

# Potential distribution modeling based on machine learning of *Sechium edule* (Jacq.) Sw. in Japan

**Jorge Cadena-Iñiguez**

Colegio de Postgraduados Campus San Luis Potosi

**Luis Ángel Barrera-Guzmán**

Universidad del Valle de Puebla

**Daniel Alejandro Cadena-Zamudio** (✉ [cadena.daniel@inifap.gob.mx](mailto:cadena.daniel@inifap.gob.mx))

INIFAP: Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias <https://orcid.org/0000-0002-6972-7414>

**Gustavo Antonio Arévalo Galarza**

Universidad Autonoma Chapingo

**Juan Francisco Aguirre Medina**

Universidad Autónoma de Chiapas: Universidad Autonoma de Chiapas

---

## Research Article

**Keywords:** chayote, machine learning, bioclimatic variables, AUC, TSS, temperature

**Posted Date:** August 14th, 2023

**DOI:** <https://doi.org/10.21203/rs.3.rs-3243661/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Genetic Resources and Crop Evolution on October 24th, 2023. See the published version at <https://doi.org/10.1007/s10722-023-01739-w>.

# Abstract

Species distribution models identify regions with ideal environmental characteristics for the establishment and proliferation of species. The chayote is a crop that originated and domesticated in Mexico; however, it is cultivated in different parts of the world due to its nutritional and pharmaceutical importance. The objective of this research was to locate the potential distribution of *S. edule* in Japan supported on seven machine learning models, to also determine which bioclimatic variables influence its distribution, and which are the most suitable regions for its establishment. Thirty-one occurrence points, elevation, and the bioclimatic variables bio1, bio3, bio4, bio7, bio8, bio12, bio14, bio15, and bio17 were used to infer the models. Hundred percent of the occurrence points coincided with the Cfa climate distributed in Acrisol (60.9%), Andosol (17.4%), Cambisol (13%), Fluvisol (4.35%), and Gleysol (4.35%) soil. The Maxent model reported the highest AUC value (0.93), while the GLM obtained the best TSS value (0.84); the SVM model reported the largest suitability area  $\geq 0.5$  with 100,394.4 km<sup>2</sup>. Temperature-related variables were the major contributors to the models and the ones explaining the distribution limits of *S. edule* in Japan. The coastal eastern prefectures of Kantō, Chūbu, Kinki, Chūgoku, Kyūshū, and Shikoku regions showed a suitability  $\geq 0.5$ .

## Introduction

*Sechium edule* [(Jacq.) Swartz] is a native crop of the southern mountainous region of Mexico, its fruits known as "chayote", "chocho", "cidrayota", "gayota", "papa del aire", "cayota", to mention some examples, are used as a food source and in pharmaceutical research (Newstrom 1990; Lira 1995; Cadena-Iñiguez et al. 2013; Aguiñiga-Sánchez et al. 2015, 2017; Salazar-Aguilar et al. 2017; Rosado-Pérez et al. 2019). Mexico is the country with the highest *S. edule* production worldwide with a total annual production of 195,190 tons and an average yield of 66 t ha<sup>-1</sup> (SIACON 2020).

Due to its nutraceutical properties, chayote has been introduced in different regions of the world for its cultivation and research. In India, 74 varieties have been studied with morphological and molecular markers (RAPD and ISSR) (Verma et al. 2017); morphological characterization, and DAMD-type molecular markers (Jain et al. 2017), and evaluation of nutritional content (Mishra and Das 2015). Morphological and biochemical studies have been conducted in Spain (Castro-Rodríguez et al. 2015); in Italy, it has been reported that *S. edule* fruits contain 26 compounds classified into phenolic acids, flavonoids, and polar compounds (Díaz-de Cerio et al. 2019), just to cite some examples. *Sechium edule* is gaining importance worldwide, there is already a first draft of its genome by Nanopore third-generation sequencing, which has an approximate size of 606.02 megabases with 28,237 protein-coding genes (Fu et al. 2021).

Chayote crops thrive in the mesophyll mountain forest, evergreen or semi-evergreen forest (Lira et al. 2009b); grows at altitudes of 300–2000 m a.s.l., with 80–85% relative humidity, well-distributed annual rainfalls of 1500–2000 mm, and suitable temperatures of 13–21 °C (Lira 1996). However, these range of climatic conditions may vary depending on the intraspecific variation of *S. edule* (Cadena-Iñiguez and Arévalo-Galarza 2011). The main use of chayote is for alimentary purposes, hence the importance of its introduction worldwide. However, to introduce it in other regions, it is necessary to assess its potential distribution and climatic suitability.

Species distribution models (SDMs) are based on using Geographic Information Systems (GIS) and statistical tools to predict suitable geographical areas in which species can thrive under certain environmental conditions (Mateo et al. 2011). In recent years, SDM's based on machine learning are capable of processing a large amount of information and their predictive capacity facilitates its application into bioclimatic modeling (Bhattacharya 2013). Machine learning is a research area that derives from artificial intelligence for extracting information which presents models that provide a detailed description of the data and allows making predictions about them (Ryo et al. 2021). Although there are many machine learning techniques, the maximum entropy model (Maxent) is one of the most widely employed, due to its efficiency and computation simplicity (Phillips et al. 2006). To improve and enrich the discussion on SDM's, it is convenient to execute different techniques to obtain multiple perspectives on a distribution, in addition, multiple models can be assembled and so collect information from each specific model (Schmitt et al. 2017).

