International Journal of Botany

Species Distribution Modelling of Family Sapindaceae in West Africa

T.O. Adeyemi, O.T. Ogundipe and J.D. Olowokudejo
Molecular Systematics Laboratory, Department of Botany, University of Lagos,
Akoka, Lagos, Nigeria

Abstract: Members of the family Sapindaceae provides various benefits in combating increasing malnutrition and poverty in West Africa. Due to their multiple uses, their high nutritional and medicinal value, members of Sapindaceae have been identified as one of the most important forest species to be conserved and valued in Africa. Therefore, it is important to study the potential future distribution of this species and determine strategies for conservation. In order to achieve these, suitability of sites in Africa was evaluated for potential conservation using spatial environmental data in MAXENT modelling framework. A total of 153 geo-referenced records of Sapindaceae were assembled from herbarium records and fieldwork and climatic data were acquired from the Worldclim Database. The main variables that contributed towards predicting the species distribution were annual precipitation and temperature seasonality. Results suggest that the distribution model was excellent with training AUC value of 0.974 and test AUC value of 0.960 confirming the wide distribution of Sapindaceae in West Africa. Also, the environmental variables that affected the model the most are minimum temperature of the coldest month (35.4%), precipitation of the coldest quarter (16.9%) and precipitation of wettest month (14.9%). Recommendations for different conservation strategies include *in situ* conservation in Protected Areas; *ex situ* conservation in seed banks and conservation through 'sustainable utilization'.

Key words: Sapindaceae, Africa, conservation, modelling, maxent, AUC

INTRODUCTION

Climate change is an increasing concern worldwide; issues such as crop failure and food security are progressively becoming more worrisome. The fruit pulp, leaves and seeds of the several genera in the family Sapindaceae are a good source of vitamins and minerals (Metcalfe and Chalk, 1957; Chapman, 1984) and are common ingredients in traditional herbal remedies in rural and urban Africa (Odugbemi and Akinsulire, 2006). Apart from food, they also serve as shade to local people (Watson and Dallwitz, 2008). Owing to the combination of health claims (e.g. anti-inflammatory (Sofidiya *et al.*, 2007), anti-fungal (Adesegun *et al.*, 2008), anti-diarrhoeal, (Antwi *et al.*, 2009), aphrodisiac (Muanya and Odukoya, 2008), pre-biotic and antioxidant functions (Ripa *et al.*, 2010), for eye infections (Pendota *et al.*, 2008), high calcium content) and food technological functions (high pectin and fiber content), fruit pulp of several genera has been identified as a suitable candidate for a new generation of functional foods and drinks (Chapman, 1984). Some yield valuable most widely used is Tulipwood (*Harpullia pendula*). Many other members of the Sapindaceae are important nuts, or sources of beverages, or oils. Saponins are present in the fruits, seeds, barks, twigs, leaves, pericarp and other tissues of several species; the bark of *Jagera pseudorhus* Blume contains large amounts and is used as a foaming agent and as a fish poison. Some types such as *Sapindus saponaria* are used as soap substitutes in the tropics. The fruit of *Dodonaea viscosa* (hop bush) has been used as a substitute for hops in the brewing of beer.

Also, due to formation of residential, industrial and other layouts, the natural slopes, vegetation and land cover gets disturbed however, species composition can be predicted by local environmental conditions (Ali and Kauser, 2006). Although members of Sapindaceae are not commonly cultivated, they can be successfully propagated. Investigating potential sites for cultivation is therefore an important first step towards the conservation of the representative species. In this study, geo-referenced Sapindaceae locality data and spatial environmental data were used in MAXENT, the species distribution modelling framework to predict potential sites for cultivation in Western Africa.

Corresponding Author: T.O. Adeyemi, Molecular Systematics Laboratory, Department of Botany, University of Lagos, Akoka, Lagos, Nigeria

MATERIALS AND METHODS

Species data: A total of 153 Sapindaceae growing localities (without duplicates) were assembled from diverse sources. About 32% of the localities came from recent fieldwork while 68% were herbarium records. In order to reduce the effect of temporal bias and to match the species data with the environmental climatic data (produced over the period 1950-2000), herbarium records collected before 1950 were eliminated.

Environmental data: Growing potential was assessed by characterising the climate at the presence locations. Twenty spatial datasets from the WorldClim database (at 5' resolution) (Hijmans *et al.*, 2005) were used. The Worldclim dataset included altitude and 19 bioclimatic variables derived from temperature and rainfall. Although 21 spatial datasets were used in the beginning, some were eliminated during modelling due to their low contribution to the final model. In agricultural systems, abiotic factors are more limiting than biotic factors (such as competition or dispersal); therefore no biotic factors were included in this study. Also, it was assumed that at a continental scale, climatic variables were much more limiting than human population.

