

Análisis de vacíos en la colecta de especies del género Passiflora L. en Colombia

María Victoria Díaz López

Tutora: Anna Muñoz Bolas

Profesor: Albert Solé Ribalta

Universitat Oberta de Catalunya

Maestría en Ciencia de Datos

21 de diciembre de 2023

Contenido

- 1 Introducción
- 2 Objetivos
- 3 Enfoque y metodología
 - Modelos de distribución de especies (SDM)
 - Conservación ex situ
 - Validación
 - Mapa final de vacío
- 4 Diseño e implementación del trabajo
 - Resultados
- 5 Conclusiones
- 6 Bibliografía

Introducción

Contexto y justificación

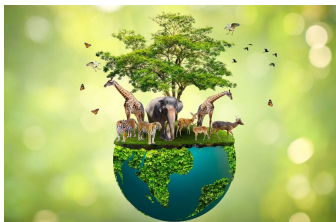


Figura: Importancia de la biodiversidad



Figura: Semillas del Futuro (CIAT)

"La biodiversidad está bajo amenaza de extinción y erosión". Maxted (2001)

Variedades tradicionales!

Objetivos

Generales

- Analizar los vacíos en conservación de especies de *Passiflora* L. en Colombia.
- Encontrar una herramienta de modelación de distribución de especies que obtenga las mejores predicciones y el mejor rendimiento.

Específicos

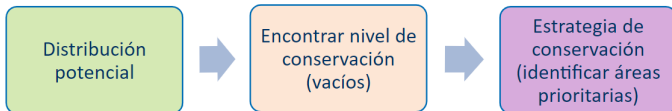
- Revisar la literatura y escoger las(s) especie(s) a analizar.
- Ajustar y comparar tres métodos de modelación de la distribución espacial de especies de plantas.
- Determinar la importancia de diferentes variables medioambientales en el desarrollo de las especies escogidas.
- Encontrar la distribución geográfica potencial las especies a lo largo del territorio colombiano.
- Hallar las regiones de Colombia en donde el estatus de conservación ex situ de las especies es deficiente.
- Mapear la distribución potencial de las especies.
- Mapear las regiones con alta prioridad en conservación de las especies.
- Calcular el porcentaje área de la distribución potencial en el que se considera hay un buen estatus de conservación.

Enfoque y metodología

Análisis de vacíos:

Método de análisis del grado de conservación, con el fin de identificar especies insuficientemente representadas en los sistemas de conservación (Maxted et al. 2008).

Compara la representación geográfica de la colección ex situ con una distribución potencial basada en datos ecogeográficos y de bancos de germoplasma.



Área de estudio



Figura: Mapa político de Colombia

Datos del cultivo

Especie	NV	Departamento	Municipio	Longitud	Latitud
P. edulis f. flavicarpa	Maracuyá	Valle del Cauca	Vijes	-76.43049	3.70980
P. edulis f. flavicarpa	Maracuyá	Valle del Cauca	Vijes	-76.43050	3.70980
P. edulis f. flavicarpa	Maracuyá	Valle del Cauca	Yumbo	-76.45440	3.67276
P. edulis f. flavicarpa	Maracuyá	Valle del Cauca	Zarzal	-75.99471	4.35821
P. edulis f. flavicarpa	Maracuyá	Valle del Cauca	Ginebra	-76.24273	3.72610
P. edulis f. flavicarpa	Maracuyá	Valle del Cauca	Ginebra	-76.27950	3.74140
P. edulis f. flavicarpa	Maracuyá	Valle del Cauca	Ginebra	-76.27979	3.74030
P. edulis f. flavicarpa	Maracuyá	Valle del Cauca	La Unión	-76.11743	4.53064
P. edulis f. flavicarpa	Maracuyá	Valle del Cauca	Toro	-76.05132	4.66970
P. edulis f. flavicarpa	Maracuyá	Valle del Cauca	Roldanillo	-76.14670	4.42686

Passiflora L.

80/610 comestibles



13 especies en Colombia

Fuentes

Alianza Internacional de Bioversity y CIAT.
Universidad Nacional de Colombia.

