Species distribution modeling and conservation assessment of the northwestern Argentinian highland papayas under global change scenarios

SUPPLEMENTARY MATERIAL

Table S1. Environmental conditions for probabilities higher than 0.8 of each species.

Tabla S1. Condiciones ambientales para probabilidades superiores a 0.8 de cada especie.

Environmental variable	*V. glandulosa*	*V. quercifolia*
Annual mean temperature (°C)	<18	<17
Mean diurnal range (°C)	>13.5	<13
Temperature seasonality	<3900	>5000
Annual precipitation (mm)	>650	>700
Precipitation of driest month (mm)	2.5-22.5	7.5-22.5

Figure S1. Occurrence records (germplasm accessions and herbarium/sighting records) for *V. glandulosa* (A) and *V. quercifolia* (B) in northwestern Argentina.

Figura S1. Registros de ocurrencia (accesiones de germoplasma y registros de herbario/observación) para *V. glandulosa* (A) y *V. quercifolia* (B) en el noroeste argentino.

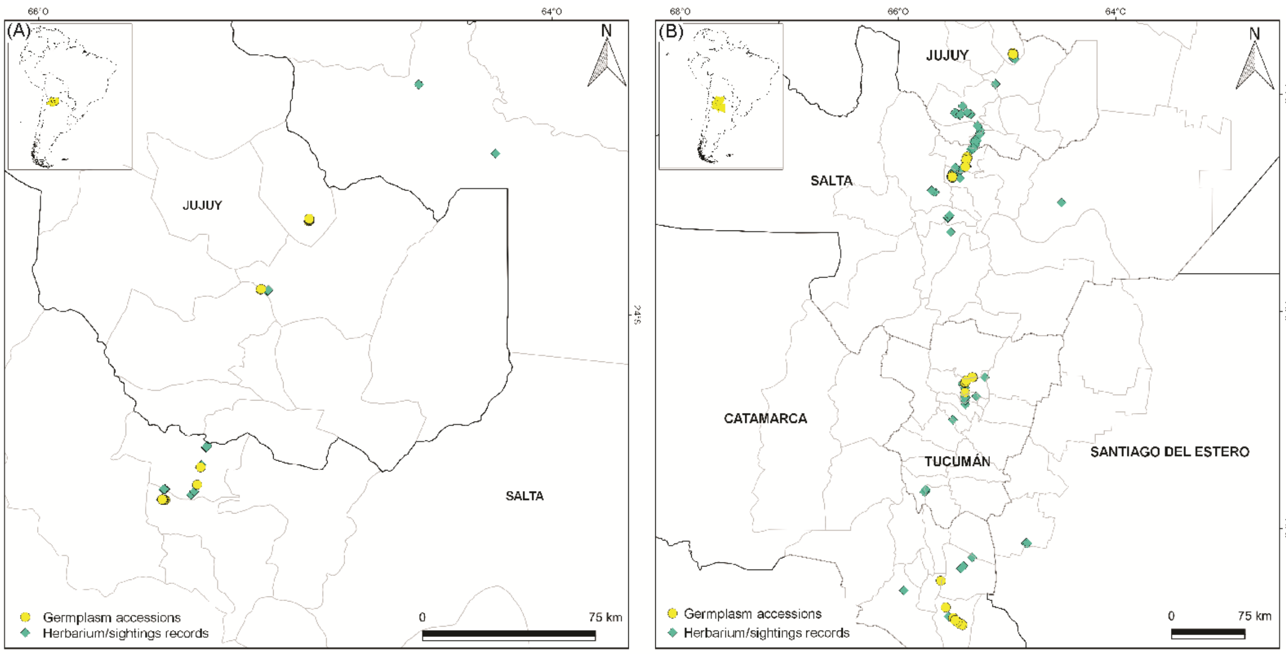


Figure S2. Jackknife analysis of *V. glandulosa* model variable contribution.

Figura S2. Análisis de Jackknife de la contribución de las variables al modelo de *V. glandulosa*.

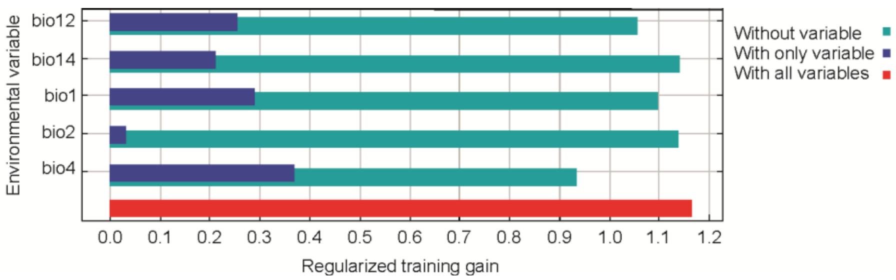


Figure S3. Jackknife analysis of *V. quercifolia* model variable contribution

Figura S3. Análisis de Jackknife de la contribución de las variables al modelo de *V. quercifolia*.

