

P.G. Jones¹, L. Guarino², A. Jarvis³



UNIDAD DE INFORMACION Y
DOCUMENTACION

Abstract

FloraMap is a specialized GIS program which was developed to map the predicted distribution, or areas of possible climatic adaptation, of organisms in the wild. The climate at the locations where individuals of a given taxon were observed is assumed to be representative of the environmental range of the organism. The climate at these locations is extracted from a set of interpolated climate grids, and used as a calibration set to calculate a probability surface. The system can be used for a number of different plant genetic resources applications besides predicting species distributions. Some of these are described, along with future plans for development of the software.

225425

Introduction

Management of plant genetic resources (PGR) both generates and uses data. The analysis of these data is crucial to the effectiveness of the PGR management process, and can add significantly to the value of genetic resources. Many of these data are geo-referenced, i.e., corresponding to specific locations on the earth's surface, which means that they can be analyzed – and linkages made between them and other geo-referenced data from sources external to the PGR process – using Geographic Information Systems (GIS). A GIS is a database management system that can simultaneously handle data representing spatial objects and their attribute data. GIS can play an important role in the management of large and complex PGR datasets (Guarino et al., 2001). However, examples of the use of GIS to address PGR problems are still relatively few. Reasons for this probably include the relatively high cost and complexity of much GIS software and the perception that geo-referenced data can be difficult and expensive to obtain.

¹ Land Use Consultant, Centro Internacional de Agricultura Tropical (CIAT).

² Scientist, International Plant Genetic Resources Institute (IPGRI).

³ Visiting Researcher, International Plant Genetic Resources Institute (IPGRI).



UNIDAD DE INFORMACION Y
DOCUMENTACION

While increasing familiarity with GIS within the PGR conservation community, and advances in information technology, will no doubt eventually overcome these problems, PGR workers need the power associated with these methods now, not at some time in the future. It is with this in mind that CIAT, CIP and IPGRI have been working to develop easy-to-use GIS applications (including software and data) aimed at solving specific PGR problems. This paper introduces one of these, called FloraMap, and describes its possible usefulness to PGR conservation and use programmes. Another paper in this series describes a software tool called DIVA developed by CIP and IPGRI.

FloraMap (Jones and Gladkov, 1999) is a GIS application developed by CIAT (International Centre for Tropical Agriculture, Cali, Colombia) primarily for the prediction of the distribution of organisms in the wild, when little or nothing is known of the physiology of the species involved, so that analytical or simulation models cannot be used. This is important to PGR workers because data on the geographic distribution of species are often scanty, and it would clearly be useful to be able to use whatever little is known to identify areas where a species has not been previously recorded but where it might still be expected to occur. FloraMap assumes that the climate at the points of observation and/or collection of a species (herbarium specimens, germplasm accessions etc.) is representative of the environmental range of the organism. The climate at these points is used as a calibration set to compute a climate probability model, which is then used to assess the likelihood of other sites being climatically suitable for the species.

Below we briefly describe the methods used in FloraMap, the software interface, its possible uses and future developments. The methods used in FloraMap are described in detail by Jones et al. (1997), and also in the user manual (Jones and Gladkov, 1999).

How FloraMap works

The base climate data. FloraMap includes – and uses in its analyses – climatic data from a 10-minute grid (corresponding to 18km at the Equator). The grids were derived by

interpolation from thousands of meteorological stations. A simple algorithm based on the inverse square of the distance between the five nearest stations and the interpolated point is used. Thirty-six climatic variables are used: 12 monthly averages for temperature, rainfall, and diurnal temperature range. Temperature is standardised with elevation using the NOAA TGP-006 (NOAA, 1984) digital elevation model (DEM) and a lapse rate model (Jones, 1991). Rainfall and diurnal temperature range remain independent of elevation.

In order to adjust for geographic differences in the timing of major seasons, a 12-point Fourier transform is applied (for further information, see Jones et al., 1997). This is done to allow for a better comparison among climates. For example, the climate of a location in the northern hemisphere may be very similar to that of a location in the southern hemisphere, except for the irrelevant fact that the rainy season starts in a different month (e.g., July versus January). Six frequencies are fitted to the climate data. The phase angle of the first frequency for the rainfall and temperature records are combined as a function of latitude and the resulting angle is subtracted from each climate record. The Fourier transformation thus rotates the climate data to a common time frame to make them directly comparable.

FloraMap currently has climate grids for Latin America and the Caribbean, Africa, Asia, Europe and the coterminous USA. The European grid uses data courtesy of Dr Ian Makin IWMI and the USA grid data from climate grids developed by Dr Chris Daly of Oregon State University.