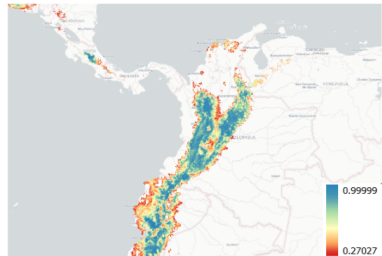
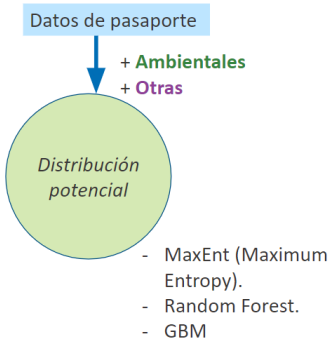
Variables predictoras

50 predictores potenciales, espacialmente explícitos.

- Climáticas.
- Topográficas.
- Historia evolutiva de los cultivos (por ejemplo, distancia a asentamientos humanos).
- Variables atropogénicas (por ejemplo, densidad de población y accesibilidad, grupos étnicos).

SDM

Los predictores espaciales asociados a la especie se utilizan para modelar la probabilidad de aparición geográfica de la misma, dentro del área de estudio.



0: Baja probabilidad de presencia.

1: Alta probabilidad de presencia.

SDM - Pseudoausencias

Muestra aleatoria de $n = 10$ veces más grande la cantidad de puntos de presencia.

- SVM: Zonas potencialmente adecuadas para la presencia de la especie.
- Lugares a más de 5KMs de distancia de las presencias.
- Muestra aleatoria en lugares seleccionados.

SDM - Selección de variables

- PCA. Descarte de variables con menor contribución.
- VIF. Descarte de variables con valor menor a 10. (teóricamente).
- Construcción de la base de datos final.

SDM - Calibración de parámetros MaxEnt

- Entrenamiento de 100 modelos MaxEnt (trainMaxEnt).
- Menor AICc. Caso de empate: Menos features, menos coeficientes y mayor parámetro de regularización.

Se probaron parámetros de regularización en secuencia de 0.5 a 6 de cinco en cinco. Los features ajustados fueron: linear, product, quadratic, hinge.

SDM - Tuning de hiperparámetros (GBM y RF)

Validación cruzada.

SDM - Ajuste

Ajuste de los modelos, usando $k = 5$ folds para validación cruzada, con un 80 % de los datos para entrenamiento y 20 % para testeo.

El modelo final será la mediana de los 5 modelos ajustados.

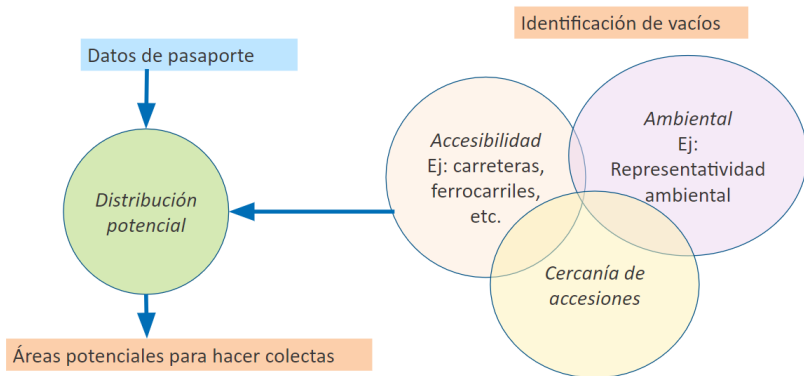
Ensemble

Combinación de las tres predicciones mediante la media ponderada por el AUC de cada modelo.

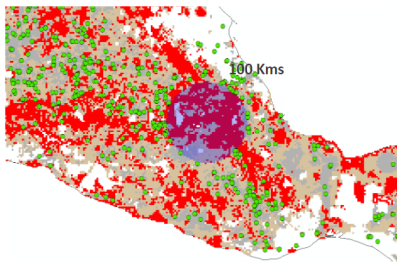
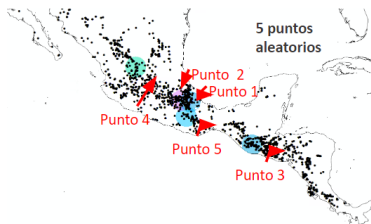
Evaluación

- Área bajo la curva ROC.
- Sensibilidad.
- Especificidad.
- cAUC.

