Using GIS-Based Tools and Distribution Modeling to Determine Sweetpotato Germplasm Exploration and Documentation Priorities in Sub-Saharan Africa

Arthur Villordon¹

LSU AgCenter, Sweet Potato Research Station, 130 Sweet Potato Road, Chase, LA 71324

Wambui Njuguna

Department of Horticulture, Oregon State University, Corvallis, OR

Simon Gichuki

Kenya Agricultural Research Institute, Biotechnology Centre, P.O. Box 57811, Nairobi

Philip Ndolo

KARI-Kakamega Research Station, Kakamega, Kenya

Heneriko Kulembeka and Simon C. Jeremiah

MAF Department of Research and Development, ARI-Ukiriguru, P.O. Box 1433, Mwanza, Tanzania

Don LaBonte

LSU AgCenter, Department of Horticulture, 137 JC Miller Hall, Baton Rouge, LA 70803

Bernard Yada and Phinehas Tukamuhabwa

Department of Crop Science, Makerere University, Kampala, Uganda

Robert O. M. Mwangi

National Agricultural Research Organization, Namulonge Agric. & Animal Production Res. Institute, Box 7084 Kampala, Uganda

Additional index words. *Ipomoea batatas*, geographic information systems, model comparison

Abstract. Detailed information on the geographic distribution of a crop is important in planning efficient germplasm conservation strategies but is often not available, particularly for minor crops. Using germplasm collection data from Kenya, Tanzania, and Uganda, we used distribution modeling to predict the distribution of sweetpotato [*Ipomoea batatas* L. (Lam.)] in sub-Saharan Africa. We used a consensus modeling approach using the following algorithms: genetic algorithm for rule set prediction (GARP), maximum entropy, BIOCLIM, and DOMAIN. The predicted distribution encompasses known sweetpotato production areas as well as additional areas suited for this crop species. New geographic areas where at least three models predicted presence were in Angola, Cameroon, Central African Republic, The Congo, Democratic Republic of Congo, Gabon, Ghana, Angola, Ethiopia, Mozambique, Rwanda, and the Central African Republic. This information can be used to fill gaps in current sweetpotato germplasm collections as well as to further enhance the current presence-only based distribution model. Our approach demonstrates the usefulness of considering several models in developing distribution maps.

East Africa is widely considered as one of the secondary centers of diversity for sweetpotato, *Ipomoea batatas* L. (Lam.) (Austin,

1988; Gichuki et al., 2000; Huamán and Zhang, 1997; Yen, 1982). Farmers in the region typically grow landraces, and hybrid-

ization (Aldrich, 1963; MacDonald, 1965) and mutation (Aldrich, 1963) occasionally give rise to new varieties. This diversity is a potential source of desirable germplasm for distribution in the region as well as for use in breeding programs. However, regional conflicts, natural disasters, and the persistent threat of insect pests and disease endanger this genetic resource. This underscores the need to document the germplasm diversity in the region and to promote exploration, conservation, and improvement activities. An obstacle to efficient germplasm collection is that the entire distribution of a crop is often unknown. This can lead to the presence of major unnoticed gaps in a germplasm collection. It might strongly affect the genetic diversity captured, because many rare and useful traits may occur at the fringes of a species distribution.

Modeling of a species' distribution now represents an important tool in biogeography, evolution, ecology, conservation, and invasive-species management (Anderson et al., 2003; Williams, 2005). These methods use environmental data associated with the known distribution of a given species to infer its requirements and predict its entire geographic distribution using spatially continuous environmental surfaces. Environmental data used is mostly climatic, but may also include layers characterizing geomorphology or vegetation (Parra et al., 2004). Assumptions associated with these models as well as the influence of sample size are discussed elsewhere (Guisan and Thuiller, 2005; Guisan and Zimmermann, 2000; Hirzel et al., 2001; Stockwell and Peterson, 2001). Distribution modeling has not been used much in plant genetic resources conservation or germplasm management activities. A few examples are reported for *Arachis* (Jarvis et al., 2003) and *Anthurium* (Vargas et al., 2004). We explore the application of distribution models in predicting the distribution of a cultivated species: sweetpotato in sub-Saharan Africa.

The main objective of this study was to explore the feasibility of using a consensus GIS-based distribution model to identify gaps in sweetpotato collections for sub-Saharan Africa. Additionally, we also compared the predicted distributions derived from four distribution models and assessed the influence of sample stratification and variable selection on the derived models.

Materials and Methods

Modeling algorithms and software tools.

We used the following algorithms and software tools: genetic algorithm for rule set prediction (GARP, version 1.1; <http://beta.>

Received for publication 24 Apr. 2006. Accepted for publication 23 May 2006. Approved for publication by the director of the Louisiana Agricultural Experiment Station as manuscript no. 06-66-0205. Mention of a trademark, proprietary product, or

vendor does not imply endorsement by the Louisiana State University Agricultural Center nor its approval to the exclusion of other suitable products or vendors. We thank Robert Hijmans for providing valuable suggestions and for reviewing an

early draft of the manuscript. We thank McKnight Foundation for providing funds for the conduct of this research.

¹To whom reprint requests should be addressed; e-mail avillordon@agcenter.lsu.edu.

lifemapper.org/desktopgarp; Anderson et al., 2003; Stockwell and Peterson, 1999), maximum entropy (MaxEnt, version 2.0; <http://www.research.att.com/~phillips/density.html>; Phillips et al., 2004, 2005), BIOCLIM (Busby, 1986; Nix, 1986), and DOMAIN (Carpenter et al., 1993). We used BIOCLIM and DOMAIN as implemented in the DIVA-GIS software (version 5.2; Hijmans et al., 2001, 2002; www.diva-gis.org). These models were selected because they were appropriate for presence only data, represent a variety of statistical approaches, and are available for public download from their respective web sites.

Climate and elevation files. Global climate (WORLDCLIM version 1.3; Hijmans et al., 2005) and elevation grid layers were downloaded from <http://bioge.berkeley.edu/worldclim/worldclim.htm>. These layers had a resolution of 2.5 min (≈ 4 km) and included 19 bioclimatic variables (Table 1). The climate data spanned the years from 1950 to 2000. ArcView GIS (version 3.3, ESRI, Redlands, Calif.) and the Spatial Analyst extension (version 3.2, ESRI) were used to open and display these files. Using the ArcView script (GarpDatasets.avx) that came with the GARP installation, all of the files were clipped to match the study area (African continent) and exported as ASCII-formatted grids.

Sample data. The modeling experiments and evaluations were performed following this approach (Fielding and Bell, 1997): 1) presence data were split into training and testing sets, 2) the training set was used to build models, 3) the testing set was used to perform quantitative evaluation, and 4) pooled presence data were used to build a final model. Using this approach, sweetpotato germplasm collection data from Kenya, Uganda, and Tanzania (the “sample locality”) were compiled in a spreadsheet file and georeferenced. Assignment of longitude and latitude coordinates was performed manually or through the coordinate assignment utility in DIVA-GIS. Gazetteers for Kenya, Uganda, and Tanzania were downloaded from www.diva-gis.org (last accessed Feb. 2006). Georeferenced data were converted into comma-separated value (CSV) files and then transformed into point shapefiles using the DIVA-GIS import utility. The collection sites were then displayed along with country- and continent-level boundary data and visually examined to detect and remove outliers and other inaccurate locations (Fig. 1). This dataset (97 geographically unique records) was subsequently labeled the total data set (TOT). To generate the training (TRAIN) and testing (TEST) data, the modeling (modeling $\rightarrow$ evaluation $\rightarrow$ sample points) option in DIVA-GIS was used. We performed three repetitions (repetition = replication) and used the default 75% of the TOT to generate the TRAIN. The TEST was generated using DIVA-GIS’ “random absence data” feature with the following options: “random from the input file area” (bounding box) and the number of pseudoabsence points equal to the