Species distribution modelling: For each plant sample, GPS coordinates of the collection sites were recorded and subjected to modelling using the Maximum Entropy Programme (MAXENT) (Phillips and Dudik, 2008). Maxent predicts the potential species distribution by estimating the probability distribution of maximum entropy across a specified region, subject to a set of constraints that represent the missing information (lack of absence data) about the target distribution (Phillips *et al.*, 2006). The Maxent method is currently considered to be the most accurate approach to modelling presence-only data (Elith *et al.*, 2006; Pearson *et al.*, 2007).

Model validation: Model performance was evaluated using several methods. First, model performance was determined by assigning a subset of the presence records for training and using the remaining records to test the resulting model. A good model should predict correctly the presence of Sapindaceae in the test locations. As performance can vary depending upon the particular set of data selected for building the model and for testing it, 10 random partitions of the presence records were made to assess the average behaviour of Maxent, following Phillips *et al.* (2006). Each partition was created by

randomly selecting 75% of the total presence records to build the model and the remaining 25% of presence records were used for testing. However, the full set of presence records were used to build the final model to obtain the best estimate of the species distribution (for all records model, West Africa model). Second, Receiver Operating Characteristic (ROC) analysis was used to evaluate how well the Maxent model compared to a random prediction. The area under the ROC curve (AUC) serves as a measure of model performance in terms of sensitivity versus specificity. The sensitivity for a particular threshold is the fraction of all positive instances that are classified as present and specificity is the fraction of all negative instances that are classified as not present. The value of the AUC is typically between 0.5 (random) and 1.0. The closer the AUC value to 1, the better the model performance. Moreover, the success of the model was also evaluated by visually examining how well the mapped probability values matched the presence records. A good model should produce regions of high probability that cover the majority of presence records and areas of low probability should contain few or no presence points.

RESULTS

Potential distribution of Sapindaceae in West Africa: All records model, the resulting Maxent cultivation potential map for Africa predicted members of Sapindaceae to occur in most of West Africa. For the West Africa model, the cultivation potential map predicted Sapindaceae to grow in most of the Sahel from Senegal to south-west Cameroon and in southern Ghana-Nigeria (Fig. 1).

Model performance: All the ten generated training/testing models showed a high level of performance when compared to random (where the AUC would be 0.5). Training gained in this model ranged from 2.139 - 2.668 (Fig. 2) while the AUC value was 0.974 (Fig. 3). Test AUC values was lower but close to training AUC i.e., 0.960 (Fig. 3) while the test gained was 2.117 (Fig. 4). The training/testing models correctly predicted most of the test locations in all models and standard deviation is 0.007.

Variable contribution: The model indicated that the presence of members of Sapindaceae was mainly associated with minimum temperature of the coldest month (35.4%), precipitation of the coldest quarter (16.9%) and precipitation of wettest month (14.9%) (Table 1). Maxent's jackknife test of variable importance also
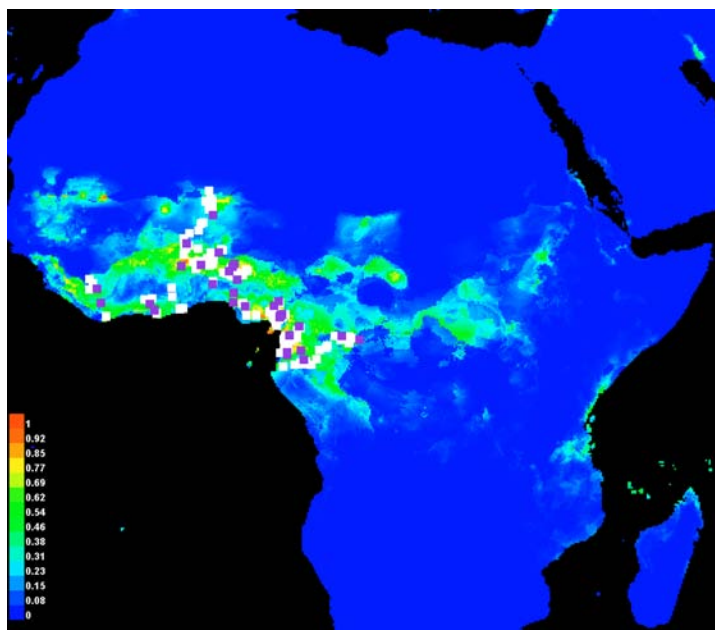
Fig. 1: Species distribution: Training vs. test locations, Green: The typical condition of where species are found, Blue: Low probability of suitable condition, Red: High probability of suitable condition, White dots: The presence locations used for training, while violet dots show test locations
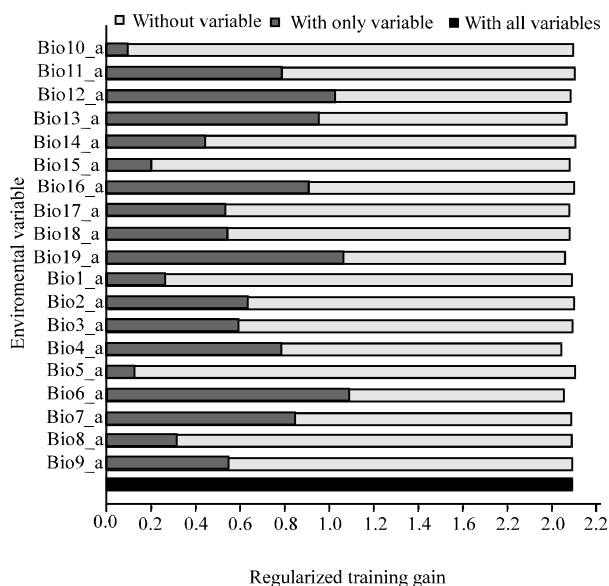