Análisis de representación ex situ de las especies



Validación



Límites: Promedios de los valores que maximizan la predicción (AUC) de los cinco vacíos en cada métrica.

Rendimiento de las predicciones: AUC, sensibilidad y especificidad.

Predicción de vacío

Los píxeles con valores por encima de los límites serán marcados con valor de 1, e indican alta probabilidad de vacío en conservación, mientras que los otros con valores por debajo de los límites serán marcados con 0 e indican baja o poca probabilidad de vacío, por lo que representarían áreas donde el cultivo está “bien conservado”.

Suma:

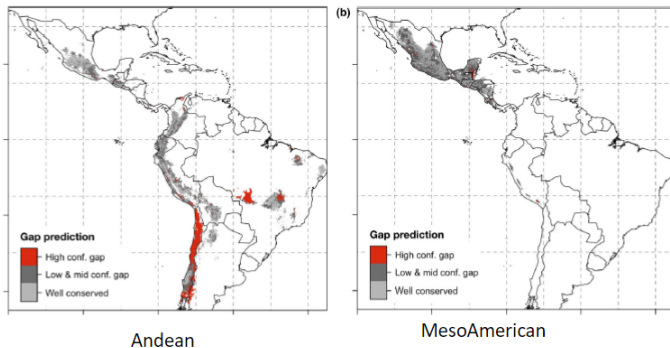


Figura: Frijol común. Ramírez-Villegas et al., 2020

Diseño e implementación del trabajo

21 de diciembre de 2023

Modelos de distribución de especies - Maracuyá

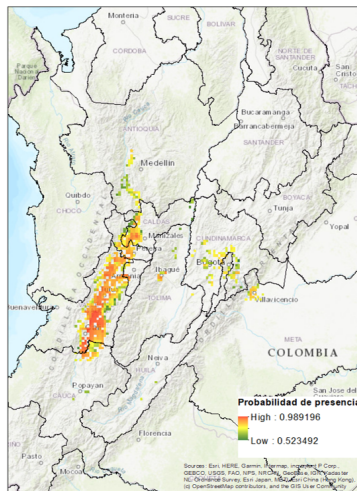


Figura: Distribución potencial del Maracuyá. Ensemble

Evaluación de los modelos

MAXENT					
Fold	1	2	3	4	5
Sensibilidad	0.857143	0.9	0.875	0.888889	0.6
Especificidad	0.924051	0.921053	0.923077	0.935065	0.9625
Threshold	0.225	0.155	0.187	0.217	0.221
Corr. Mathews	0.616474	0.69359	0.647149	0.704103	0.51673
Índice de Kappa	0.589359	0.674595	0.623279	0.688782	0.514286
AUC	0.971067	0.964474	0.983974	0.97114	0.875
RANDOM FOREST					
Fold	1	2	3	5	5
Sensibilidad	0.857143	0.6	0.875	0.888889	0.6
Especificidad	0.987342	1	1	1	1
Threshold	0.327	0.347	0.347	0.367	0.348
Índice de Kappa	0.844485	0.726115	0.926995	0.93475	0.738462
AUC	0.971067	0.896053	0.996795	0.974026	0.9

Evaluación de los modelos

GBM					
Fold	1	2	3	4	5
Sensibilidad	0.857143	0.6	0.75	0.888889	0.4
Especificidad	0.987342	1	1	1	1
Threshold	0.453	0.39	0.44	0.404	0.472
Índice de Kappa	0.844485	0.726115	0.844765	0.93475	0.556522
AUC	0.909584	0.843421	0.945513	0.948052	0.7925

Evaluación de modelos - Ensemble

AUC

AUC = 0.952

cAUC

cAUC = 0.546

Modelos de distribución de especies - Granadilla

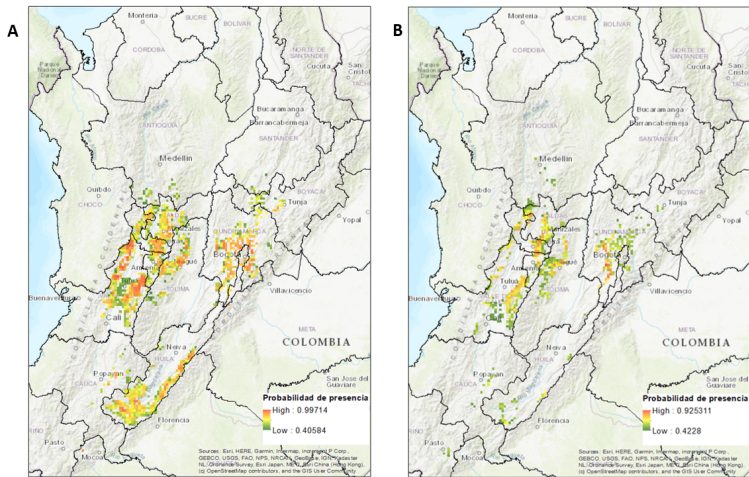


Figura: Distribución potencial de la Granadilla. (A) MaxEnt. (B) Random Forest.

Modelos de distribución de especies - Granadilla

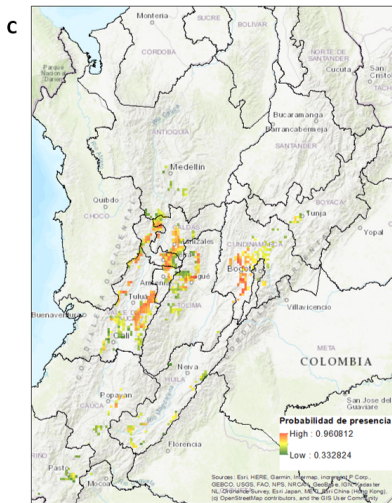


Figura: Distribución potencial de la Granadilla.(C) GBM

Modelos de distribución de especies - Granadilla

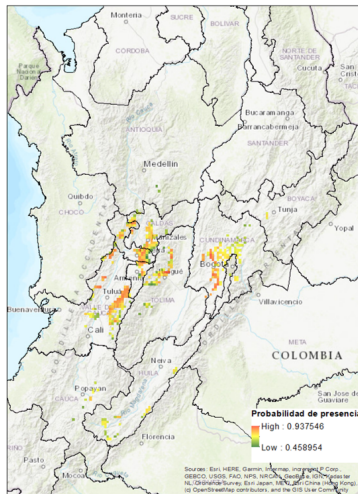


Figura: Distribución potencial de la Granadilla.Ensemble

Evaluación de los modelos

MAXENT					
Fold	1	2	3	4	5
Sensibilidad	0.888889	0.666667	1	1	0.833333
Especificidad	0.878788	0.972222	0.984615	0.956522	1
Threshold	0.299	0.216	0.222	0.208	0.341
Corr. Mathews	0.608956	0.557177	0.9461	0.798549	0.906232
Índice de Kappa	0.769585	0.473684	0.801762	0.747475	0.268775
AUC	0.962963	0.986111	0.996923	0.987923	0.977941
RANDOM FOREST					
Fold	1	2	3	4	5
Sensibilidad	0.888889	0.666667	0.7	0.833333	0.166667
Especificidad	0.939394	0.958333	1	0.971015	1
Threshold	0.401	0.414	0.38	0.396	0.423
Índice de Kappa	0.724062	0.473684	0.801762	0.747475	0.268775
AUC	0.969697	0.976852	0.993846	0.992754	0.921569

Evaluación de los modelos

GBM					
Fold	1	2	3	4	5
Sensibilidad	0.888889	0.666667	0.7	0.833333	0.166667
Especificidad	0.954546	0.958333	1	0.971015	1
Threshold	0.299	0.216	0.222	0.208	0.341
Índice de Kappa	0.769585	0.473684	0.801762	0.747475	0.268775
AUC	0.959596	0.986111	0.981539	0.983092	0.879902