Table 1. BIOCLIM variables used in the modeling experiments.²

BIO1 = Annual mean temperature
BIO2 = Mean diurnal range (mean of mo.ly (max temp – min temp))
BIO3 = Isothermality (P2/P7) (* 100)
BIO4 = Temperature seasonality (standard deviation *100)
BIO5 = Maximum temperature of warmest month
BIO6 = Minimum temperature of coldest month
BIO7 = Temperature annual range (P5–P6)
BIO8 = Mean temperature of wettest quarter
BIO9 = Mean temperature of driest quarter
BIO10 = Mean temperature of warmest quarter
BIO11 = Mean temperature of coldest quarter
BIO12 = Annual precipitation
BIO13 = Precipitation of wettest month
BIO14 = Precipitation of driest month
BIO15 = Precipitation seasonality (coefficient of variation)
BIO16 = Precipitation of wettest quarter
BIO17 = Precipitation of driest quarter
BIO18 = Precipitation of warmest quarter
BIO19 = Precipitation of coldest quarter

²WORLDCLIM Version 1.3 climate layers were downloaded from <http://bioge.berkeley.edu/worldclim/worldclim.htm>. (accessed Dec. 2005).

Fig. 1. Map of region showing location of sweetpotato germplasm distribution data used to build the models. Highlighted areas represent concentrations of occurrence data used for generating stratified samples.

presence points. We exported (modeling $\rightarrow$ external models $\rightarrow$ export $\rightarrow$ Maxent/GARP format) the TRAIN to a csv file for performing the GARP and Maxent modeling experiments.

To assess the effects of data partitioning and the relative importance of each environmental layer, we performed modeling experiments using partitioned occurrence data as well as reduced climate layers. Data partitioning was accomplished by splitting the TOT into five partitions, each partition representing concentrations of *I. batatas* points from the sample locality (Fig. 1). Each partition was subjected to the “sample points” procedure generating five subsamples (two samples each from Kenya and Uganda; one from Tanzania) that were aggregated to generate the “stratified” sample. Environmental layer selection was performed using two procedures: the variable selection procedure in MaxEnt and the removal of redundant variables through correlation analysis. The MaxEnt selection procedure consisted of jackknifing the variables and

comparing the training gain associated with each variable to the overall model derived from all variables. The following variables were selected through this procedure: temperature annual range (TAR), precipitation of wettest month (PWM), precipitation of driest month (PDM), precipitation of wettest quarter (PWQ), precipitation of warmest quarter (PHQ), and precipitation of coldest quarter (PCQ). We identified correlated variables by randomly sampling 500 points across locations in Uganda, Tanzania, and Kenya and subjected the sampled points to correlation analysis in SAS. We selected the following variables using this procedure: annual mean temperature (AMT), isothermality (ISO), temperature seasonality (TSO), temperature annual range (TAR), annual precipitation (APR), and precipitation seasonality (PRS).

Modeling parameters. The default settings were used for performing modeling experiments with each software application. The settings for GARP (version 1.3.3) were the following: number of iterations (1000),

convergence factor (0.01), and number of tasks (20). We also turned on the “best subset” setting and stacked the resulting prediction layers to display the results. The

default settings also used for the DIVA-GIS (version 5.2.0.2) implementations for the BIOCLIM and DOMAIN algorithms. The following were the default settings for Max-

Ent (version 2): convergence threshold (1.0E-5), maximum iterations (500), regularization multiplier (1), output format (cumulative), and output file type (.asc).