Computing the climate probability model. The input database file provided by the user should include an accession identifier, latitude and longitude coordinates in decimal degrees, and altitude if available. The 36 climate variables described above are first extracted for the pixel in which each accession in the input file is located, and principal components analysis (PCA) carried out on the resulting data. PCA identifies new variables or dimensions (principal components), related to the original variables, but which account for most of the variance in climates among the accession locations in the fewest number of dimensions. The PCA is performed on the variance-covariance matrix

since the Fourier analysis has already transformed the variables to comparable scales. FloraMap then fits a multivariate-Normal distribution to the principal component scores of the accessions (the user specific how many dimensions to include).

Mapping and exporting the probability surface. Finally, FloraMap uses the parameters of this multivariate-Normal distribution to calculate the probability of the climate at each pixel in selected continental coverage(s) belonging to the probability distribution defined by the calibration set, in the space of dimension determined by the number of principal components chosen by the user. The final result is a probability surface for the continent(s) under consideration. These probability surfaces are ESRI® Shapefiles, with each pixel defined as a square polygon. They can be readily exported to other GIS applications, a feature useful for carrying out further analyses and producing individually crafted output maps.

FloraMap analysis can thus be summarized as follows:

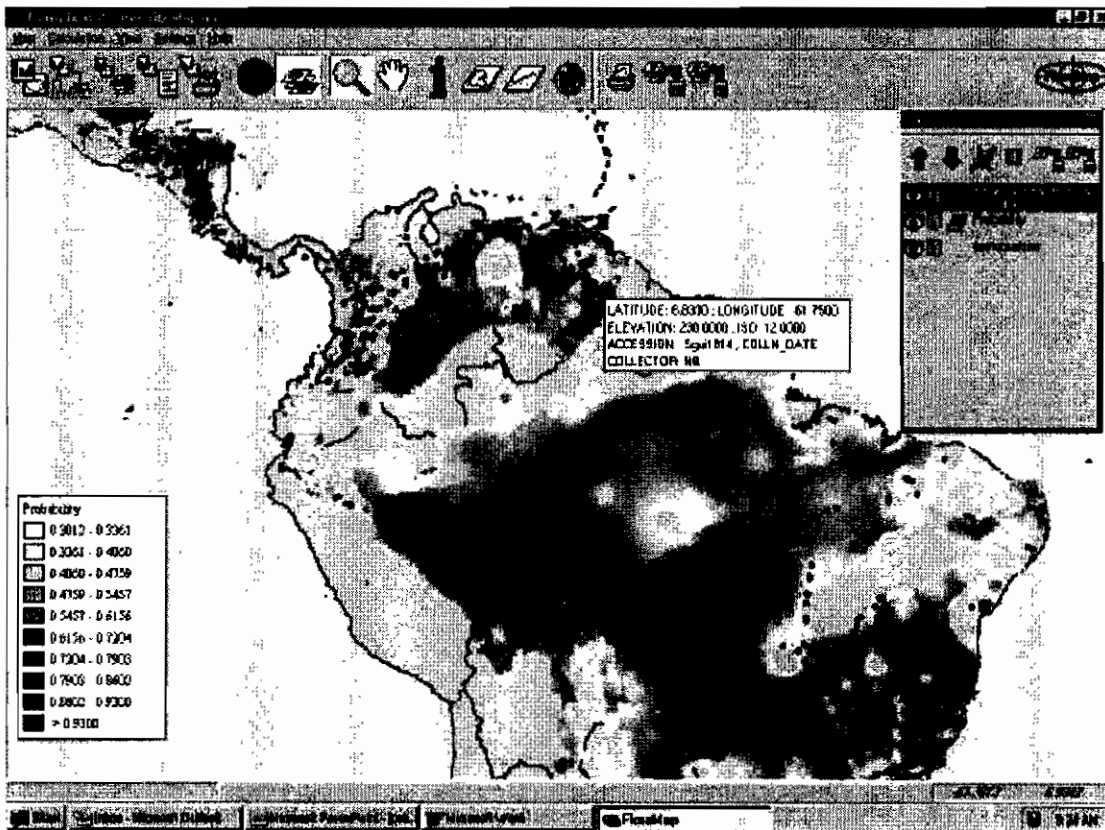
1. Input and checking of accession locality data
2. Extraction of data for 36 climate variables for each accession
3. PCA of the extracted climate data
4. Fitting of multivariate-Normal distribution to the resulting PCA scores
5. Calculation, for each pixel in the selected continent, of its probability of belonging to this multivariate-Normal distribution
6. Mapping of the probability surface for the selected continent

The process is managed by the user through a series of distinct but interrelated tools, as described in the next section.