Evaluación de modelos - Ensemble

AUC

$AUC = 0.979$

cAUC

$cAUC = 0.527$

Importancia de variables

Para el Maracuyá, las variables más importantes son:

- Accesibilidad.
 - Irrigación.
 - Densidad poblacional
 - Distancia a ríos.
 - Precipitación en el trimestre más caluroso.
- Relación con la intervención humana y su cercanía a la civilización

Para la Granadilla, las variables más importantes son:

- Variabilidad mensual de evapotranspiración potencial.
- Irrigación.
- Accesibilidad.
- Densidad poblacional
- Presión del vapor de agua en el ambiente.

Importancia de un clima menos húmedo, más que de la intervención humana.

Predicción de vacío final

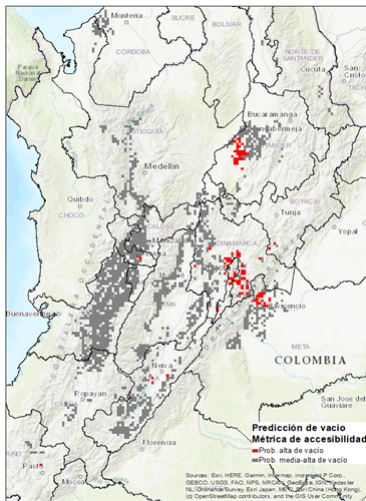


Figura: Maracuyá

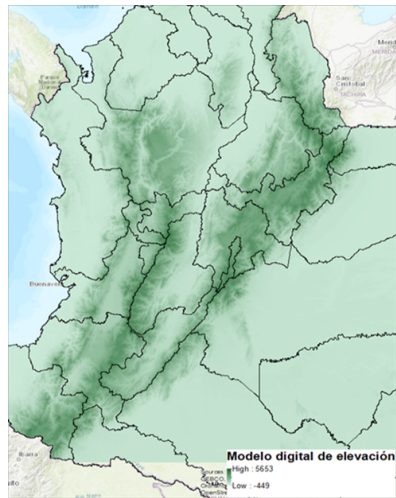


Figura: Modelo Digital de Elevación

Predicción de vacío final

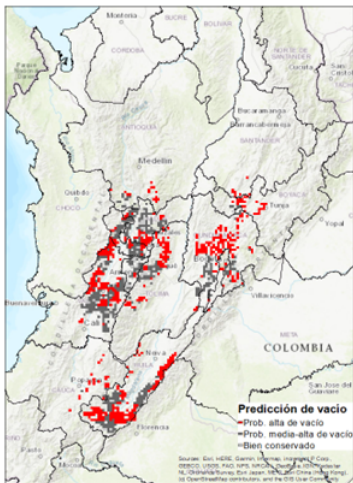


Figura: Granadilla

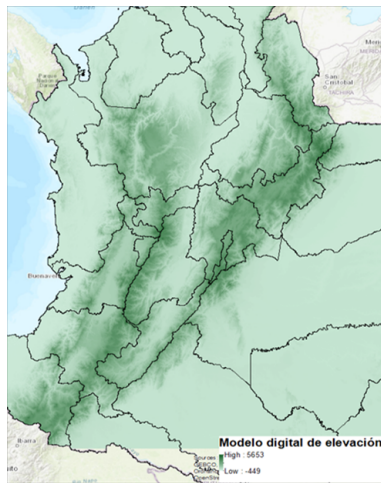


Figura: Modelo Digital de Elevación

Estado de conservación

- Entre el 19.7 % y el 49.9 % de la distribución potencial predicha de la Granadilla se encuentra bien conservado, y el área bajo la curva ROC del modelo de vacío fue de un 75.04 %.
- Entre el 70.3 % y el 93.9 % de la distribución potencial predicha del Maracuyá se encuentra bien conservado, y el área bajo la curva ROC del modelo de vacío fue de un 92.8 %.