Fig. 2: Jack-knife of regularized training gain for Sapindaceae

Fig. 3: Sensitivity vs. specificity test on Sapindaceae

suggested that the variables which contributed the most to the model were good predictors as they had the most

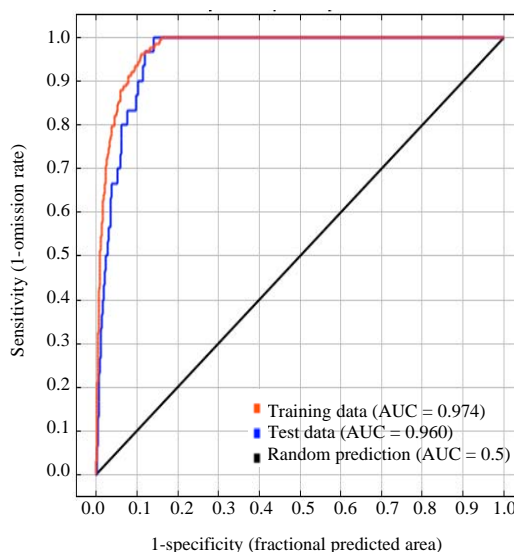information not contained in other variables and they could describe the distribution of Sapindaceae on their own (without the other variables).

Table 1: Variable contributions to the MAXENT model

| Variable	Contribution (%)
Annual mean temperature	2.3
Mean diurnal range (mean of monthly max.-min. temp.)	1.7
Isothermality (P2/P7) (* 100)	0.6
Temperature seasonality (standard deviation *100)	4.3
Max temperature of warmest month	1.1
Min temperature of coldest month	35.4
Temperature annual range (P5-P6)	4.6
Mean temperature of wettest quarter	1.9
Mean temperature of driest quarter	0.2
Mean temperature of warmest quarter	2.3
Mean temperature of coldest quarter	2.3
Annual precipitation	3.6
Precipitation of wettest month	14.9
Precipitation of driest month	0.3
Precipitation seasonality (coefficient of variation)	1.7
Precipitation of wettest quarter	2.3
Precipitation of driest quarter	2.6
Precipitation of warmest quarter	1.2
Precipitation of coldest quarter	16.9
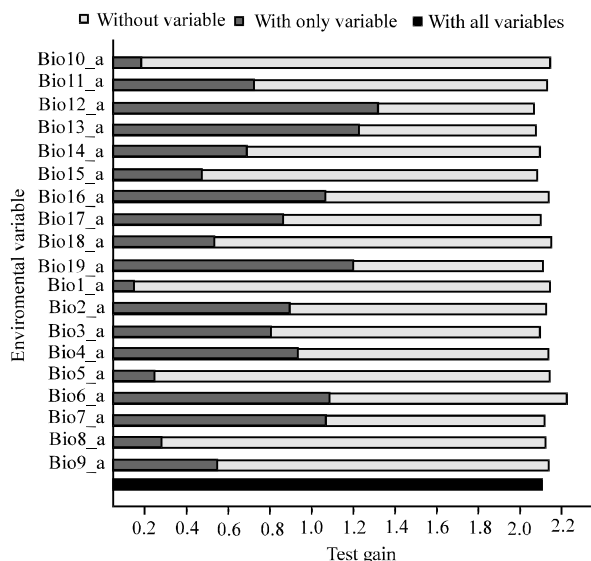
Fig. 4: Jack-knife of regularized test gain for Sapindaceae

DISCUSSION

Although members of the family Sapindaceae have been recorded to be widely distributed in Africa, their occurrence is being threatened by high rate of deforestation and agricultural practices leading to loss of forest and threatened status of the family as recorded in the IUCN R.L. (2008). MAXENT, a machine-learning method based on the principle of maximum entropy, are used to predict distribution for each species of butterfly under current and future climates (Sofy *et al.*, 2012). Data from this model is used to test the efficiency of the modeling program through the evaluation of the area under the Receiving Operator Curve (AUC) (Elith *et al.*, 2006; Phillips *et al.*, 2006).

