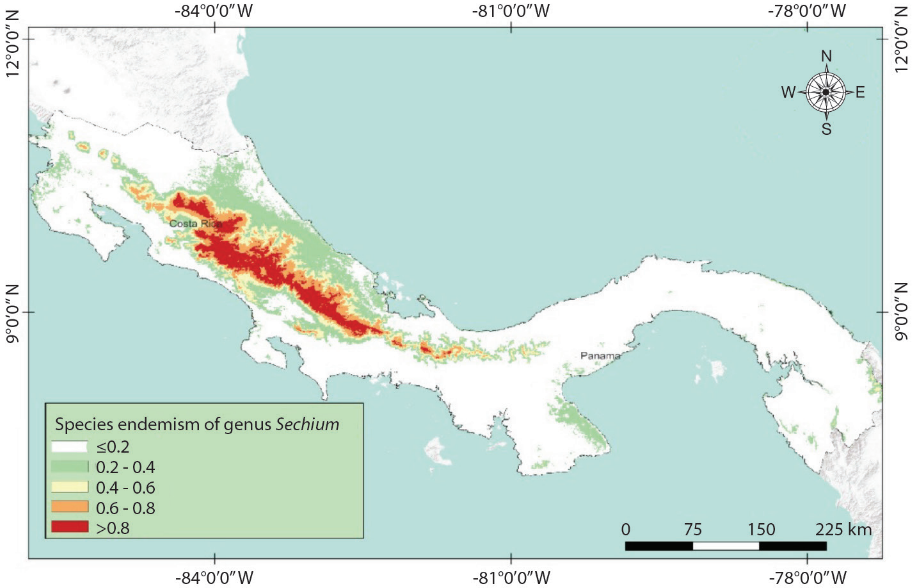

Fig. 7. Map of the endemism of Central American *Sechium* species in Costa Rica and Panama. The caption shows endemism percentages.

which indicates low diversity levels. The areas with less diversity are the coasts and most of the Panamanian territory.

Fig. 7 shows the endemism percentages, which are very similar to the diversity results recorded in Fig. 6. Endemism is measured as a proportion of the total species within a given area. It shows the percentage of the local species that can be endangered if their habitats are destroyed or threatened (Shipley & McGuire, 2022). The greatest concentration of endemism in Costa Rica can be found in the central and mountain area, while in Panama they are more scattered and less concentrated. Therefore, the Cordillera de Talamanca is a notable region due to its high concentration of endemism.

DISCUSSION

The predominant climates of the occurrence records were type A and C, this is because Sechium species have a wide phenotypic plasticity and can adapt to different environments; on the other hand, mountainous areas favor a rich variety of climates where species can diversify in morphological, genetic and physiological aspects. Andosol is the predominant soil for the Central American *Sechium* species. This type of soil has a high organic matter content, a volcanic origin, and a high-water retention and cation exchange capacity. The SVM classification analysis and the Kruskal-Wallis test failed to divide the species into groups or significantly different means. In particular, the Kruskal-Wallis test was only able to differentiate the *Sechium* species into precipitation related variables (bio13: wettest precipitation month; bio15: precipitation seasonality; and bio6: wettest quarter precipitation). Most occurrence records showed low consistency levels, suggesting that all of them can cluster in a single group; however, the NbClust package can only calculate the parameters based on a minimum of two clusters. There are no well-defined groups of the species. Their distribution is limited to Costa Rica and Panama and, consequently, their edaphoclimatic conditions are very similar and they probably interact in the same ecological niches.

Most of the occurrence records of the species are distributed in different clusters —*i.e.*, one cluster can include several points of occurrence of distinct species. They occupy similar ecological niches, due to their limited distribution in Costa Rica and Panama. For instance, 65 % and 35 % of the 147 occurrence records of *S. pittieri* were included in cluster 1 and cluster 2, respectively (Fig. 4B). The silhouette coefficients of the two clusters were 0.42 and 0.38 (Fig. 4D), respectively, which is lower than the threshold (0.5), suggesting that all the occurrence records could be grouped in a single cluster. The high altitude and precipitation fluctuations resulted in a multivariate sensitivity analysis. In addition, soil and climate types play a major role in the distribution of the species (Fig. 5B). The high diversity levels in the mountains could be influenced by the variety of microclimates and the available habitats. This situation is different in low diversity areas, where factors such as urbanization or a greater agricultural activity could reduce the possibilities of finding *Sechium* species.