Species distribution models applications in agriculture are quite diverse. In Egypt, protected geographical areas have been delimited to conserve medicinal plants with a Maxent model (Kaky and Gilbert 2016); Williams et al. (2009) applied generalized linear models, artificial neural networks, random forest, and Maxent to establish suitable areas for six restricted distribution species in northern California, USA. Santos-González et al. (2017) made predictions with the Maxent algorithm to forecast the distribution of the specific variation of *S. edule* in different climate change scenarios for the year 2050. Other applications include conservation research for endangered species such as orchids (Wang et al. 2015), the combination of genotyping methods and species distribution models to make phylogenetic inferences (Ball et al. 2020), potential areas for forest restoration (Gastón et al. 2014); monitoring of pollinating agents (Polce et al. 2013) and prediction of invasive species for crops (Lantschner et al. 2018), to mention some examples.

The objective of this research was to determine the potential distribution of *S. edule* in Japan by applying machine learning models, as well as to determine which are the bioclimatic variables that influence its distribution and which are the ideal regions for its development, to strengthen the Nagoya protocol agreements between Japan (University of Tsukuba) and Mexico (Interdisciplinary Research Group on *Sechium edule* in Mexico).

## Materials and methods

### Data on occurrence and edaphoclimatic information

Eighty-one *S. edule* occurrence points present in Japan were obtained from the Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>) database; however, this number was reduced to 31 because some occurrence points had missing or repeated coordinates. Despite few occurrence points, Hijmans and Elith (2013) argue that for model consistency there must be at least 20 occurrence points. The 19 bioclimatic variables of WorldClim version 2.1 from the 1970–2000 period in a  $\sim 1 \text{ km}^2$  spatial resolution were used (Table 1) (Fick and Hijmans 2017). Raster models with  $\sim 1 \text{ km}^2$  spatial resolution were then added for altitude (meters) (Fick and Hijmans 2017); Köppen-Geiger climate classification (Beck et al. 2018), and soil types from Harmonized World Soil Database version 1.2 (Fischer et al. 2008). Raster values for the

22 layers were obtained with the Point Sampling Tool from QGIS version 3.18.1 (QGIS Development Team 2020).

Table 1  
List of variables used in the *S. edule* model development for Japan\*

| Variable                                                                         | Description                              | Unit |
|----------------------------------------------------------------------------------|------------------------------------------|------|
| Bio1*                                                                            | Annual Mean Temperature                  | °C   |
| Bio2                                                                             | Mean Diurnal Range                       | °C   |
| Bio3*                                                                            | Isothermality                            | °C   |
| Bio4*                                                                            | Temperature Seasonality                  | °C   |
| Bio5                                                                             | Maximum Temperature of the Warmest Month | °C   |
| Bio6                                                                             | Minimum Temperature of the Coldest Month | °C   |
| Bio7*                                                                            | Temperature Annual Range                 | °C   |
| Bio8*                                                                            | Mean Temperature of Wettest Quarter      | °C   |
| Bio9                                                                             | Mean Temperature of Driest Quarter       | °C   |
| Bio10                                                                            | Mean Temperature of Warmest Quarter      | °C   |
| Bio11                                                                            | Mean Temperature of Coldest Quarter      | °C   |
| Bio12*                                                                           | Annual Precipitation                     | mm   |
| Bio13                                                                            | Precipitation of Wettest Month           | mm   |
| Bio14*                                                                           | Precipitation of Driest Month            | mm   |
| Bio15*                                                                           | Precipitation Seasonality                | mm   |
| Bio16                                                                            | Precipitation of Wettest Quarter         | mm   |
| Bio17*                                                                           | Precipitation of Driest Quarter          | mm   |
| Bio18                                                                            | Precipitation of Warmest Quarter         | mm   |
| Bio19                                                                            | Precipitation of Coldest Quarter         | mm   |
| Elevation*                                                                       | Digital Elevation Model                  | m    |
| * Variables selected in the distribution models of the <i>S. edule</i> in Japan. |                                          |      |

## Distribution and configuration models

All the listed packages were executed in Rstudio (R Core Team 2020). For the 19 bioclimatic and elevation variables, a Pearson correlation analysis was performed to eliminate those variables with coefficients  $\geq 0.8$  to avoiding affecting the models and their complexity (Phillips et al. 2006). With the SSDM package (Schmitt et al. 2017) in the default configuration, seven machine learning models were executed: Generalized linear

model (GLM), Generalized boosted regressions model (GBM), Classification tree analysis (CTA), Random Forest (RF), Support vector machines (SVM), Maxent, Artificial neural network (ANN) and an assembled model of the seven models mentioned above. In the case of the Maxent model, the ENMeval package was used to optimize it (Muscarella et al. 2014), the hinge and threshold functions were disabled to avoid over-fitting of the response curves. The product, linear and quadratic functions were activated with a 1.25 regularization multiplier. In the same SSDM package (Schmitt et al. 2017), a Jackknife test was performed to quantify each of the variable's contributions to the distribution models. The evaluation and performance of the models were calculated with the area under the curve (AUC) and True Skill Statistics (TSS). area under the curve values > 0.5 denote optimal models, while TSS values > 0.4 indicate good model performance and reliability (Eskildsen et al. 2013). To improve the AUC values, point distance sampling was employed, *i.e.*, the distance of presence and absence records was calculated for the test (30%) and training (70%) data (Hijmans and Elith 2013). Area under the curve and TSS values were averaged. Output layers were edited in QGIS version 3.18.1 (QGIS Development Team 2020). The calculation area (km<sup>2</sup>) of the the ideal suitability of  $\geq 0.5$  was executed with the "raster" package (Hijmans 2020).