Species distribution modelling with climatological data analysis using nineteen (19) standard environmental variables from the WorldClim database revealed that the species distribution is most affected by the minimum temperature of the coldest month (bio 6) with a variable contribution of 35.4%. This is followed by precipitation of coldest quarter (bio 19) and precipitation of the wettest month (bio 13) with 16.9 and 14.9% contribution, respectively. These results support the location of Sapindaceae mostly in the tropics rather than the temperate region.

Also, from the AUC curve generated, it can be concluded that this present model is more reliable as it predicts the species distribution better than a typical random model as shown in Fig. 3. This makes it useful for future projections. Recommendations for different conservation strategies include *in situ* conservation in Protected Areas; *ex situ* conservation in seed banks; and conservation through 'sustainable utilization'.

ACKNOWLEDGMENTS

REFERENCES

Adesegun, A.S., A.B.H. Coker and T.M. Hamann, 2008. Antifungal triterpenoid saponins from *Lecaniodiscus cupanioides*. Res. J. Phytochem., 2: 93-99.

Ali, S.M. and S. Kauser, 2006. Plant communities analysis of selected urban flora of Islamabad. J. Applied Sci., 6: 177-182.

Antwi, S., O.N.K. Martey, K. Donkor and L.K.N.A. Okine, 2009. Anti-diarrhoeal activity of *Blighia sapida* (Sapindaceae) in rats and mice. J. Pharmacol. Toxicol., 4: 117-125.

Chapman, K.R., 1984. Lychee-*Litchi chinensis* Sonn. In: Tropical Tree Fruits for Australia, Page, P.E. (Ed.). Queensland Department of Primary Industries, Brisbane, Australia, pp: 179-191.

Elith, J., C.H. Graham, R.P. Anderson, M. Dudik and S. Ferrier *et al.*, 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29: 129-151.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol., 25: 1965-1978.

Metcalfe, C.R. and L. Chalk, 1957. Anatomy of Dicotyledons. Vol. 1, Clarendon Press, Oxford, UK., Pages: 725.

Muanya, C.A. and O.A. Odukoya, 2008. Lipid peroxidation as index of activity in aphrodisiac herbs. J. Plant Sci., 3: 92-98.

Odugbemi, T. and O. Akinsulire, 2006. Medicinal Plants According to Family Names. In: Outlines and Pictures of Medicinal Plants from Nigeria, Odugbemi, T. (Ed.). University of Lagos Press, Akoka, Yaba, Nigeria, pp: 117-161.

Pearson, R.G., C.J. Raxworthy, M. Nakamura and A.T. Peterson, 2007. Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. J. Biogeogr., 34: 102-117.

Pendota, S.C., D.S. Grierson and A.J. Afolayan, 2008. An ethnobotanical study of plants used for the treatment of eye infections in the Eastern Cape province, South Africa. Pak. J. Biol. Sci., 11: 2051-2053.

Phillips, S.J. and M. Dudik, 2008. Modelling of species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography, 31: 161-175.

Phillips, S.J., R.P. Anderson and R.E. Schapire, 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell., 190: 231-259.

Ripa, F.A., M. Haque and I.J. Bulbul, 2010. *In vitro* antibacterial, cytotoxic and antioxidant activities of plant *Nephelium longan*. Pak. J. Biol. Sciences, 13: 22-27.

Sofidiya, M.O., O.A. Odukoya, A.J. Afolayan and O.B. Familoni, 2007. Survey of anti-inflammatory plants sold on herb markets in Lagos Nigeria. Int. J. Bot., 3: 302-306.

Sofy, A.R., A.A. Mousa, A.M. Soliman and K.A. El-Dougdoug, 2012. The Limiting of climatic factors and predicting of suitable habitat for citrus gummy bark disease occurrence using GIS. Int. J. Virol., 8: 165-177.

Watson, L. and M.J. Dallwitz, 2008. The grass genera of the world: Descriptions, illustrations, identification and information retrieval; including synonyms, morphology, anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution and references. The Grass Genera of the World.