On the one hand, the mountains of Costa Rica, such as the Cordillera de Talamanca, Cordillera de Guanacaste, Cordillera de Tilarán, and the Central Volcanic Range, usually have microclimates and specific temperature, light, soil diversity, and humidity conditions that favor the development of *Sechium* and other species. On the other hand, the strong geographical isolation is also a major endemism factor in this area (Noroozi et al., 2018; Peñas et al., 2005). Mountain ranges are also sensitive to climate change (La Sorte & Jetz, 2010), which impacts several major bioclimatic variables, such as mean daily temperature, seasonality, and precipitation (Chang et al., 2024). This situation can trigger phenological and aptitude changes that damage the diversity of the species (Munson & Sher, 2015). In addition, factors such as agricultural expansion, new crops, deculturization, urban developments, recreation, and tourism impact endemism areas (Wani et al., 2023).

Wild populations of *S. edule* can be found in Mexico, particularly in Veracruz. They are distributed in the cloud forest, where the specific humidity and shade conditions allow their development. The fruits of these species have different morphological characteristics (Villanueva-Jiménez, 2012). Likewise, the cloud forest of the Costa Rican Mountain ranges is the characteristic habitat of *Sechium* and other species. Their association with other plants enables most of these species to climb tall trees. Regarding endemism and diversity, in some cases, there are few points of occurrence; this is a key indicator that limits the perception of the actual distribution of the species under study.

The multivariate analysis (k-medoids and principal components) showed a slight clustering trend of the species, suggesting that Central American *Sechium* is mainly distributed under the same bioclimatic conditions. However, this was not the case regarding the soil type variable. The analysis of variance did not show significant differences for most of the studied variables, unlike the precipitation variables.

The references included in this study suggest that the high humidity and low temperature conditions required by these species, the Central American *Sechium* species should be included in genetic improvement programs. They can be used to improve not only the domesticated species, but also to obtain cucurbitacin and pharmaceuticals. Within Mexico, successful interspecific hybridization (*S. edule* × *S. compositum*) has been carried out. In addition, several edible domesticated species and their wild and bitter ancestor were subjected to intraspecific hybridization, obtaining genotypes of outstanding hybrid vigor regarding the number of secondary metabolites (Aguiñiga-Sánchez et al., 2015; Avendaño-Arrazate et al., 2014; Cadena-Iñiguez et al., 2008). Compatibility tests between species of interest should be carried out for this type of research. In addition, studies about the morphological, molecular, and biochemical characteristics of the populations of this genus should be conducted.

Central American *Sechium* species thrive under similar bioclimatic conditions; however, significant precipitation differences were recorded. These species develop in volcanic soils, high humidity conditions, and highly fluctuating altitude ranges, but mainly in C type climates. These characteristics are ideal for the improvement of *Sechium* domesticated species. They are also interesting from a bioprospecting point of view, mainly in the research about their biochemical components (*e.g.*, cucurbitacins). The Costa Rican Mountain ranges are the main habitat of *Sechium* species.

Ethical statement: The authors declare that they all agree with this publication and made significant contributions; that there is no conflict of interest of any kind; and that we followed all pertinent ethical and legal procedures and requirements. All financial sources are fully and clearly stated in the acknowledgments section. A signed document has been filed in the journal archives.

REFERENCES

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

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*(6), 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Avendaño-Arrazate, C. H., Cadena-Iñiguez, J., Arévalo-Galarza, M. L. C., Cisneros-Solano, V. M., Morales-Flores, F. J., & Ruiz-Posadas, L. M. (2014). Mejoramiento genético participativo en chayote. *AgroProductividad*, *7*, 30–39.

Baldwin, R. A. (2009). Use of maximum entropy modeling in wildlife research. *Entropy*, *11*(4), 854–866 https://doi.org/10.3390/e11040854

Barrera-Guzmán, L. A., Legaria-Solano, J. P., Cadena-Iñiguez, J., & Sahagún-Castellanos, J. (2021). Phylogenetic relationships among Mexican species of the genus *Sechium* (Cucurbitaceae). *Turkish Journal of Botany*, *45*(4), 302–314. https://doi.org/10.3906/bot-2007-18

Bayliss, S. L. J., Mueller, L. O., Ware, I. M., Schweitzer, J. A., & Bailey, J. K. (2022). Stacked distribution models predict climate-driven loss of variation in leaf phenology at continental scales. *Communications Biology*, *5*(1), 1213. https://doi.org/10.1038/s42003-022-04131-z

Beck, J. (2012). Predicting climate change effects on agriculture from ecological niche modeling: Who profits, who loses? *Climatic Change*, *116*(1–2), 177–189.

Bedair, H., Shaltout, K., & Halmy, M. W. A. (2023). Stacked machine learning models for predicting species richness and endemism for Mediterranean endemic plants in the Mareotis subsector in Egypt. *Plant Ecology*, *224*(12), 1113–1126. https://doi.org/10.1007/s11258-023-01366-6

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R package for cluster validation. *Journal of Statistical Software*, *25*(4), 1–22. https://doi.org/10.18637/jss.v025.i04

Cadena-Iñiguez, J., Avendaño-Arrazate, C. H., Soto-Hernández, M., Ruiz-Posadas, L. M., Aguirre-Medina, J. F., & Arévalo-Galarza, L. (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

Chang, A., Wu, T., Li, B., Jiao, D., Wang, Y., He, D., Jiang, Z., & Fan, Z. (2024). Distribution pattern of species richness of endemic genera in mountainous areas of Southwest China and its influencing factors. *Sustainability*, *16*(9), 3750. https://doi.org/10.3390/su16093750

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). *NbClust*: An R Package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(6), 1–36. https://doi.org/10.18637/jss.v061.i06

Cross, H., Lira, S. R., & Motley, T. J. (2006). Origin and diversification of chayote. In T. J. Motley, N. Zerega, & H. Cross (Eds.), *Darwin's harvest: New approaches to the origins, evolution, and conservation of crops* (pp. 171–194). Columbia University Press.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J. P., & Guisan, A. (2011). Predicting spatial patterns of plant species richness: A comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, *17*(6), 1122–1131. https://doi.org/10.1111/j.1472-4642.2011.00792.x

Fick, S. E., & 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., van Velthuizen, H., & Shah, M. (2002). *Global agro-ecological assessment for agriculture in the 21st century: Methodology and results*. International Institute for Applied Systems Analysis (IIASA) & Food and Agriculture Organization of the United Nations (FAO).

Global Biodiversity Information Facility. (2024). *GBIF Occurrence Download*. https://www.gbif.org/es/

Gelfand, A. E. (2022). Spatial modeling for the distribution of species in plant communities. *Spatial Statistics*, *50*, 100582. https://doi.org/10.1016/j.spasta.2021.100582

Hijmans, R. J. (2020). *raster: Geographic data analysis and modeling. R package* (Version 3.3-13) [Software]. https://CRAN.R-project.org/package=raster

Hijmans, R. J., & Elith, J. (2013). *Species distribution modeling with R*. R CRAN Project.

Kassambara, A. (2017). *Multivariate analysis I. Practical guide to cluster analysis in R. Unsupervised Machine Learning* (1st ed.). STHDA.

La Sorte, F. A., & Jetz, W. (2010). Projected range contractions of montane biodiversity under global warming. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1699), 3401–3410. https://doi.org/10.1098/rspb.2010.0612

Lê, S., Josse, J., & Husson, F. (2008). *FactoMineR*: An R package for multivariate analysis. *Journal of Statistical Software*, *25*(1), 1–18. https://doi.org/10.18637/jss.v025.i01

Lira, R., Caballero, J., & Dávila, P. (1997). A contribution to the generic delimitation of *Sechium* (Cucurbitaceae, Sicyinae). *Taxon*, *46*(2), 269–282. https://doi.org/10.2307/1224097

Lira, R., & Nee, M. (1994). A new species of *Sechium* sect. *Frantzia* (Cucurbitaceae, Sicyeae, Sicyinae) from México. *Brittonia*, *51*(2), 204–209. https://doi.org/10.2307/2666628

Lira, S. R. (1995). *Estudios taxonómicos en el género* Sechium *P. Br. Cucurbitaceae* [Doctoral thesis, Universidad Nacional Autónoma de México]. UNAM Repository. https://repositorio.unam.mx/contenidos/82785

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

Mateo, R. G., Felicísimo, A. M., & Muñoz, J. (2011). Species distributions models: A synthetic revision. *Revista Chilena de Historia Natural*, *84*, 217–240. http://dx.doi.org/10.4067/S0716-078X2011000200008

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2019). *E1071: Misc functions of the department of statistics (E1071), probability theory group (formerly: E1071) TU Wien. R package* (Version 1.7) [Software]. https://CRAN.R-project.org/package=e1071

Monge, J. E., & Loría, M. (2017). Caracterización de frutos de cinco genotipos de tacaco [*Sechium tacaco* (Pittier) C. Jeffrey] en Costa Rica. *Tecnología en Marcha*, *30*(3), 71–84. https://doi.org/10.18845/tm.v30i3.3274