Model evaluation. DIVA-GIS was used to assess model accuracy for all four models. The TEST was used to generate evaluation files and draw ROC curves (Fielding and Bell, 1997) and determine the area under the curve (AUC) statistic. AUC is a threshold independent statistic. A value of 0.5 indicates that the model is no better than random, whereas an AUC of one indicates that it perfectly distinguishes between presence and absence. The AUC values were subjected to analysis of variance using SAS version 8.2 (SAS Institute, Cary, N.C.). The use of AUC to assess the accuracy of species distribution models has been reviewed by Fielding and Bell (1997).

Table 2. Effect of sample stratification and variable selection on area under the curve (AUC) values of four niche-based species distribution models.^z

Sampling ^x Environmental Layers ^w	Mean AUC ^y							
	DOMAIN		MAXENT		BIOCLIM		GARP	
	S	NS	S	NS	S	NS	S	NS
Full	0.99 a	0.95 bc	0.95 bcd	0.95 cd	0.94 cd	0.89 hg	0.90 fhg	0.91 efg
Reduced-VS	0.99 a	0.95 bc	0.94 ed	0.93 ed	0.93 ed	0.88 hi	0.81 m	0.82 lm
Reduced-CORR	0.99 a	0.97 ab	0.86 jk	0.86 jk	0.87 ij	0.84 kl	0.78 n	0.75 o

^zThe following niche-based species distribution models were used: BIOCLIM, DOMAIN, GARP, and MAXENT.

^yMean of three replicates. Mean separation within columns and rows by LSD at $P \leq 0.01$.

^xS = stratified; NS = nonstratified.

^wFull = BIOCLIM variables; Reduced-VS, variable selected by the jackknife-based procedure in MAXENT; Reduced-CORR, variable selected through a correlation analysis performed on 500 BIOCLIM points from Uganda, Kenya, and Tanzania.

Fig. 2. Potential distribution maps of *Ipomoea batatas* germplasm in sub-Saharan Africa using four niche-based modeling algorithms: A) DOMAIN, B) MAXENT, C) GARP, and D) BIOCLIM. Legend values for A, B, and D represent default values when displayed using DIVA-GIS. Legend values for C were modified to show intervals between stacked layers. Number of samples $n = 97$.

Consensus model. The TOT was used to generate distribution models with each algorithm. Threshold kappa values (TKV) were estimated using the DIVA-GIS model evaluation procedure. The kappa statistic is a chance corrected assessment of model performance (Monserud and Leemans, 1992). To compute kappa, the model results have to be classified as true (presence) or false (absence); by doing this for several thresholds, TKV, the threshold at which kappa is highest, is found. The TKV value was used to reclassifying the model output to 0 (predicted absence) and 1 (predicted presence), i.e., values below the TKV were reclassified as "0," whereas those above it were reclassified as "1". The reclassified distribution maps were added to generate a consensus predicted species distribution map with values that ranged from 0 to 4 (0 = no model predicts presence, 4 = concurrence among all models for predicting presence).

Results and Discussion

Accuracy of individual distribution models. We found that sample stratification generally increased the accuracy (higher AUC values) of BIOCLIM and DOMAIN algorithms across environmental layers (full and reduced-VS), but did not influence the results of the GARP and MAXENT-based models (Table 2). Sample stratification minimizes sampling bias by ensuring that regions with concentrations of sample points are not overrepresented in the training data set. The sensitivity of the DOMAIN model to collection bias has also been observed previously (Vargas et al., 2004). The inclusion or exclusion of environmental layers did not influence the AUC values for DOMAIN and BIOCLIM models. This response may be in

part the result of the fact that these models assign equal weights to each variable and generally do not address potential interactions among environmental layers (Farber and Kadmon, 2003). On the other hand, GARP and MAXENT responded to the environmental selection procedure with higher AUC values when all environmental layers were included in the model. The Maxent-based variable selection procedure (jackknife-based) used in this experiment identified several rainfall-related variables, suggesting the important role of moisture in defining areas that are suitable for species presence.

The accuracy of species distribution models vary according to species, sample number, environmental variables, and related factors (Segurado and Araujo, 2004). Using presence data from a limited geographic area, the GARP-based model (stratified sample/complete environmental layers) predicted the largest TKV-defined mean geographic area (2,216,950 km²) but showed the lowest AUC values across environmental layer and sampling categories. Commission errors or overprediction, i.e., prediction of occurrence in areas where the species is absent, generally decreases AUC values. Stockman et al. (2006) have demonstrated that GARP can exhibit a large degree of overprediction. On the other hand, commission errors may also be the result of areas that are potentially habitable regions but lack occurrence data resulting from biotic, historical, and other reasons (Anderson et al., 2003). In contrast, the DOMAIN model had the most restrictive mean TKV-defined area (244,032 km²) that excluded several locations that the other three algorithms predicted as occurrence sites. The mean TKV-defined area for BIOCLIM and MAXENT were 1,579,493 km² and 1,704,375 km², respectively. Although both algorithms appeared to reach consensus for

certain geographic locations, BIOCLIM predicted presence in certain sites that were excluded by MAXENT (Fig. 2). These results indicate that a distribution map based on a single model (i.e., BIOCLIM, DOMAIN, MAXENT) will exclude sites that represent potential species presence. On the other hand, a GARP-based distribution model will potentially include sites that represent species absence, resulting in inefficient use of resources for germplasm exploration and documentation.

Consensus distribution model. There is general agreement that no single modeling algorithm is better than the other (Guisan and Zimmermann, 2000; Segurado and Araujo, 2004). We used an approach similar to what Burgman et al. (2005) describes as a strategy that does not focus on finding a true prediction, but rather explores models and modeling frames that lead to the consideration of robust, satisfactory alternatives. In our case, we combined several algorithms to identify a central tendency (Gelfand et al., 2005; Thuiller 2004) to develop a consensus potential geographic distribution model (Fig. 3A). Outside of the sample locality (Uganda, Kenya, and Tanzania), the various algorithms achieved virtual consensus (score of 3–4) in predicting the likely geographic presence of *I. batatas* in various locations of the study area: south central Ethiopia and the region bounded by Cameroon, the Congo, and Gabon. Isolated pockets of potential distribution were also predicted in Angola, Ghana, Democratic Republic of Congo, Mozambique, Central African Republic, and Rwanda. These regions generally encompass areas of documented occurrences of *I. batatas*. Huaccho and Hijmans (2000) compiled information from various sources and mapped the global distribution of sweetpotatoes. The georeferenced location of sweetpotatoes for

Fig. 3. (A) Consensus potential species distribution map for *Ipomoea batatas* in sub-Saharan Africa. This map was generated by stacking the reclassified distribution models generated by BIOCLIM, DOMAIN, GARP, and MAXENT. (B) Georeferenced known locations of sweetpotato production areas in Africa. Shapefile provided by Robert Hijmans based on Huaccho and Hijmans (2000).

Africa is shown in Fig. 3B. In general, the consensus predicted distribution map encompasses currently known production areas as documented by Huaccho and Hijmans (2000), related information from FAO (FAO-STAT, 2005), the System-Wide Information Network for Genetic Resources database (SINGER, 2005), and other reports of sweetpotato distribution in the region.

The information derived from the consensus prediction map can be used in a number of ways. First, it can assist in planning germplasm management activities by identifying potential priority areas (geographic locations with consensus score of 3–4) for germplasm exploration and documentation. Specifically, it can serve as a tool to verify that all known areas of distribution have been adequately explored (Williams et al., 2002) and identify gaps in the current record. For example, the SINGER database (SINGER, 2005) shows sweetpotato germplasm collection records for Cameroon, Kenya, Rwanda, and Uganda. Thus, other consensus sites, e.g., Mozambique, Ghana, and Democratic Republic of Congo, appear to be underrepresented. In addition, the consensus model predicts a high probability of occurrence for certain locations in the Central African Republic. However, there are no current FAO estimates of sweetpotato production for this country (CAF, Fig. 3B) or entries for it in the SINGER database. Future sweetpotato germplasm exploration and collection activities might focus on these putatively “underrepresented” regions. Such exploration and collection activities also offer an opportunity to improve the accuracy of the current model, validate the consensus distribution model by generating “absence data” (Margules and Austin, 1994), and further explore presence–absence-based modeling algorithms. The current model was developed using a geographically limited training data set. Additional data can enhance the current model’s accuracy. For example, the current consensus model predicts presence in a small area in Madagascar. However, it is reported that sweetpotatoes are present in other areas of the island (Randrianaivoarivony, personal communication, 2006). Moreover, the information can be used along with other data in developing models that predict the influence of climate change on species distribution (Guisan and Thuiller, 2005; Martinez-Meyer, 2005).

Conclusions

Four niche-based species distribution models were combined to develop a consensus predicted geographic distribution for sweetpotatoes in sub-Saharan Africa. The algorithms reached virtual consensus in predicting species presence for certain locations. The geographic distribution map identified areas that are currently underrepresented in publicly available germplasm collection and crop production databases. This information can be used by stakeholders in setting prior-

ities for future germplasm exploration and documentation activities.

Literature Cited

- Aldrich, D.T. 1963. The sweetpotato crop in Uganda. *East African Agriculture and Forestry Journal* 29:42–49.
- Anderson, R.P., D. Lew, and A.T. Peterson. 2003. Evaluating predictive models of species’ distributions: criteria for selecting optimal models. *Ecol. Modell.* 162:211–232.
- Austin, D.F. 1988. The taxonomy, evolution and genetic diversity of sweetpotatoes and related wild species. In: *Exploration, Maintenance, and Utilization of Sweetpotato Genetic Resources*. Report of the First Planning Conference 1987. International Potato Center, Lima, Peru. p. 27–59.
- Burgman, M.A., D.B. Lindenmayer, and J. Elith. 2005. Managing landscapes for conservation under uncertainty. *Ecology* 86:2007–2017.
- Busby, J.R. 1986. A biogeographical analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. *Aust. J. Ecol.* 11:1–7.
- Carpenter, G., A.N. Gillson, and J. Winter. 1993. DOMAIN: a flexible modeling procedure for mapping potential distributions of plants and animals. *Biodivers. Conserv.* 2:667–680.
- Farber, O. and R. Kadmon. 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecol. Modell.* 160:115–130.
- FAOSTAT. 2005. <http://faostat.fao.org/>. Accessed Feb. 2006.
- Fielding, A.H. and J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24:38–49.
- Gelfand, A.E., A.M. Schmidt, S. Wu, J.S. Silander, A. Latimer, and A.G. Rebelo. 2005. Modelling species diversity through species level hierarchical modelling. *J. R. Stat. Soc. C Appl. Stat.* 54:1–20.
- Gichuki, S.T., M. Berenyi, J. Schmidt, J. Glöbl, and K. Burg. 2000. Evaluation of genetic diversity of sweetpotato cultivars with RAPD and AFLP molecular markers. *African Potato Assoc. Conf. Proc.* 5:19–21.
- Guisan, A. and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8:993–1009.
- Guisan, A. and N.E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecol. Modell.* 135:147–186.
- 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.
- Hijmans, R.J., L. Guarino, C. Bussink, and E. Rojas. 2002. DIVA-GIS, version 2. A geographic information system for the analysis of biodiversity data. Manual. International Potato Center, Lima, Peru.
- Hijmans, R.J., L. Guarino, M. Cruz, and E. Rojas. 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genet. Resour. Newsl.* 127:15–19.
- Hirzel, A.H., V. Helfer, and F. Metral. 2001. Assessing habitat suitability models with a virtual species. *Ecol. Modell.* 145:111–121.
- Huaccho, L. and R.J. Hijmans. 2000. A Geo-Referenced Database of Global Sweetpotato Distribution. Production Systems and Natural Resource Management Department Working Paper No. 4. CIP, Lima, Peru.
- Huamán, Z. and D.P. Zhang. 1997. Sweetpotato. In: D. Fuccillo, L. Sears, and P. Stapleton (eds.), *Biodiversity in Trust*. Cambridge Univ. Press, Cambridge, U.K. p. 29–38.
- Jarvis, A., M.E. Ferguson, D. Williams, L. Guarino, P. Jones, H.T. Stalker, J.F.M. Valls, R. Pittman, C. Simpson, and P. Bramel. 2003. Biogeography of Wild *Arachis*. Assessing conservation status and setting future priorities. *Crop Sci.* 43:1100–1108.
- MacDonald, A.S. 1965. Variation in open pollinated sweetpotato seedlings in Buganda, East Africa. *E. Afr. Agr. J.* 31:183–188.
- Margules, C.R. and M.P. Austin. 1994. Biological models for monitoring species decline: the construction and use of databases. *Philos. Trans. R. Soc. London* 344:69–75.
- Martinez-Meyer, E. 2005. Climate change and biodiversity: some considerations in forecasting shifts in species’ potential distributions. *Biodiversity Informatics* 2:42–55.
- Monserud, R.A. and R. Leemans. 1992. Comparing global vegetation maps with the Kappa statistic. *Ecol. Modell.* 62:275–293.
- Nix, H. 1986. A biogeographical analysis of Australian elapid snakes. *Atlas of Elapid Snakes of Australia*. Australian Government Publishing Service, Canberra, Australia, p. 4–15.
- Parra, J.L., C.C. Graham, and J.F. Freile. 2004. Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography* 27:350–360.
- Phillips, J.S., R.P. Anderson, and R.E. Schapire. 2005. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190:231–259.
- Phillips, J.S., M. Dudik, and R.E. Schapire. 2004. A Maximum entropy approach to species distribution modeling. In: *Proc. 21st International Conference on Machine Learning*, Banff, ACM Press, N.Y. p. 655–662.
- Segurado, P. and M.B. Araujo. 2004. An evaluation of methods for modeling species distributions. *J. Biogeogr.* 31:1555–1568.
- SINGER. 2005. <http://singer.grinfo.net/>. Accessed Feb. 2006.
- Stockman, A.K., D.A. Beamer, and J.E. Bond. 2006. An evaluation of a GARP model as an approach to predicting the spatial distribution of non-vagile invertebrate species. *Diversity and Distributions* 12:81–89.
- Stockwell, D. and A. Peterson. 1999. The GARP modeling system: problems and solutions to automated spatial prediction. *Int. J. Geogr. Inf. Sci.* 13:143–158.
- Stockwell, D. and A. Peterson. 2001. Effect of sample size on accuracy of species distribution models. *Ecol. Modell.* 148:1–13.
- Thuiller, W. 2004. Patterns and uncertainties of species range shifts under climate change. *Glob. Change Biol.* 10:2020–2027.
- Vargas, J.H., T. Consiglio, P.M. Jorgensen, and T.B. Croat. 2004. Modelling distribution patterns in a species-rich plant genus, *Anthurium* (Araceae), in Ecuador. *Diversity Distrib.* 10:211–216.
- Williams, K.A. 2005. An overview of the U.S. National Plant Germplasm System’s Exploration Program. *HortScience* 40: 297–301.
- Williams, P.H., C.R. Margules, and D.W. Hilbert. 2002. Data requirements and data sources of biodiversity priority area selection. *J. Biosci.* 27:327–338.
- Yen, D.E. 1982. Sweetpotato in historical perspective. In: R.L. Villareal and T.D. Griggs (eds.), *Sweet Potato*. Proceedings of the First International Symposium, AVRDC Publ. No. 82-172. p. 17–30.