The FloraMap user interface

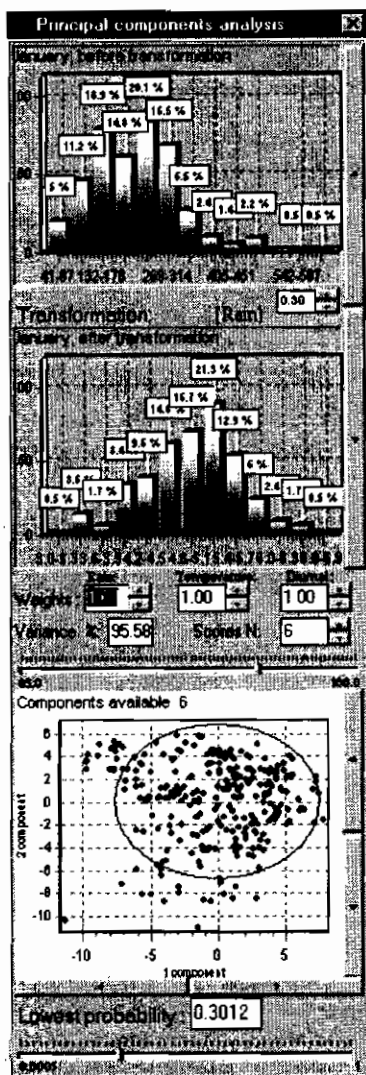
The FloraMap user interface was developed using C++ and ESRI's MapObject software. It is described below in terms of each of the tools that comprise it.

The Map Window Tool (Fig. 1). The map window is the central tool in FloraMap. It displays the map and has controls for managing the map layers, including probability surfaces and accession points used in the model. The political boundaries of Latin America, Africa and Asia can be displayed as background to the probability layers. Roads, rivers and cities can be overlaid as aids in geographically representing the study areas. Control of colours and characteristics of all layers is available through the Layer Control Window. There are facilities for saving and reloading complete maps, or individual map layers.

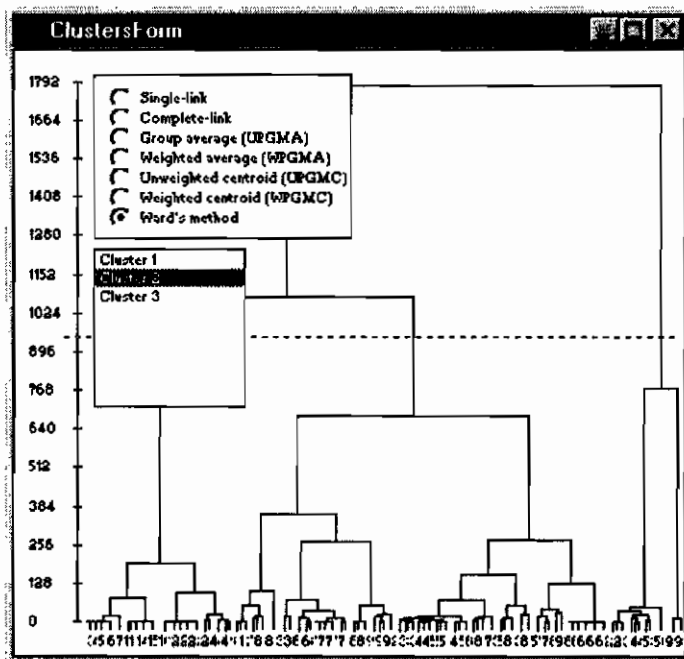
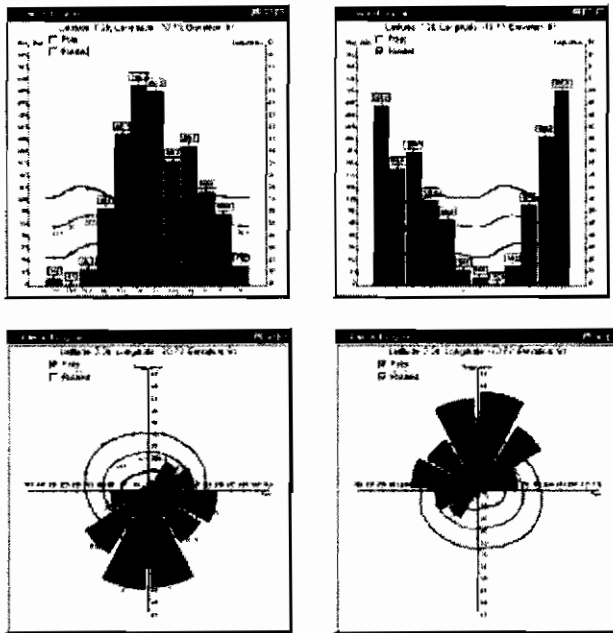