### Spatial decisions for five varieties of *Sechium edule*

The purpose of zoning is to separate areas with similar potentialities and limitations to achieve the best development of crops. Specific programs can be formulated to provide the most effective support for each area. The parameters used in the definition focus on the climatic and edaphic requirements of the crops and the management systems in which they are grown.

Each zone has a similar combination of constraints and potentials for land use and serves as a reference point for potentially increasing production or limiting resource degradation. To carry out this procedure, a set of climate, soil, and site data (quantitative, nominal, classification) is required to locate the best areas for *S. edule* cultivation. With these elements a decision tree is formed that helps to build decision scenarios. Based on the data from González-Santos et al. (2016) in this work, 19 bioclimatic variables (WorldCim, v. 1.0) were used, to which a principal component analysis was carried out and the most relevant varietal groups were evaluated. The BIO1 Average annual temperature (°C) and BIO2 Mean daytime range (°C) data were the most relevant for the crop. It was modeled for the genotypes *nigrum xalapensis*, *virens levis*, *nigrum spinosum*, *nigrum minor* and *albus dulcis*, because Japan has a material transfer agreement with these varieties (Fig. 1).

**Figure 1** Representative fruits of five edible varieties of *Sechium edule* (Jacq.) Sw. A: *var. nigrum xalapensis*, B: *var. virens levis*, C: *var. nigrum spinosum*, D: *var. nigrum minor* and E: *var. albus dulcis*

## Results

The 31 occurrence points in Japan of *S. edule* occurred in the Cfa climate (temperate, no dry season, and hot summer). As for soil types, Acrisol (60.9%), Andosol (17.4%), Cambisol (13%), Fluvisol (4.35%) and Gleysol (4.35%) were observed. The distribution patterns were inferred by layers bio1, bio3, bio4, bio7, bio8, bio12, bio14, bio15, bio17, and elevation.

The seven distribution models along with the assembly showed optimal values in terms of their AUC and TSS. The Maxent model obtained the highest AUC value (0.93), while the GLM obtained the best TSS value (0.84) (Table 2). The SVM model obtained the largest land area for suitability  $\geq 0.5$  with 100,394.4 km<sup>2</sup> (Table 2), while GLM obtained 175.5 km<sup>2</sup>. The Maxent models and the assembly reported similar territorial extensions. Regarding the Jackknife test, the greatest contribution was obtained by the bioclimatic variables related to temperature: bio1, bio3, bio4, and bio5 (Table 3). The CTA model assigned the same percentage of contribution for each one of the variables.

Table 2  
Statistical parameters for the evaluation of the models and area with suitability > 0.5

| Model    | AUC  | Sensibility | Specificity | TSS  | Suitability $\geq 0.5$    |
|----------|------|-------------|-------------|------|---------------------------|
| GLM      | 0.91 | 0.95        | 0.89        | 0.84 | 175.5 km <sup>2</sup>     |
| GBM      | 0.78 | 0.78        | 0.78        | 0.56 | 48,259.6 km <sup>2</sup>  |
| CTA      | 0.85 | 0.71        | 0.99        | 0.70 | 65,768.7 km <sup>2</sup>  |
| RF       | 0.78 | 0.78        | 0.78        | 0.56 | 85,950.9 km <sup>2</sup>  |
| SVM      | 0.78 | 0.78        | 0.78        | 0.56 | 100,394.4 km <sup>2</sup> |
| MAXENT   | 0.93 | 0.78        | 0.80        | 0.58 | 17,139.1 km <sup>2</sup>  |
| ANN      | 0.60 | 0.50        | 0.64        | 0.14 | 61,536.4 km <sup>2</sup>  |
| Ensemble | 0.82 | 0.82        | 0.82        | 0.64 | 13,636.5 km <sup>2</sup>  |

Table 3  
Contribution (%) of the bioclimatic variables to the distribution models using the Jackknife test

| Model    | Bio1  | Bio3  | Bio4  | Bio7  | Bio8  | Bio12 | Bio14 | Bio15 | Bio17 | Elevation |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----------|
| GLM      | 8.19  | 27.86 | 28.65 | 27.51 | 0     | 1.41  | 1.89  | 2.04  | 0.48  | 1.95      |
| GBM      | 12.24 | 0.42  | 79.08 | 1.77  | 1.89  | 0.95  | 1.93  | 1.08  | 0.21  | 0.43      |
| CTA      | 10    | 10    | 10    | 10    | 10    | 10    | 10    | 10    | 10    | 10        |
| RF       | 26.54 | 2.58  | 26.27 | 10.7  | 10.71 | 2.72  | 1.91  | 7.7   | 5.53  | 5.34      |
| SVM      | 10.6  | 7.37  | 5.71  | 8.44  | 1.84  | 25.2  | 12.09 | 12.66 | 11.8  | 4.29      |
| MAXENT   | 40.5  | 5.05  | 5.85  | 5.98  | 3.7   | 20.3  | 2.19  | 2.5   | 3.78  | 10.15     |
| ANN      | 19.28 | 9.68  | 11.05 | 18.66 | 9.68  | 0     | 12.49 | 9.59  | 9.57  | 0         |
| Ensemble | 15.9  | 11.31 | 19.47 | 15.02 | 4.93  | 8     | 5.89  | 7.7   | 4.72  | 7.07      |

All seven models and their ensemble exhibited suitability areas in the eastern coastal region of Japan, mainly in the Kantō, Chūbu, Kinki, Shikoku, Chūgoku, and Kyūshū prefectures (Fig. 1). In Fig. 2A the GLM showed 0.25 suitability for the Hokkaido region; however, in the rest of the models, except for ANN (Fig. 2G) there was no suitability for this region. The GBM (Fig. 2B), ACT (Fig. 2C), RF (Fig. 2D), and MVS (Fig. 2E) models showed zones of suitability  $\geq 0.5$  for Ibaraki, Chiba, Saitama, Tokyo, Kanagawa, Yamanashi, Shizuoka, Aichi, Mie, Nara, Wakayama prefectures and the prefectures of Shikoku and Kyūshū. The Maxent models maps (Fig. 2F) and the ensemble (Fig. 2H) exhibited similarities on their suitability areas and their ideal areas  $\geq 0.5$  are lower than the GBM, ACT, RF, ANN, and MVS models. The ANN model (Fig. 2G) was the only one that showed high suitability zones for the western coastal areas of Japan, specifically in Yamagata, Niigata, Toyama, Ishikawa, and Fukui prefectures, the ANN model also showed high suitability in Sapporo.

**Figure 2** Distribution models of *S. edule* in Japan. A: GLM. B: GBM. C: CTA. D: RF. E: SVM. F: Maxent. G: ANN. H: Assembly

At the varietal group level, each one has different environmental preferences. Variables with significant differences were selected, leaving Bio13, Bio3 and Bio4 out of the analysis. Therefore, due to temperature differences in Japan, where the winter is long and low temperatures, as well as very humid summers, therefore the relative weight assigned to minimum temperatures were more relevant than humidity.

The combination of the variables was carried out with a medium risk, giving all the weight of importance to the minimum temperatures and their ranges of variation. Due to the above, a reduction was obtained in the aptitude surfaces with general data. The most adaptive regional trends are southern and southeastern Japan, with persistent warm temperatures in early autumn. The following table shows a hypothetical approximation of the adaptation of the varietal groups, all of them in the southern part of the country.

Table 4 shows the potential area of distribution with adequate characteristics for five varieties of *S. edule*, considering 19 bioclimatic variables in consensus with González-Santos et al. (2016). The variety with the greatest distribution potential is *albus dulcis* in the middle and southern part of Japan, followed by *nigrum minor*, *nigrum spinosum*, *virens levis* and *nigrum xalapensis* with the lowest value (Table 4, Fig. 3).

| Table 4                                                                                                                 |                          |                             |                          |                          |
|-------------------------------------------------------------------------------------------------------------------------|--------------------------|-----------------------------|--------------------------|--------------------------|
| Potential distribution area with suitable agroecological conditions for five varieties of <i>Sechium edule</i> in Japan |                          |                             |                          |                          |
| <i>var. nigrum xalapensis</i>                                                                                           | <i>var. virens levis</i> | <i>var. nigrum spinosum</i> | <i>var. nigrum minor</i> | <i>var. albus dulcis</i> |
| Surface (km <sup>2</sup> )                                                                                              |                          |                             |                          |                          |
| 2,225                                                                                                                   | 13,959                   | 32,681                      | 39,820                   | 51,737                   |

**Figure 3** Distribution models of five edible varieties of chayote (*Sechium edule*) for Japan

## Discussion

The soil type variable was not considered for inferring the models, because it is not a limiting factor in its distribution and the multiple phenotypic variants of chayote allow it to adapt to a wide range of soils; however, like other crops, excessive salinity and humidity can cause phytopathological and physiological problems (Andrade-Luna et al. 2016; Aguirre-Medina et al. 2021; Cadena-Iñiguez et al. 2021). At the same time, climate type is highly correlated with the mean annual temperature (bio1) and mean annual precipitation (bio12); moreover, all occurrence points coincided in Cfa climate, which except for Hokkaidō and Tōhoku regions, is the predominant climate type in Japan. The Hokkaidō and Tōhoku regions are not suitable for *S. edule* distribution due to their below 0°C temperatures. At the occurrence points, an average of 15.68 °C and annual precipitation of 1950 mm were observed, which are within the optimal range for the cultivated chayote (Lira et al. 2009b; Cadena-Iñiguez et al. 2011).

Bioclimatic variables related to temperature are the most important contributors to *S. edule* distribution models for Japan. The phenotypic variability of *S. edule* related to the fruit morphological variables could suggest that *S. edule* varietal complexes require different bioclimatic conditions. In this regard, González-Santos et al. (2017) found no significant differences between yellow (*albus dulcis*) and green chayotes (*nigrum xalapensis* and *virens levis*) in function of the bio1, bio3, bio4, bio5, bio7, bio8, bio9, and bio10 variables. However, significant differences were found between *albus dulcis* and chayotes with ancestral characteristics, *i.e.*, small fruit (*nigrum minor*) and spiny texture (*nigrum spinosum*) for the bio1, bio2, bio6, bio8, bio9, bio10, and bio11 variables.

Although *S. edule* varietal complexes may have different ecological adaptations, the SDMs can be considered to represent the entire species (Mateo et al. 2011). The coniferous and temperate broadleaf forests in Japan might offer the bioclimatic conditions for *S. edule* varietal complexes growth according to the models presented here; however, the establishment of cultivated populations of *S. edule* in Japan requires monitoring and evaluation, in these populations, of their agronomic response. Other factors such as human influence, distribution of pollinating agents, competition with weeds or invasive plants are usually important in SDM's; however, incorporating this data is complicated due to the lack of information and studies of the dynamics they exert on the species, making these variables go unnoticed (Mateo et al. 2011).

An agroclimatic region is a zone with interrelated characteristics between climate and crop systems, and understanding these interrelationships helps to make better decisions in the agronomic management of crops (White et al. 2001). Climatic aptitude studies for cultivated species have recently increased in order to project the best establishment areas that ensure success (Montes-Galban, 2020); for example, Quiroz-Antunez et al. (2022) evaluated climate change scenarios for *Theobroma cacao* L. and *Coffea arabica* L. determining the distribution and fitness surfaces, in addition to comparing the current and future fitness of both crops, in such a way that the impact of climate change was quantified. The results indicated that the future suitability of the land may decrease the potential areas for the establishment of both crops. In this regard, an important feature in the search for agroclimatic aptitude for *Sechium* in Japan is due to the fact that there is a material transfer agreement under the Nagoya Protocol for five edible varieties, for which it is relevant to identify the areas potential for its establishment and reduce the level of risk and loss (Chandra and Idrisova 2011).

# Conclusions

Machine learning-based SDM's exhibited adequate results in terms of their AUC and TSS, with the Maxent, GML, and CTA models standing out. The Kantō, Chūbu, Kinki, Chūgoku, Shikoku, and Kyūshū regions and their coastal eastern prefectures are the most suitable for *S. edule* distribution. Temperature-related bioclimatic variables are the main responsible for the *S. edule* distribution in Japan. The infraspecific variation of chayote (five varieties ) can thrive in coniferous and temperate broadleaf forests; however, its cultivation should be monitored and evaluated in terms of its agronomic response.

# Declarations

## Conflicts of interest

All the authors declare that they have no conflict of interest.

## Human and animal rights

The authors declare that no research involving human participants and/or animals was conducted.

# References

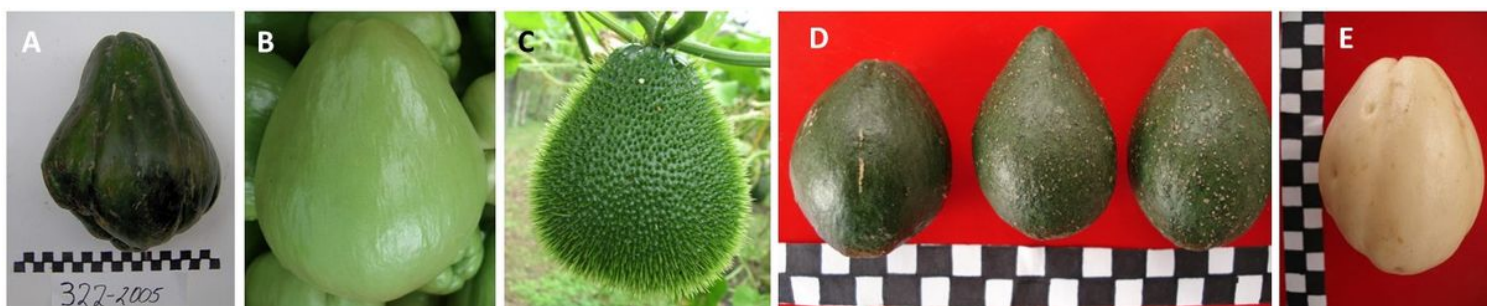
1. Andrade-Luna MI, Espinosa-Victoria D, Gómez-Rodríguez O et al (2016) Severity of a *Phytophthora capsici* strain in chayote *Sechium edule* plants at growth chamber level. Rev Fito Mex 35 (1), 40-57. <https://doi.org/10.18781/R.MEX.FIT.1607-3>
2. Aguiñiga-Sánchez I, Cadena-Iñiguez J, Santiago-Osorio E et al (2017) Chemical analyses and in vitro and in vivo toxicity of fruit methanol extract of *Sechium edule* var. *nigrum spinosum*. Pharm Biol 55, 1638-1645. <https://doi.org/10.1080/13880209.2017.1316746>
3. Aguiñiga-Sánchez I, Soto-Hernández M, Cadena-Iñiguez J et al (2015) Fruit extract from a *Sechium edule* hybrid induce apoptosis in leukaemic cell lines but not in normal cells. Nutr Cancer 67, 250-257. <https://doi.org/10.1080/01635581.2015.989370>
4. Aguirre-Medina JF, Cadena-Iñiguez J, Olguín-Hernández G et al (2021) Co-Inoculation of *Sechium edule* (Jacq.) Sw. Plants with *Rhizophagus intraradices* and *Azospirillum brasilense* to Reduce *Phytophthora capsici*. Damage Agric 11, 391. <https://doi.org/10.3390/agriculture11050391>
5. Ball JW, Robinson TP, Wardell-Johnson GW et al (2020) Fine-scale species distribution modelling and genotyping by sequencing to examine hybridisation between two narrow endemic plant species. Sci. Rep 10, 1562. <https://doi.org/10.1038/s41598-020-58525-2>
6. Beck HE, Zimmermann NE, McVicar TR et al (2018) Present and future Köppen-Geiger climate classification maps at 1-km resolution Sci Data 5, 1-12. <https://doi.org/10.1038/sdata.2018.214>
7. Bhattacharya M 2013 Machine learning for bioclimatic modelling. Int J Adv Comput Sci Appl 4, 1–8.
8. Cadena-Iñiguez J, Arévalo-Galarza MLC (2011) Las variedades de Chayote (*Sechium edule* (Jacq.) Sw.) y su comercio mundial, 1 st ed. ed. bba, Montecillo, Texcoco.

9. Cadena-Iñiguez J, Soto-Hernández M, Torres-Salas A et al (2013) The antiproliferative effect of chayote varieties (*Sechium edule* (Jacq.) Sw.) on tumour cell lines. JMPR 7, 455-460.  
<https://doi.org/10.5897/JMPR12.866>
10. Cadena-Iñiguez J, Olguín-Hernández G, Camacho-Tapia M et al (2021) First Report of *Neoverysiphe sechii* Causing Powdery Mildew on *Sechium edule* in San Luis Potosi, Mexico. Plant Dis 106:1069.  
<https://doi.org/10.1094/PDIS-07-21-1432-PDN>
11. Cadena-Iñiguez J, Soto-Hernández RM, Arévalo-Galarza ML et al (2011) Caracterización bioquímica de variedades domesticadas de chayote *Sechium edule* (Jacq.) Sw. comparadas con parientes silvestres. Rev. Chapingo. Ser. Hortic. 17, 45-55.
12. Chandra A, Idrisova A (2011). Convention on Biological Diversity: a review of national challenges and opportunities for implementation. Bio and Conser, 20, 3295-3316.
13. Castro-Rodríguez JM, Toledo-Díaz AM, Rodríguez-Galdón B et al (2015) Caracterización morfológica y composición química de chayotas (*Sechium edule*) cultivadas en las Islas Canarias (España) Archivos Latinoamericanos de Nutrición 65, 2-17.
14. Díaz-de Cerio E, Fernández-Gutiérrez A, Gómez-Caravaca AM et al (2019) New insight into phenolic composition of chayote (*Sechium edule* (Jacq.) Sw.). Food Chem 295, 514-519.  
<https://doi.org/10.1016/j.foodchem.2019.05.146>
15. Eskildsen A, Roux PC le, Heikkinen RK et al (2013) Testing species distribution models across space and time: high latitude butterflies and recent warming. Glob. Ecol. Biogeogr. 22, 1293-1303.  
<https://doi.org/10.1111/geb.12078>
16. Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. I J Climatol 37, 4302-4315. <https://doi.org/10.1002/joc.5086>
17. Fischer G, Nachtergaele F, Prieler S et al (2008) Global Agro-ecological Zones Assessment for Agriculture.
18. Fu A, Wang Q, Mu J et al (2021) Combined genomic, transcriptomic, and metabolomic analyses provide insights into chayote (*Sechium edule*) evolution and fruit development. Hort. Res. 8, 1-15.  
<https://doi.org/10.1038/s41438-021-00487-1>
19. Gastón A, Garcia-Viñas J, Bravo JA et al (2014) Species distribution models applied to plant species selection in forest restoration: Are model predictions comparable to expert opinion?. New For 1-13.  
<https://doi.org/10.1007/s11056-014-9427-7>
20. González-Santos R, Cadena-Iñiguez J, Morales-Flores FJ et al (2017) Prediction of the effects of climate change on *Sechium edule* (Jacq.) Swartz varietal groups in Mexico. Genet Resour Crop Evol 64, 791-804. <https://doi.org/10.1007/s10722-016-0401-4>
21. Hijmans RJ (2020) raster: Geographic Data Analysis and Modeling.
22. Hijmans RJ Elith J (2013) Species distribution modeling with R.
23. Jain JR, Timsina B, Satyan KB et al (2017) A comparative assessment of morphological and molecular diversity among *Sechium edule* (Jacq.) Sw. accessions in India. 3 Biotech 7, 106.  
<https://doi.org/10.1007/s13205-017-0726-5>

24. Kaky E, Gilbert, F (2016) Using species distribution models to assess the importance of Egypt's protected areas for the conservation of medicinal plants. *J. Arid Environ.* 135, 140-146.  
<https://doi.org/10.1016/j.jaridenv.2016.09.001>
25. Lantschner V, Vega G, Corley J (2018) Predicting the distribution of harmful species and their natural enemies in agricultural, livestock and forestry systems: an overview. *Int. J. Pest Manag.* 65, 1-17.  
<https://doi.org/10.1080/09670874.2018.1533664>
26. 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., Rome, Italy.
27. Lira R, Eguiarte LE, Montes-Hernández S (2009a) 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.
28. Lira R, Téllez O, Dávila P (2009b) The effects of climate change on the geographic distribution of Mexican wild relatives of domesticated Cucurbitaceae. *Genet Resour Crop Evol* 56, 691-703.  
<https://doi.org/doi:10.1007/s10722-008-9394-y>
29. Lira SR (1995) Estudios taxonómicos en el género *Sechium* P. Br. Cucurbitaceae (Tesis Doctoral). Universidad Nacional Autónoma de México, México, D.F.
30. Mateo RG, Felicísimo AM, Muñoz J (2011) Species distributions models: A synthetic revision. *Rev. Chil. de Hist. Nat.* 84, 217-240. <http://dx.doi.org/10.4067/S0716-078X2011000200008>
31. Mishra L, Das P (2015) Nutritional Evaluation of Squash (*Sechium Edule*) Germplasms Collected from Garo Hills of Meghalaya – North East India *Int. j. environ. agric. biotech* 8, 971-975.  
<https://doi.org/10.5958/2230-732X.2015.00111.4>
32. Montes-Galban, E. (2020). Zoning the potential agricultural use in the middle basing of Lujan river, Argentina. *Revista Geográfica Digital* 19(38) 65-80. <https://doi.org/10.30972/geo.19386204>
33. Muscarella R, Galante PJ, Soley-Guardia M et al (2014) ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for ecological niche models. *Methods Ecol. Evol.* 5, 1198-1205. <https://doi.org/10.1111/2041-210X.12261>
34. Newstrom L (1990) Origin and evolution of chayote, *Sechium edule*, in: Bates, D.M., Robinson, R.W., Jeffrey, C. (Eds.), *Biology and Utilization of the Cucurbitaceae*. Cornell University Press, New York, USA, pp. 141-149.
35. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190, 231-259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
36. Polce C, Termansen M, Aguirre-Gutiérrez J et al (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>
37. Quiroz-Antunez UG, Monterroso-Rivas AI, Calderón-Vega MF et al (2022). Aptitud de los cultivos de café (*Coffea arabica* L.) y cacao (*Theobroma cacao* L.) considerando escenarios de cambio climático. *LA GRANJA. Revista de Ciencias de la Vida*, 36(2), 60-74. <https://doi.org/10.17163/lgr.n36.2022.05>
38. QGIS Development Team (2020) QGIS Geographic Information System. Open Source Geospatial Foundation Project.

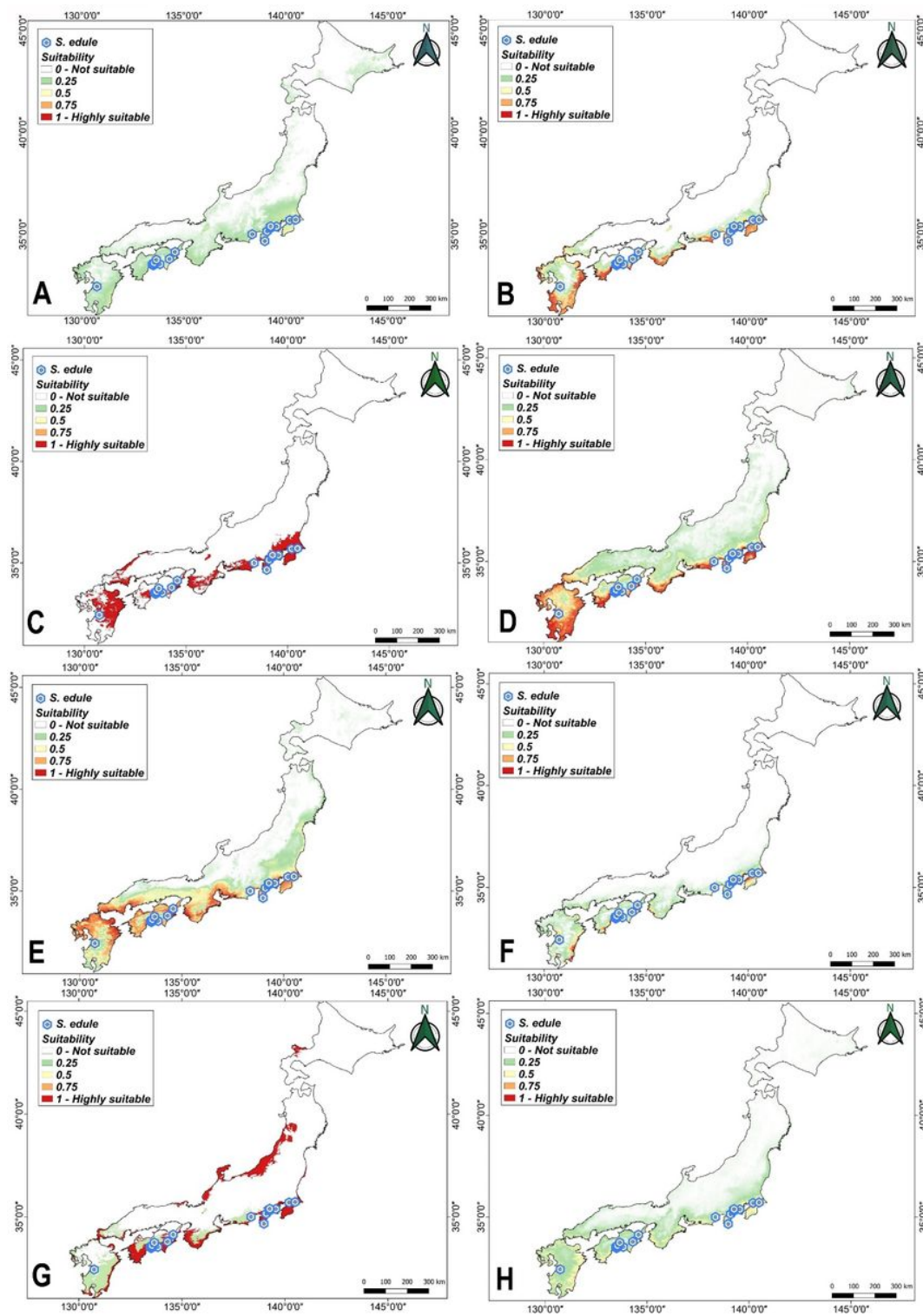
39. R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
40. Rosado-Pérez J, Aguiñiga-Sánchez I, Santiago-Osorio E et al (2019) Effect of *Sechium edule* var. *nigrum spinosum* (Chayote) on Oxidative Stress and Pro-Inflammatory Markers in Older Adults with Metabolic Syndrome: An Exploratory Study Antioxidants (Basel) 8, 146. <https://doi.org/10.3390/antiox8050146>
41. Ryo M, Angelov B, Mammola S et al (2021) Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. Ecography 44, 199-205. <https://doi.org/10.1111/ecog.05360>
42. Salazar-Aguilar S, Ruiz-Posadas LDM, Cadena-Lñíguez J et al (2017) *Sechium edule* (Jacq.) Swartz, a New Cultivar with Antiproliferative Potential in a Human Cervical Cancer HeLa Cell Line. Nutrients 9, E798. <https://doi.org/10.3390/nu9080798>
43. Schmitt S, Pouteau R, Justeau D et al (2017) ssdm: An r package to predict distribution of species richness and composition based on stacked species distribution models. Methods Ecol. Evol. 8, 1795-1803. <https://doi.org/10.1111/2041-210X.12841>
44. SIACON (2020) Sistema de Información Agroalimentaria de Consulta Nueva Generación. SIAP, México.
45. Verma VK, Pandey A, Jha AK et al (2017) Genetic characterization of chayote [*Sechium edule* (Jacq.) Swartz.] landraces of North Eastern Hills of India and conservation measure. Physiol Mol Biol Plants 23, 911-924. <https://doi.org/10.1007/s12298-017-0478-z>
46. Wang HH, Wonkka CL, Treglia ML et al (2015) Species distribution modelling for conservation of an endangered endemic orchid. AoB PLANTS 7. <https://doi.org/10.1093/aobpla/plv039>
47. White DH, Lubulwa GA, Menz K et al (2001). Agro-climatic classification systems for estimating the global distribution of livestock numbers and commodities. Environm internat 27(2-3), 181-187. [https://doi.org/10.1016/S0160-4120\(01\)00080-0](https://doi.org/10.1016/S0160-4120(01)00080-0)
48. Williams J, Seo C, Thorne J et al (2009) Using species distribution models to predict new occurrences for rare plants. Divers. Distrib. 15, 565-576. <https://doi.org/10.1111/j.1472-4642.2009.00567.x>