Morales, A. J. (1994). Morfología general del tacaco, *Sechium tacaco* (Cucurbitaceae). *Revista de Biología Tropical*, *42*(1-2), 59-71. https://archivo.revistas.ucr.ac.cr/index.php/rbt/article/view/22462

Munson, S. M., & Sher, A. A. (2015). Long-term shifts in the phenology of rare and endemic Rocky Mountain plants. *American Journal of Botany*, *102*(8), 1268–1276. https://doi.org/10.3732/ajb.1500156

Newstrom, L. E. (1990). Origin and evolution of chayote, *Sechium edule*. In C. Jeffrey (Ed.), *Biology and utilization of the Cucurbitaceae* (pp. 141–149). Cornell University Press.

Noroozi, J., Talebi, A., Doostmohammadi, M., Rumpf, S. B., Linder, H. P., & Schneeweiss, G. M. (2018). Hotspots within a global biodiversity hotspot—Areas of endemism are associated with high mountain ranges. *Scientific Reports*, *8*(1), 10345. https://doi.org/10.1038/s41598-018-28504-9

Olguín-Hernández, G., Valdovinos-Ponce, G., Cadena-Iñiguez, J., & Arévalo-Galarza, M. L. C. (2013). Etiología de la marchitez de plantas de chayote (*Sechium edule*) en el Estado de Veracruz. *Revista Mexicana de Fitopatología*, *31*(2), 161–169.

Patil, I. (2021). Visualizations with statistical details: The ‘ggstatsplot’ approach. *The Journal of Open Source Software*, *6*(61), 3167. https://doi.org/10.21105/joss.03167

Peñas, J., Pérez-García, F. J., & Mota, J. F. (2005). Patterns of endemic plants and biogeography of the Baetic high mountains (south Spain). *Acta Botanica Gallica*, *152*(3), 347–360. https://doi.org/10.1080/12538078.2005.10515494

Phillips, S. J. (2010). A brief tutorial on Maxent. *Lessons in Conservation*, *3*, 108–135.

Qazi, A. W., Saqib, Z., & Zaman-ul-Haq, M. (2022). Trends in species distribution modelling in context of rare and endemic plants: A systematic review. *Ecological Processes*, *11*(1), 40. https://doi.org/10.1186/s13717-022-00384-y

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

R Core Team. (2020). *R: A language and environment for statistical computing* (Version 1.3.1093) [Software]. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research. R package*

(Version 2.0.9) [Software]. Northwestern University. https://CRAN.R-project.org/package=psych

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

Sebastian, P., Schaefer, H., Lira, R., Telford, I. R. H., & Renner, S. S. (2012). Radiation following long-distance dispersal: The contributions of time, opportunity and diaspore morphology in Sicyos (Cucurbitaceae). *Journal of Biogeography*, *39*(8), 1427–1438. https://doi.org/10.1111/j.1365-2699.2012.02695.x

Shipley, B. R., & McGuire, J. L. (2022). Interpreting and integrating multiple endemism metrics to identify hotspots for conservation priorities. *Biological Conservation*, *265*, 109403. https://doi.org/10.1016/j.biocon.2021.109403

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

Villanueva-Jiménez, J. A. (2012). Las variedades del chayote (*Sechium edule* (Jacq.) Sw) y su comercio mundial. *Agricultura, Sociedad y Desarrollo*, *9*(4), 481–482.

Wani, Z. A., Akhter, F., Ridwan, Q., Rawat, Y. S., Ahmad, Z., & Pant, S. (2023). A bibliometric analysis of studies on plant endemism during the period of 1991–2022. *Journal of Zoological and Botanical Gardens*, *4*(4). 692–710. https://doi.org/10.3390/jzbg4040049

Wunderlin, R. P. (1976). Two new species and a new combination in *Frantzia* (Cucurbitaceae). *Brittonia*, *28*(2), 239–244. https://doi.org/10.2307/2805833

Yilan, L., & Rutong, Z. (2015). *clustertend: Check the clustering tendency. R package* (Version 1.4.) [Software]. https://CRAN.R-project.org/package=clustertend

Zwiener, V. P., & Alves, V. A. (2023). Community-level predictions in a megadiverse hotspot: Comparison of stacked species distribution models to forest inventory data. *Journal of Plant Ecology*, *16*(3), 099. https://doi.org/10.1093/jpe/rtac099