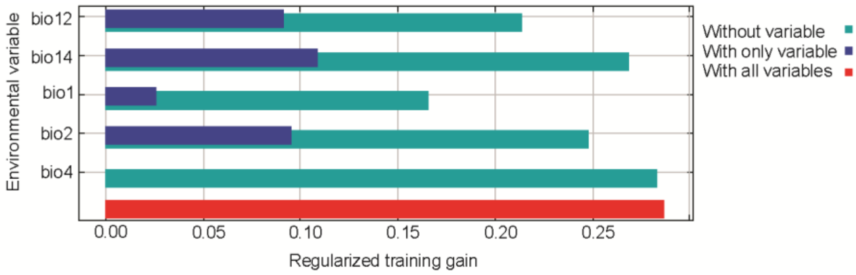


Figure S4. Response curves of the five bioclimatic variables for *V. glandulosa*. Bio1: annual mean temperature (temperature*10); bio2: mean diurnal range; bio4: temperature seasonality; bio12: annual precipitation; bio 14: precipitation of the driest month.

Figura S4. Curvas de respuesta de las cinco variables bioclimáticas para *V. glandulosa*. Bio1: temperatura media anual (temperatura*10); bio2: rango diurno medio; bio4: estacionalidad de la temperatura; bio12: precipitación anual; bio 14: precipitación del mes más seco.

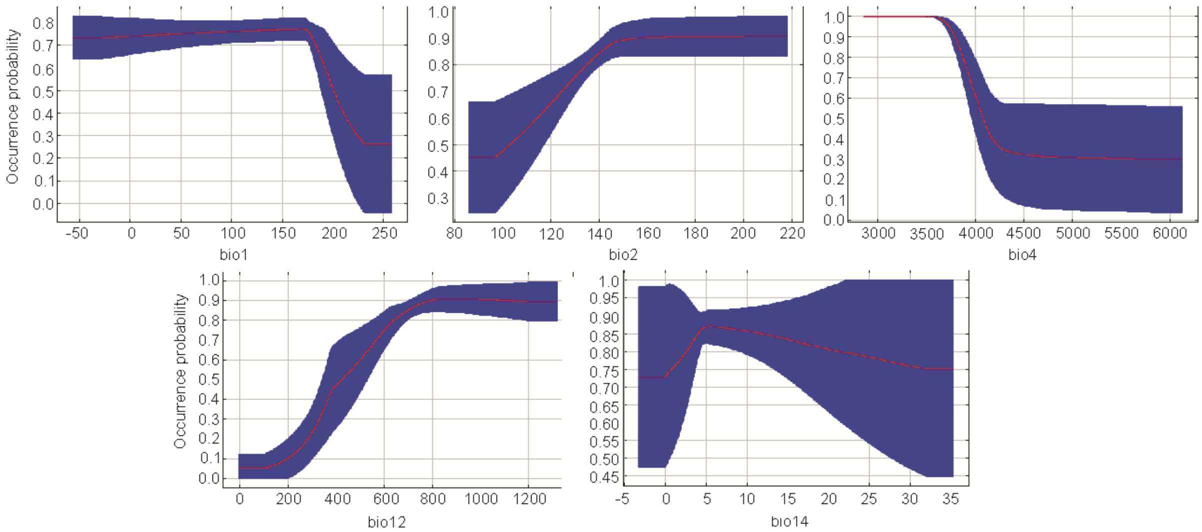
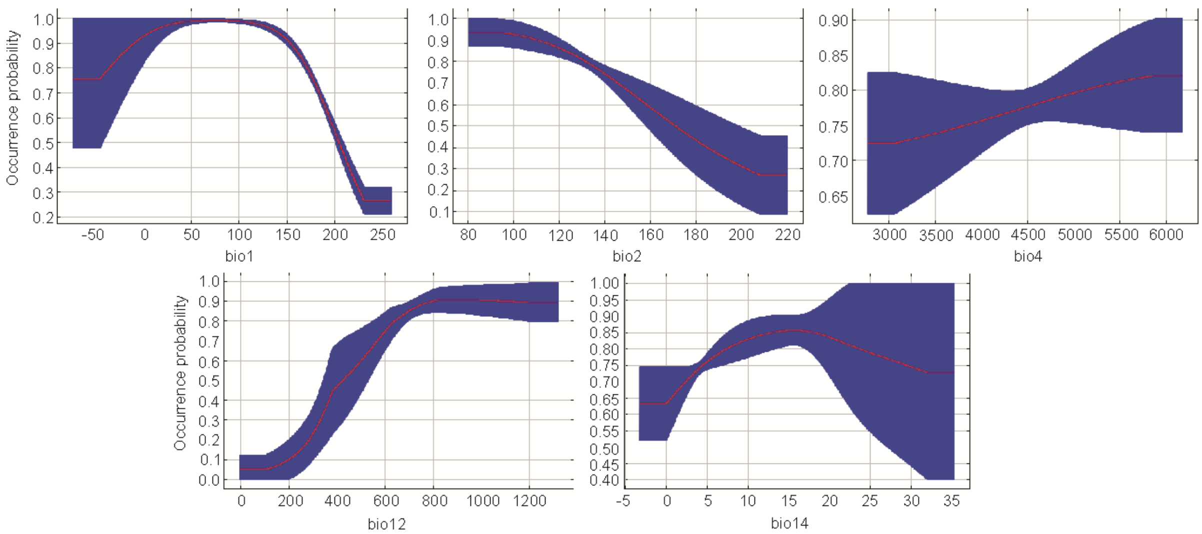