## Figures



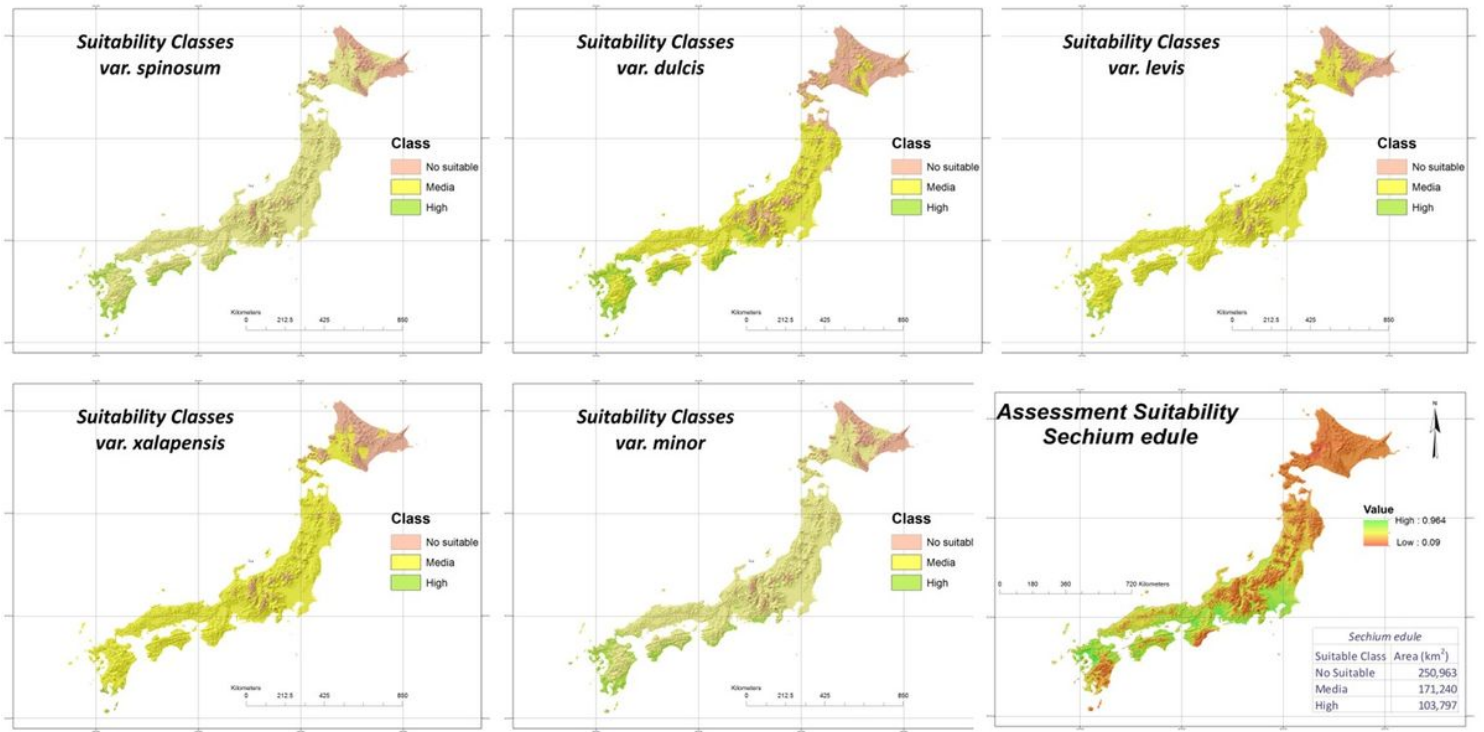
**Figure 1**

Representative fruits of five edible varieties of *Sechium edule* (Jacq.) Sw. A: *var. nigrum xalapensis*, B: *var. virens levis*, C: *var. nigrum spinosum*, D: *var. nigrum minor* and E: *var. albus dulcis*



**Figure 2**

Distribution models of *S. edule* in Japan. A: GLM. B: GBM. C: CTA. D: RF. E: SVM. F: Maxent. G: ANN. H: Assembly



**Figure 3**

Distribution models of five edible varieties of chayote (*Sechium edule*) for Japan