The Principal Components Window Tool (Fig. 2). This window controls the principal components analysis. The scaling, weighting and transformation of the variates, choice of components and map probability limits are all at the control of the user, using handy sliders or by entering the values desired. A scatter diagram shows the distribution of the accession points in any combination of the principal component dimensions. A powerful

tool for selecting sub-sets of the data works merely by drawing round the points in question. This also works from the Map Window.



The Cluster Analysis Tool (Fig. 3). An accession set may be found to contain sub-groups with significantly distinct climatic ranges. In such cases, fitting a model to the data as a whole may give spurious results. The Cluster Analysis Tool gives the user access to seven powerful clustering techniques to investigate the possibility of clustering. When clusters are found, the user can point and select a cluster and calculate the probability model for those accessions only, repeating for each cluster.



The Climate Diagram Tool (Fig. 4). The user can view the climate of a point from the map, an accession point from the scatter diagram or map and also the mean climate for groups of points from the scatter diagram or the map. The climate diagram can be displayed in Cartesian or polar co-ordinates and can show the data either by real date, or in rotated form.

The varied uses of FloraMap

The primary use of FloraMap is, as the subtitle of the user manual states, predicting the distribution of plants and other organisms in the wild from perhaps very limited distribution data. Comparing such predicted distribution with the localities where accessions were collected could, for example, be useful to germplasm collectors trying to identify gaps in existing collections. There have been few attempts to test the accuracy of the predicted distributions, but Jones et al. (1997a) did find some success with wild *Phaseolus* in South America.

However, prediction of natural distribution and gap identification are not the only use that PGR workers can make of FloraMap. The data that comes with it, and the software's analytical facilities, also allow other PGR problems to be tackled, mainly through the "retro-classification" of accessions. This is the term coined by Steiner and Greene (1996) to describe the process of obtaining data on the environment at a site of collection not in the field, but by overlaying the locations of the sites on different thematic base maps of the appropriate scale and reading off the map attributes, most conveniently by using GIS. Steiner and Greene (1996) give an example of the usefulness of retro-classification using the *Lotus* collection of the U.S. National Plant Germplasm System (NPGS) (see also Greene et al. 1999). The following applications may be highlighted:

- checking data quality
- predicting climatic adaptation in other areas
- identifying groups of accessions with distinct climatic adaptations (ecotypes)
- comparing climatic adaptation among *a priori* groups of accessions

Checking data quality. The PCA scatter diagram is useful in identifying possible errors in the data. Any accessions which are obvious outliers in the PCA diagram should be checked for any inaccuracies in the latitude and longitude information. FloraMap also flags accessions whose coordinates cause it to fall in a body of water. If required, it can move them to the nearest point on the coastline.

Predicting climatic adaptation in other areas. The model describing the climatic adaptation of the species need not be applied only to the putative region of its distribution in the wild. FloraMap allows the user to apply the model to other continents, for example. This means that the distribution of species which are promising as forages and which are native to Africa, say, can be used to predict where they might be expected to yield a good response on introduction to South America or Southeast Asia. Initial trial material could be targeted precisely to areas where they would be expected to be best adapted climatically. Jones et al. (2000) presents a study of cross-continental climatic adaptation in *Desmodium* and Libreros et al. (2000) a study of adaptation of South American fruit species in Mexico.

Identifying groups within collections. We have seen how the clustering tool allows the user to investigate the possible existence of groups of accessions with distinct climatic adaptations within the collection. When such groups are strongly defined, it is necessary to fit individual probability distributions separately to the accessions in each group, rather than to the collection as a whole, in order for the model to work correctly. However, the PGR worker may well be interested in the existence of such groupings quite apart from their importance in fitting the probability model, because such groups may be associated with genetic differences of significance for genetic resources conservation or use. In studies of *Stylosanthes* (Jones et al. 1997b) and *Passiflora* (Segura et al., 1998; Segura et al. 1999), for example, climatic clusters were found to differ with regard to specific isozymes. Climatic groupings could be used in developing core collections. Also, material with distinct climatic adaptation within the collection may need to be multiplied, characterized and evaluated at different sites.

Comparing climatic adaptations. Groupings may already have been defined within a collection, say on the basis of characterization or evaluation data. In such cases, it might be interesting to the PGR worker to compare the climatic adaptation of these *a priori* groups of accessions. For example, FloraMap was used in conjunction with data on a collection of wild *Arachis* species to define and compare the climatic adaptations of different species (Ferguson et al., in prep). Users interested in adaptation to particular

climatic conditions would now know which of the species are most suitable for their purposes.

What FloraMap does NOT do

Useful as it is proving to be for various applications, users should be fully aware of the limitations of FloraMap. The software uses a particular model to map the potential climatic “envelopes” where an organism could exist. First, it is worth reiterating that the “envelope” that is developed is purely a climatic one. FloraMap takes no account of factors such as soil and dispersal mechanism. Though climate is widely considered to be one of the primary determinants of plant distribution, it is clearly not the only one. However, the probability surfaces can be imported into other GIS, where they can be overlaid with other geo-referenced information, for example on soils, natural vegetation, human intervention, physical barriers etc., to refine the analysis.

Second, it should also be remembered that a different model could result in different predictions. Apart from FloraMap, other software also address the problem of predicting species distributions, though in different ways, and therefore potentially with different results. These include BIOCLIM¹ (Busby, 1991; see also GARP², Genetic Algorithm for Rule-set Production, an extension of the BIOCLIM approach) and DOMAIN³ (Carpenter et al., 1993). An instructive, large-scale example of the application of GIS to the problem of “objective prediction of the full distribution of a species from incomplete point distribution maps, based on its ecological preferences” is the work being carried out by the Royal Botanic Gardens, Kew⁴ in the context of Madagascar’s Environmental Action Plan.

¹ http://dino.wiz.uni-kassel.de/model_db/mdb/bioclim.html

² http://kaos.erin.gov.au/general/biodiv_model/ERIN/GARP/home.html

³ http://www.cgiar.org/cifor/research/intro_d.html

⁴ http://www.rbgekew.org.uk/herbarium/madagascar/plant_dis.html

Finally, FloraMap does not give a once-and-for-all answer. As more distribution data becomes available, the results of the analysis may well change significantly, in particular if the original accessions were in some way a climatically biased sample of the real distribution of the species. For more detailed examination of geographic biases in accessions datasets see Hijmans et al. (1999).

The FloraMap user community

The Users Group. FloraMap currently has a users group of over 100 individuals throughout the world. To join the group listserver, it is simply necessary to send a message to: listserv@cgiar.org with the following text in the body of the message: **subscribe FloraMap <your email address>**. The listserver is used to answer user questions, disseminate new information about FloraMap and as a platform for users world-wide to exchange experiences and hints on its use.

The Web Site. FloraMap has a Web site at <http://www.floramap-ciar.org> with information on the software, examples of its use, a growing list of Frequently Asked Questions (FAQ) with their answers, and extra climate grids not supplied with the original software package. The solution to your difficulties with FloraMap could be merely a few mouse clicks away.

The future of FloraMap

CIAT is committed to supporting FloraMap as a member of its stable of Climate Applications, and for the foreseeable future will be developing the product in response to user requests and as the algorithms are further developed. One of the present drawbacks to the system is the precision of the climate grids. At 10 minutes (18km) for Latin America, Africa and Europe, these are far from satisfactory in mountainous areas and

territory with broken relief. The aim is to provide grids based on the USGS 1km Digital Elevation Model ETOPO30 as soon as feasible.

This is quite a challenge in two respects. First the sheer effort of compiling the climate data and fitting and checking the grids is daunting. However, there is a much more important hurdle to clear. At present, FloraMap will take a considerable time, and may fail, on all but the largest, fastest, modern PCs when presented with the 2.5 arc minute grid for South East Asia. A 30 arc second (1km) grid contains 25 times the data and FloraMap will certainly not work for such a large grid at this precision. The answer to this is to restructure the data storage and retrieval for the climate grids. Work has started on this and the perhaps ambitious aim has been set of increasing the speed of FloraMap by 10,000 times while reducing storage space needed for the data. This is not going to be achieved quickly, but will hopefully be ready for the next release in 2002.

Other enhancements may include better geographic background files, automatic variable transforms, alternative climate models (the user may like to see what the analysis would look like in DOMAIN) and Boolean operations to combine multiple probability layers.

Bibliography

- Busby, J.R. 1991. BIOCLIM – a bioclimate prediction system. pp 4-68. In C.R. Margules, and M.P. Austin (eds.) Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne, Australia.
- Carpenter, G., A.N. Gillison and J. