

# Potential distribution models of *Sechium tacaco* (Pittier) C. Jeffrey in Costa Rica

Barrera-Guzmán, Luis A.<sup>1,5</sup>; Cadena-Iñiguez, Jorge<sup>2,5</sup>; Legaria-Solano, Juan P.<sup>1\*</sup>; Ramírez-Ojeda, Gabriela<sup>1,3</sup>; Sahagún-Castellanos, Jaime<sup>1</sup>; Arévalo-Galarza, Ma. de Lourdes<sup>4,5</sup>

<sup>1</sup> Universidad Autónoma Chapingo, Departamento de Fitotecnia, km 38.5 Carretera México-Texcoco, Chapingo, Texcoco, Estado de México.

<sup>2</sup> Colegio de Postgraduados, Campus San Luis Potosí, Iturbide No. 73, Salinas de Hidalgo, San Luis Potosí, México.

<sup>3</sup> Campo Experimental Centro Altos de Jalisco del Instituto Nacional de Investigaciones Agrícolas, Forestales, Agrícolas y Pecuarias (INIFAP), km 8, carretera Tepatitlán-Lagos de Moreno, Jalisco, México.

<sup>4</sup> Colegio de Postgraduados, Campus Montecillo, km. 36.5 carretera México-Texcoco, Montecillo, Texcoco, Estado de México.

<sup>5</sup> Grupo Interdisciplinario de Investigación de *Sechium edule* en México (GISEM), México.

\* Corresponding author: legarias.juan@yahoo.com

## ABSTRACT

**Objective:** Determine current and potential distribution of *S. tacaco* in Costa Rica with seven Species Distribution Models (SDM), in order to optimize the management of *S. tacaco* genetic resources, aimed at identifying patterns of geographic distribution and possible climatic adaptations allowing to have perspectives on their conservation and genetic breeding.

**Design/Methodology/Approach:** 21 points of occurrence together with 19 bioclimatic variables and altitude were used to evaluate seven machine learning models and an assembly of these. Open-source libraries running in Rstudio were used.

**Results:** Distribution models were inferred by the variables bio1, bio2, bio3, bio4, bio12, bio13, bio14, bio18 y bio19. The generalized additive model obtained the highest values of area under the curve (0.96) and True skill statistic (0.90), however, the seven models tested and the assembly showed adequate performance (AUC>0.5 and TSS>0.4). Bioclimatic variables related to temperature were the ones with the greatest contribution to the models and the main limitations in the distribution of *S. tacaco*.

**Study limitations/implications:** Possibly a greater number of occurrence points are required to evaluate distribution models.

**Findings/Conclusions:** Areas with high potential distribution suitability for *S. tacaco* are found in central valleys of Costa Rica, covering regions of the provinces of Alajuela, Cartago, San José and Puntarenas. These areas can be sources of germplasm for future conservation and breeding studies.

**Key words:** SDM, germplasm, conservation, breeding.

**Citation:** Barrera-Guzmán, L. A., Cadena-Iñiguez, J., Legaria-Solano, J. P., Ramírez-Ojeda, G., Sahagún-Castellanos, J., & Arévalo-Galarza, Ma. De L. (2021). Potential distribution models of *Sechium tacaco* (Pittier) C. Jeffrey in Costa Rica. *Agro Productividad*, 14(#). <https://doi.org/10.32854/agrop.v14i7.2006>

**Editor in Chief:** Dr. Jorge Cadena Iñiguez

**Estimated publication date:** August 2021

This work is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International license.



## INTRODUCTION

*Sechium tacaco* (Pittier) C. Jeffrey is an endemic species to the mountainous regions of Costa Rica, where it is locally known as “tacaco” (Wunderlin, 1976). Its possible wild ancestor is also distributed in this country, *Sechium talamancensis* (Wunderlin) C. Jeffrey. Fruits of the tacaco are representative of Costa Rican culture and gastronomy. Through selection, phenotypic variation of the fruits have been achieved based on their weight, equatorial



width, thickness, number of spines and longitudinal sutures. (Monge and Loría, 2017). Generally, tacaco plantations are found in altitudinal ranges of 500-1700 masl. (Wunderlin, 1976; Monge and Loría, 2017) and can reach up to 2000 masl (Lira, 1995). *S. tacaco* is a species underrepresented in germplasm banks and with problems of genetic erosion, due to ignorance of the crop, disturbance of its habitat and introduction of crops (Lira, 1995).

Species distribution models (SDM) are tools that rely on Geographic Information Systems (GIS) and data of real presences to predict areas of suitability for species, this based on their environmental characteristics (Mateo *et al.*, 2011). In general, the methodology used by the SDMs consists of compiling geographic locations of the species of interest; later, spatial data of edaphoclimatic variables are obtained according to the points of occurrence of the species. Spatial data obtained are processed using statistical techniques that can predict suitable areas for the distribution of species. (Hijmans and Elith, 2013). SDMs can be classified into descriptive methods, such as Bioclim, Domain, Mahalanobis distance and Anuclim, which only need presence data to be modeled (Mateo *et al.*, 2011). There are also discriminant techniques such as generalized linear models with their respective variants, for example, generalized additive model and generalized enhanced model; within this group are also automated learning methods such as random trees, maximum entropy and support vector machines, which are used to compute numerical regressions for prediction tasks.

Discriminant techniques based on machine learning are very flexible in terms of computational calculations, they can process a large amount of information and their results are usually more consistent compared to descriptive techniques (Mateo *et al.*, 2011; Hijmans and Elith, 2013; Schmitt *et al.*, 2017). Maximum entropy model (Maxent) is the most applied in species distribution models due to a simple interface in its programming and because it provides adequate results (Phillips *et al.*, 2006). However, it is advisable to optimize the Maxent settings to obtain the best model (Muscarella *et al.*, 2014), and above all to compare it with other algorithms to have multiple perspectives on the species in question.

SDMs can have various applications in agriculture. For example, they are useful for determining the effects of climate change on the distribution of species (Beck, 2012); monitor the presence of pests, invasive species (Lantschner *et al.*, 2018) and pollinating agents (Polce *et al.*, 2013), as well as to detect plant endemisms, threatened habitats, patterns of diversity and conservation studies, among other uses (Mateo *et al.*, 2011). All these applications are vital for the formulation of strategies that seek to maximize plant genetic resources, either to mitigate the damage caused by other species and to prioritize vulnerable areas rich in germplasm (Flores-Tolentino *et al.*, 2019). Regarding genus *Sechium* P. Br., there are few studies on species distribution models, and only *S. edule* y *S. tacaco* are cultivated.

Authors such as González-Santos *et al.* (2017) predicted with Maxent that by year 2050 some varietal complexes of *S. edule* could lose more than half of their current distribution, which is an alarming panorama. In addition, in high-risk areas there are some wild populations that are important to understand the phylogenetic processes of Mexican species of the genus. On the other hand, studies of *S. tacaco* are limited to its morphology (Monge and Loría, 2017) and its phylogeny (Sebastian *et al.*, 2012).

The objective of this study was to determine the current and potential distribution of *S. tacaco* in Costa Rica with seven SDM, in order to guide the management aimed at optimizing geographic distribution patterns and possible climatic adaptations that allow have perspectives on their conservation and genetic breeding.

## MATERIALS AND METHODS

### Occurrence data and environmental information

For this article, 21 points of occurrence (latitude and longitude) of *S. tacaco* (Figure 1) were obtained from the Global Biodiversity Information Facility database (GBIF, <https://www.gbif.org/>) and Monge and Loría (2017). It was verified that occurrence points were not atypical and repeated. The 19 bioclimatic variables of WorldClim version 2.1 from period 1970-2000 with spatial resolution  $\sim 1 \text{ km}^2$  were used (Table 1) (Fick and Hijmans, 2017). Likewise, the altitude raster model with spatial resolution  $\sim 1 \text{ km}^2$  was used to obtain the elevation data in 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) were also used. The raster values of the 22 environmental layers were obtained with the Point Sampling Tool of QGIS version 3.16.2 (QGIS Development Team, 2020).



**Figure 1.** Fruits of *Sechium tacaco*, accession 1038-18 of the Germplasm Bank of *S. edule* (BANGESe). Collector: Arévalo-Galarza, M.L. image by Jorge Cadena Iñiguez.

### Species Distribution Models (SDMs)

All variables and statistical programs were executed in Rstudio (R Core Team, 2020). For raster data of the 20 environmental variables (19 WorldClim variables and altitude), Pearson correlation was calculated and those variables with correlations  $\geq 0.8$  were eliminated, to prevent collinearity and avoid affecting the models (Feng *et al.*, 2019). Climate and soil types variables were used as descriptive information and were omitted in the predictions of the distribution models; in the case of climate type, it is related to some variables such as annual precipitation (bio12) and average annual temperature (bio1); regarding soil types, there is a lack of information regarding its quality; however, by excluding these variables, optimal results can be achieved in SDM (Evans *et al.*, 2010).

**Table 1.** Bioclimatic variables used for SDMs of *S. tacaco* in Costa Rica.

Variable	Description	Unit
Bio1*	Average annual temperature	°C
Bio2*	Median diurnal Temperature range	°C
Bio3*	Isothermality	-
Bio4*	Temperature seasonality	-
Bio5	Maximum temperature of warmest month	°C
Bio6	Minimum temperature of 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	-
Bio16	Precipitation of wettest quarter	mm
Bio17	Precipitation of driest quarter	mm
Bio18*	Precipitation of warmest quarter	mm
Bio19*	Precipitation of coldest quarter	mm
Altitude	Digital elevation model	m

\* Variables selected in SDM of *S. tacaco* in Costa Rica.

Machine learning techniques or SDM widely used according to literature were used due to their high performance and optimal results (Mateo *et al.*, 2011), including the following algorithms: Generalized Linear Model (GLM), Generalized Additive Model (GAM), Generalized Power Regression Model (GBM), Classification Tree Analysis (CTA), Maxent, Random Forest (RF) and Vector Machines of Support (SVM). For the execution of these seven models the package SSDM was used (Schmitt *et al.*, 2017) with its default settings and gam dependencies (Wood, 2017), 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). However, to optimize the Maxent model, the ENMeval package was used (Muscarella *et al.*, 2014) with the following configuration: threshold and hinge functions were deactivated to avoid overfitting the response curves, the quadratic and product functions were also deactivated and the regularization multiplier was 1.25 with a linear function. Evaluation of the models was quantified with area under the curve (AUC), where values >0.5 indicate adequate models; however, to eliminate the spatial classification bias studied by Lobo *et al.* (2007), a point distance sampling was carried out, that is, the difference in distances for presences and absences was calculated in training (75%) and test (25%) data (Hijmans and Elith, 2013). To verify the performance of the models, the kappa and TSS statistics were calculated.

Kappa quantifies the proportion of correctly predicted points after eliminating the probability of random coincidence, its value ranges from  $-1$  to  $1$ ; values close to  $1$  indicate excellent model performance and values close to  $-1$  indicate poor performance. To correct dependence on the prevalence of kappa, the TSS ((Specificity + Sensitivity)  $-1$ ) was calculated, which optimizes and corrects kappa problems. TSS performance criteria are the same as kappa (Allouche *et al.*, 2006).

Additionally, Jackknife test was carried out to observe the contribution of each environmental variable to SDM, which were calculated together with the statistics AUC, kappa, TSS, sensitivity, specificity, proportion of correct predictions (PCP) and an assembly of the seven algorithms with the same SSDM package (Schmitt *et al.*, 2017). All statistics were averaged from the test results and training data. The raster of the seven models and the assembly were exported with the raster package (Hijmans, 2020) to QGIS Development Team versión 3.16.2 (2020).

## RESULTS AND DISCUSSION

Five climate types were found in the environments associated with the points of occurrence of the species: Am (tropical, monsoon, 38.1%), Af (tropical, rainforest, 23.8%), Aw (tropical, savanna, 19%), Cwb (temperate, dry winter, warm summer, 14.3%) and Cfb (temperate, no dry season, warm summer, 4.76%). Soil types found were andosol (38.1%), cambisol (33.3%) and alisol (28.6%). Bioclimatic variables to evaluate SDMs were bio1, bio2, bio3, bio4, bio12, bio13, bio14, bio18 and bio19, which resulted from the correlation analysis. Variables related to temperature showed little variation, for example, bio1 obtained a range of 17.65-21.3 °C, with an average of  $\sim 20$  °C. For the variables related to precipitation, broader ranges were obtained, for example, for bio12 there were records with a range of 2200-4000 mm of precipitation.

GAM model obtained the highest AUC (0.96) and TSS (0.90). In general, AUC for the seven algorithms was  $>0.8$ , thus indicating suitable models. TSS was within optimal performance model range of  $0.4 > TSS < 0.7$  (Allouche *et al.*, 2006). The assembly of the seven models was adequate in terms of AUC, kappa and TSS. PPC got values  $>0.8$  and was considered acceptable (Table 2).

**Table 2.** Statistical parameters for SDMs of *S. tacaco* in Costa Rica.

Model	AUC	PPC	Sensibility	Specificity	kappa	TSS
GLM	0.89	0.88	0.91	0.88	0.22	0.79
GAM	0.96	0.93	0.98	0.92	0.37	0.90
GBM	0.83	0.81	0.83	0.83	0.66	0.66
CTA	0.83	0.83	0.91	0.75	0.66	0.57
RF	0.85	0.83	0.83	0.83	0.66	0.66
Maxent	0.92	0.88	0.83	0.88	0.023	0.71
SVM	0.91	0.92	0.91	0.91	0.83	0.82
Assemble	0.83	0.83	0.83	0.84	0.42	0.67

Generalized Linear Model (GLM), Generalized Additive Model (GAM), Generalized Potentiated Regression Model (GBM), Classification Tree Analysis (CTA), Random Forest (RF), Maximum entropy (Maxent), Support Vector Machines (SVM).

Regarding Jackknife test, the variables with the greatest contribution in the SDMs were bio1, bio2, bio3 and bio14. In the assembly of the seven models, the variable with the greatest contribution was bio1. AAC model assigned the same contribution for all predictor variables (Table 3).

**Table 3.** Contribution (%) of environmental variables to SDMs of *S. tacaco* in Costa Rica.

Model	Bio1	Bio2	Bio3	Bio4	Bio12	Bio13	Bio14	Bio18	Bio19
GLM	0.22	4.01	8.01	0.1	6.54	12.36	4.76	37.99	26.07
GAM	23.92	0.29	40.19	14.8	1.68	4.47	1.24	1.65	11.73
GBM	26.63	30.42	2.11	18.04	1.66	0.64	0.21	1.05	18.85
CTA	11.11	11.11	11.11	11.11	11.11	11.11	11.11	11.11	11.11
RF	53.74	7.85	6.61	7.9	3.25	5.03	6.85	1.99	6.74
Maxent	16.8	8.4	43.7	3.5	1.6	1.5	0.3	16	8.3
SVM	24.52	3.78	2.79	11.83	2.17	6.87	30.68	13.06	4.24
Assemble	27.48	10.01	10.54	10.10	6.07	8.57	4.72	10.41	12.04

Generalized Linear Model (GLM), Generalized Additive Model (GAM), Generalized Potentiated Regression Model (GBM), Classification Tree Analysis (CTA), Random Forest (RF), Maximum entropy (Maxent), Support Vector Machines (SVM).

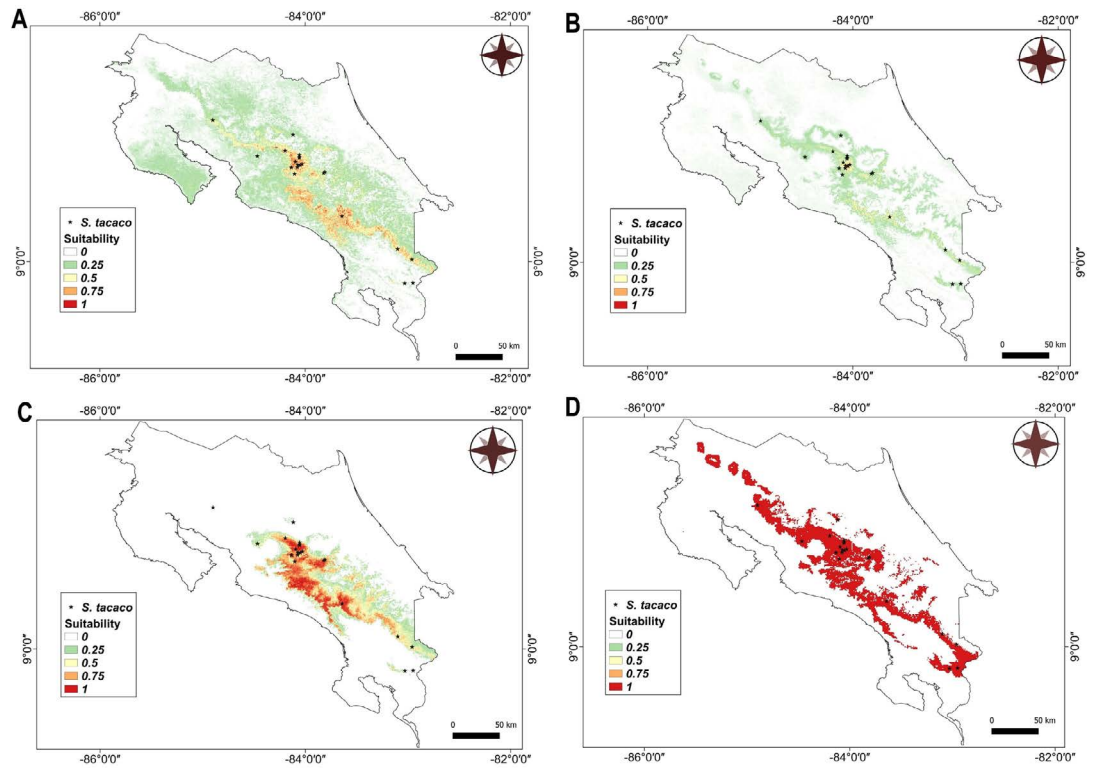
GLM and GAM models show large territorial extensions for a suitability of 0.25 (Figure 2A-B). GBM model indicated suitability with probability of 1 for the Central Valleys of Costa Rica (Figure 2C), peculiarly ACC model showed the maximum suitability in a longitudinal axis of Costa Rican territory that covers part of Alajuela, San José, Cartago, and Puntarenas provinces (Figure 2D).

RF, Maxent and SVM models showed suitability of 1 in very similar regions (Figure 3 A-C), however, RF model had a greater territorial extension in Central Valleys and in the South Pacific where Talamanca Mountain is located. The assembly of the seven models (Figure 3D) presented a potential distribution area very similar to that predicted by Maxent model.

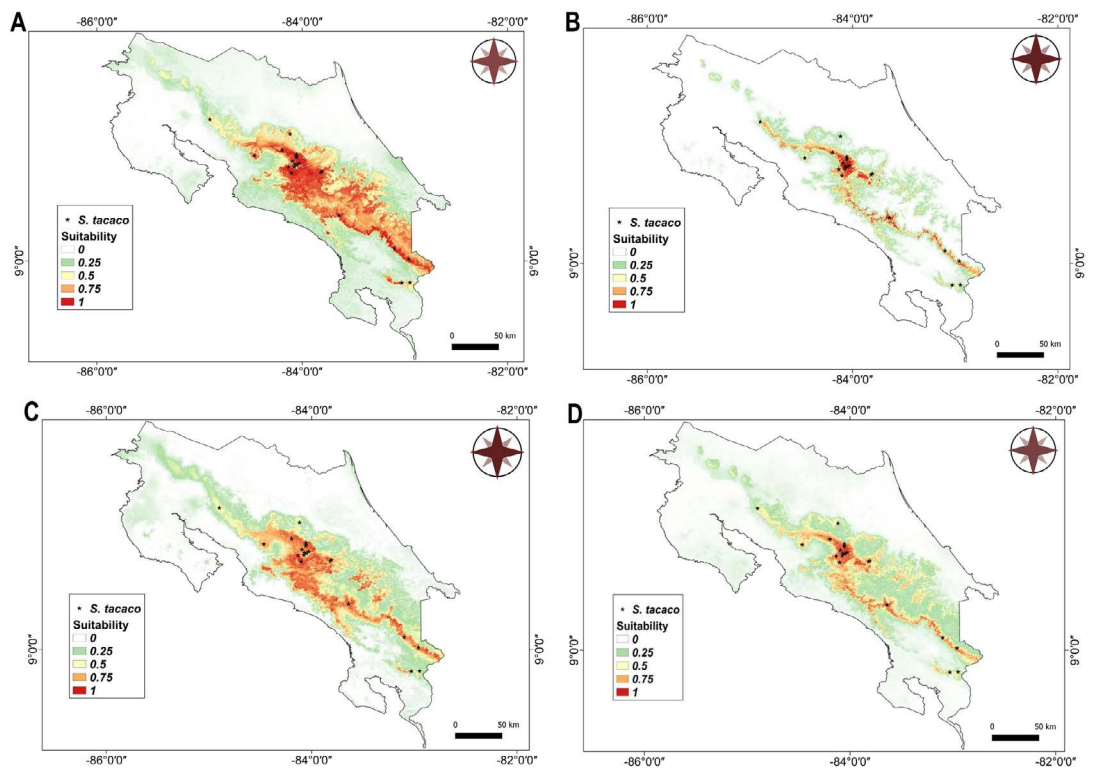
Because 80% of the Costa Rican territory has A climate type, it would be expected that the climate type is not a conditional to determine the distribution of *S. tacaco*, considering also that climate is strongly related to the bioclimatic variables bio1 and bio12.

AUC values for training and testing (0.97 and 0.93) determined a good modeling for potential distribution of *S. tacaco*. Values very close to 1 are usually indicative of restricted distribution of species (Phillips *et al.*, 2006), just as it happens for *S. tacaco*. Lira *et al.* (2018) found AUC values higher than 0.95 in wild populations of *S. edule*, which are only distributed in the states of Oaxaca and Veracruz, Mexico. Generally, endemic species tend to have low levels of genetic diversity due to the small size of their populations; however, it is necessary to evaluate them with molecular markers, since in some cases it has been found that these species may have moderate or high levels of genetic diversity (Forrest *et al.*, 2017).

Average annual temperature (bio1) plays an important role in the distribution of *S. tacaco*, although it is a semi-cultivated species, it is difficult to determine the optimal temperature and irrigation requirements as a crop due to the scarce agronomic research.



**Figure 2.** SDMs for *S. tacaco* in Costa Rica. A. Generalized Linear Model (GLM) B. Generalized Additive Model (GAM) C. Generalized Potentiated Regression Model (GBM) D. Classification Tree Analysis (CTA).



**Figure 3.** SDMs of *S. tacaco* in Costa Rica. Techniques based on machine learning, A. RF, B. Maxent, C. SVM and D. Assembling of seven models.

Figures 1 and 2, in agreement with Monge and Loría (2017), show a possible trend that *S. tacaco* may be introduced or distributed in Panama, especially with ACC, RF, Maxent y SVM models. This constitutes an important aspect to investigate, agronomic response of *S. tacaco* in other edaphoclimatic conditions. Andosol and Cambisol soil that predominate in two thirds of the occurrence points, are characterized by being of volcanic origin and having a high cation exchange capacity, which is positively correlated with the amount of organic matter (Dai *et al.*, 2018).

Distribution models are predictions of the suitability for species, they provide relevant information on geographic regions that contain germplasm of interest, which is undoubtedly important for conservation and breeding studies. The previous approach is relevant if it is considered that seeds of *Sechium* species are recalcitrant and cannot be conserved in seed banks. Due to its endocarpic nature, the seed germinates even within the fruit, and it does not show signs of senescence, since it is not carotenogenic, that is, it does not change color. In the case of *S. tacaco*, it is relevant to determine real and potential distribution areas for in situ conservation, and for the collection of specimens that contribute to maintaining its diversity.

## CONCLUSIONS

Generalized Additive Model (GAM) turned out to have the best area under the curve (AUC) and TSS ((Specificity + Sensitivity) – 1); however, all seven models and the assembly showed adequate performance. Bioclimatic variables related to temperature are the ones with the greatest contribution to models and the main limitations in the distribution of *S. tacaco*, which, being a species with restricted distribution, requires very specific climatic conditions. Areas with high suitability are found in the central valleys of Costa Rica, covering regions of the provinces of Alajuela, Cartago, San José and Puntarenas, which can be sources of germplasm for future conservation and breeding studies.

## REFERENCES

- 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), pp.1223-1232. DOI: /10.1111/j.1365-2664.2006.01214.x
- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, 5(1), pp.1-12. Doi: 10.1038/sdata.2018.214
- Beck, J. (2012). Predicting climate change effects on agriculture from ecological niche modeling: Who profits, who loses? *Climatic Change*, 116 (2), pp. 177-189. Doi: 10.1007/s10584-012-0481-x
- Evans, J. M., Fletcher, R. J., & Alavalapati, J. (2010). Using species distribution models to identify suitable areas for biofuel feedstock production. *GCB Bioenergy*, 2(2), pp.63-78. Doi: 10.1111/j.1757-1707.2010.01040.x
- Feng, X., Park, D. S., Liang, Y., Pandey, R., & Papeş, M. (2019). Collinearity in ecological niche modeling: Confusions and challenges. *Ecology and Evolution*, 9(18), pp.10365-10376. Doi: 10.1002/ece3.5555
- 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), pp.4302-4315. Doi: 10.1002/joc.5086
- Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H. T., Verelst, D., & Wiberg, D. (2008). *Global Agro-ecological Zones Assessment for Agriculture*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.
- Flores-Tolentino, M., Ortiz, E., Villaseñor, J. L., Flores-Tolentino, M., Ortiz, E., & Villaseñor, J. L. (2019). Ecological niche models as a tool for estimating the distribution of plant communities. *Revista Mexicana de Biodiversidad*, 90. Doi: 10.22201/ib.20078706e.2019.90.2829
- González-Santos, R., Cadena-Íñiguez, J., Morales-Flores, F. J., Ruiz-Vera, V. M., & 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), pp.791-804. Doi: 10.1007/s10722-016-0401-4



- Greenwell, B., Boehmke, B., Cunningham, J., & GBM Developers. (2020). *gbm: Generalized boosted regression models* (R package version 2.1.8) [Computer software]. <https://CRAN.R-project.org/package=gbm>
- Hijmans, R. J. (2020). *raster: Geographic Data Analysis and Modeling* (R package version 3.3-13) [Computer software]. <https://CRAN.R-project.org/package=raster>
- Hijmans, R. J., & Elith, J. (2013). *Species distribution modeling with R*. Disponible en: <http://www2.uaem.mx/r-mirror/web/packages/dismo/vignettes/sdm.pdf>
- Hijmans, R. J., Phillips, S. J., Leatwick, J., & Elith, J. (2017). *Dismo: Species Distribution Modeling. R package version 1.1-4*. <https://CRAN.R-project.org/package=dismo>
- 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. *International Journal of Pest Management*, 65, 1-17. Doi: 10.1080/09670874.2018.1533664
- Liaw, M., & Wiener, M. (2002). Classification and Regression by random Forest. *R News*, 2(3), pp.18-22.
- Lira, S. R. (1995). Estudios taxonómicos en el género *Sechium* P. Br. Cucurbitaceae [Tesis Doctoral]. Universidad Nacional Autónoma de México.
- 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(2), pp.145-151. Doi: 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, pp.217-240. Doi: 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 1.7-3) [Computer software]. <https://CRAN.R-project.org/package=e1071>
- Monge, P. J. E., & Loría, M. (2017). Caracterización de frutos de cinco genotipos de tacaco [*Sechium tacaco* (Pittier) C. Jeffrey] en Costa Rica | Revista Tecnología en Marcha. *Tecnología en Marcha*, 30(3), pp.71-84. Doi:10.18845/tm.v30i3.3274
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J., Uriarte, M., & 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), pp.1198-1205. Doi: 10.1111/2041-210X.12261
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3), pp.231-259. Doi: 10.1016/j.ecolmodel.2005.03.026
- Polce, C., Termansen, M., Aguirre-Gutiérrez, J., Boatman, N., Budge, G., Crowe, A., Garratt, M., Pietravalle, S., Potts, S., Ramirez, J., Somerwill, K., & Biesmeijer, J. (2013). Species Distribution Models for Crop Pollination: A Modelling Framework Applied to Great Britain. *PloS one*, 8, e76308. Doi: 10.1371/journal.pone.0076308
- QGIS Development Team. (2020). *QGIS Geographic Information System*. Open Source Geospatial Foundation Project (3.16.2) [Computer software]. <https://www.qgis.org/en/site/>
- R Core Team. (2020). *R: A language and environment for statistical computing*. *R Foundation for Statistical Computing* (1.3.1093) [Computer software]. <https://www.R-project.org/>
- Schmitt, S., Pouteau, R., Justeau, D., Boissieu, F. de, & 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), pp.1795-1803. Doi: 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), pp.1427-1438. Doi:10.1111/j.1365-2699.2012.02695.x
- Therneau, T., & Atkinson, B. (2019). *rpart: Recursive partitioning and regression trees* (R package version 4.1-15) [Computer software]. <https://CRAN.R-project.org/package=rpart>
- Wood, S. N. (2017). *Generalized additive models (Second Edition)*. Chapman and Hall CRC Press.
- Wunderlin, R. P. (1976). Two new species and a new combination in *Frantzia* (Cucurbitaceae). *Brittonia*, 28(2), pp.239-244. JSTOR. Doi: 10.2307/2805833