Figure S5. Environmental variables response curves for *V. quercifolia*. Bio1: annual mean temperature (temperature*10); bio2: mean diurnal range; bio4: temperature seasonality; bio12: annual precipitation; bio 14: precipitation of the driest month.

Figura S5. Curvas de respuesta de las variables ambientales para *V. quercifolia*. Bio1: temperatura media anual (temperatura*10); bio2: rango diurno medio; bio4: estacionalidad de la temperatura; bio12: precipitación anual; bio 14: precipitación del mes más seco.



METHODOLOGICAL DETAILS

Model settings

The modeling tuning procedure was performed using the *Run Maxent: Spatial Jackknifing tool* of SDMToolbox (Brown 2014). The procedure consists of three steps:

1. Spatial Jackknifing (or geographically structured k-fold cross-validation): the landscape is split into three regions based on spatial clustering occurrence points. Models are calibrated with k-1 spatial groups and then evaluated with the withheld group. We selected a k value of 3. In this case, models were run with following three subgroups:
- Model was calibrated with localities and background points from region AB and then evaluated with points from region C.
- Model was calibrated with localities and background points from region AC and then evaluated with points from region B.
- Model was calibrated with localities and background points from region BC and then evaluated with points from region A.

2. Independent tests of model feature classes and regularization parameters: this analysis allows for testing different combinations of four model feature class types and regularization multipliers (RM) to optimize your MaxEnt model performance. The following combinations were tested:
- RM 1 – 5 & Linear
- RM 1 – 5 & Linear and Quadratic
- RM 1 – 5 & Hinge
- RM 1 – 5 & Linear, Quadratic and Hinge
- RM 1 – 5 & Linear, Quadratic, Hinge and Product

3. Automatic model selection: finally, the best model was selected by evaluating each model's: 1. test data omission error rates (OER), 2. test data AUC, and 3. model feature class complexity. The script does this in order, chooses the model with the lowest omission rates on the test data. If many models have the identical low OER, then it selects the model with the highest AUC (using only test data to calculate this). Lastly if several models have the same low OER and high AUC, it will choose the model with simplest feature class parameters in the following order (1. linear; 2. linear and quadratic; 3. hinge; 4. linear, quadratic, and hinge; and 5. linear, quadratic, hinge, and product).

Once the best model was selected, SDMtoolbox run the final model using all the occurrence points (Brown 2014). Ten thousands background points randomly distributed in each species-specific geographical background were used with *randomseed* option. This implies that a different random train/test partition (80/20% respectively) was made and a different subset of the background was used in each model replicate. Although internal evaluation does not allow assessing the predictive performance of the models (independent evaluation data would be required for this purpose), it does provide a measure of internal consistency of the models (Araujo et al. 2011). The option clamping was selected for future projections.

Gap analysis - data acquisition

The germplasm data to calculate the SRS were obtained from international online information systems (Genesys - https://www.genesys-pgr.org/), national (INTA - https://www.inta.gov.ar) and regional repositories (Red Argentina de Bancos de Germoplasma de Plantas Nativas – Red

ARGENA - https://www.fca.unl.edu.ar/vinculacion/red-argena/, personal communication). The latter was included because it is composed of national universities and public ministries that work exclusively with native plant species with agronomic, medicinal or industrial potential.

REFERENCES

Araújo, M. B., Alagador, D., Cabeza, M., Nogués-Bravo, D., and Thuiller, W. 2011. Climate change threatens European conservation areas. Ecology letters 14(5), 484-492.

Brown, J. L. 2014. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. Methods in Ecology and Evolution 5:694–700.