Conclusiones

- Maracuyá y Granadilla aportan en gran medida al PIB Per Cápita del país y son ampliamente consumidos.
- Conocimiento de modelos de distribución de especies distintos al tradicional MaxEnt que, además de brindar resultados con igual o mejor rendimiento, son más eficientes en cuanto a costo computacional y tiempo de ejecución.
- La predicción de la distribución de las plantas cultivadas, cuya área de distribución está determinada por factores antropogénicos y medioambientales, es un reto que aún no se ha resuelto del todo en las ciencias geoespaciales.
- El Maracuyá es propio de lugares con menores altitudes y climas más húmedos, pero con menores precipitaciones en comparación con la Granadilla. Ambos se dan en zonas cafeteras.
- Tanto para Maracuyá como para Granadilla son muy importantes las variables de carácter antropogénico y accesibilidad; sin embargo, para predecir la distribución de la Granadilla son más importantes las variables de humedad de suelo (o evapotranspiración), y de precipitación.
- Los lugares en lo que es más probable encontrar un vacío en conservación son en su mayoría difíciles de acceder, pero existen otros en los que la situación de orden público no permite su cultivación.

La Granadilla es un cultivo que necesita más atención en cuanto a conservación se refiere, ya que además menos del 50 % de su distribución potencial se encuentra conservado en bancos de germoplasma o el campo colombiano, mientras que entre un 70 y 93 % de la distribución potencial del Maracuyá se encuentra bien conservado y suficientemente colectado.

Bibliografía

- Burley, F.W. 1988. Monitoring biological diversity for setting priorities in conservation. In Biodiversity. pp. 227–230. Edited by E.O. Wilson. National Academy Press, Washington, DC. 521 pp.
- Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G., Hilton-Taylor, C. (2002). Habitat Loss and Extinction in the Hotspots of Biodiversity. Conservation Biology. 16 (4). pp. 909-923. Wiley.
- Brooks, T.M., Bakarr, M.I., Boucher, T., da Fonesca, G.A.B., Hilton-Taylor, C., Hoekstra, J.M. (2004). Coverage provided by the global PA system: is it enough? Bioscience 54: 1081–1091.
- Brush, S. B. (1995). In situ conservation of landraces in centers of crop diversity. Crop Sci. 35, 346
- Buchhorn, M., Smets, B., Bertels, L., Lesiv, M., Tsendbazar, N.-E., Masiliunas, D., Linlin, L., Herold, M., Fritz, S. (2020). Copernicus Global Land Service: Land Cover 100m: Collection 3: epoch 2019: Globe (Version V3.0.1) [Data set]. Zenodo. DOI: 10.5281/zenodo.3939050
- Camacho Villa, T. C., Maxted, N., Scholten, M. and Ford-Lloyd, B. (2005). Defining and identifying crop landraces. Plant Genet. Resour. Charact. Util. 3, 373–384.

- Convention on Biological Diversity. (2010). Strategic Plan for Biodiversity 2011-2020, including Aichi Biodiversity Targets.
- Debouck, D.G., Libreros Ferla, D. 1995. Neotropical montane forests: a fragile home of genetic resources of wild relatives of new world crops. In: Churchill, S.P., Baslev, H., Forero, E., Luteyn, J.L. (eds.). Biodiversity and conservation of neotropical montane forests, New York, USA, pp. 561-577.
- Debouck, D.G., Libreros Ferla, D. 1995. Neotropical montane forests: a fragile home of genetic resources of wild relatives of new world crops. In: Churchill, S.P., Baslev, H., Forero, E., Luteyn, J.L. (eds.). Biodiversity and conservation of neotropical montane forests, New York, USA, pp. 561-577.
- Dietz, R.W. y Czech, B. (2005). Conservation deficits for the continental United States: an ecosystem gap analysis. *Conservation Biology* 19: 1478–1487.
- Ellis, C. (s.f.). Mtry in random forests.
<https://crunchingthedata.com/mtry-in-random-forests/>
- Elith, J., Leathwick, J.R. and Hastie, T. (2008), A working guide to boosted regression trees. *Journal of Animal Ecology*. 77: 802-813.
<https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Elith, J. and Leathwick, J. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40, pp. 677–697.

- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*. 17: 43-57
- Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315.
- Food and Agriculture Organization (FAO). (1998). *The State of the World's Plant Genetic Resources for Food and Agriculture*. Rome: FAO.
- Guisan, A. and Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8, pp. 993–1009.
- Harlan, JR. (1975). Our vanishing genetic resources. *Science* 188:618 – 621.
- Hawkes, JG. 1983. *The Diversity of Crop Plants*. Cambridge, MA: Harvard University Press, p. 102.
- Hurtado, A., Ocampo, J., Ceballos-Aguirre, N., García, D.J., López, W.R. 2021. *Passiflora: Genetic, Grafting and Biotechnology Approaches*. Nova Science Publishers. New York.
- Ingram, G.B. y Williams, J.T. (1993). Gap analysis for in situ conservation of crop gene pools: implications of the Convention on Biological Diversity. *Biodiversity Letters* 1(5): 141–148.
- Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E. Hole-filled. (2008). *Seamless SRTM Data V4*. International Center for Tropical Agriculture.
<https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4>

- Khoury, C. K., Amariles, D., Soto, J.S., Diaz, M.V., Sotelo, S., Sosa, C., et. al. (2019). Comprehensiveness of conservation of useful wild plants: an operational indicator for biodiversity and sustainable development targets. *Ecol Ind.* 98, 420–429.
- Konowalik, K., Nosol, A. (2021). Evaluation metrics and validation of presence-only species distribution models based on distributional maps with varying coverage. *Sci Rep* 11, 1482.
- MacNeely J.A., K.R. Miller., N.A. Reid., R.A. Mittermer., T.B. Wainer. (1990). Conserving the world's biological diversity. World Conservation Union, World Resources Institute, World Wildlife Fund – U.S. World Bank, Washington, D.C.
- May R.M. (1992). How many species inhabit the Earth? *Scientific American* 267(4): 18-24.
- Maxted, N. (2001). Ex situ, In situ conservation. *Encyclopedia of Biodiversity*. Volume 2. University of Birmingham. Academy Press. England.
- Maxted, N., Dulloo, E., V Ford-Lloyd, B., Iriondo, J.M. and Jarvis, A. (2008). Gap analysis: a tool for complementary genetic conservation assessment. *Diversity and Distributions*, 14: 1018-1030.
- National Plant Germplasm System (NPGS). (1991). GRIN—Forage Legume Data Dictionary. <http://www.ars-grin.gov/npgs/foragedd.pdf>.
- Ocampo, J.; Coppens d'Eeckenbrugge, G.; Restrepo, M.; Jarvis, A.; Salazar, M.; Caetano, C. (2007). Diversity of Colombian Passifloraceae: biogeography and an updated list for conservation. *Biota Colombiana*. 8 (1).
- O'Donell, M.S., e Ignizio, D.A. (2012). Bioclimatic Predictors for Supporting Ecological Applications in the Conterminous United States. *Geological Data Series* 691, 10 p.

- Phillips, S. J., Anderson, R.P., Schapire, R.P. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190: 231-259.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. and Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19, 181–197.
- Phillips, N.D.; Reid, N.; Thys, T.; Harrod, C.; Payne, N.L.; Morgan, C.A.; White, H.J.; Porter, S.; Houghton, J.D.R. Applying species distribution modelling to a data poor, pelagic fish complex: The ocean sunfishes. *J. Biogeogr.* 2017, 44, 2176–2187.
- Ramirez-Villegas, J., Khoury, C.K., Achicanoy, H.A., Mendez, A.C., Diaz, M.V., Sosa, C., et. al. (2020). A gap analysis modelling framework to prioritize collecting for ex situ conservation of crop landraces. *Divers. Distrib.* 26, 730–742.
- Sustainable development goals. United Nations. En: <https://sdgs.un.org/2030agenda> (2015).
- SDG Indicators, Metadata Repository, Goal 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture. United Nations.
- Title, P. O. and Bemmels, J. B. (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* 41, 291–307.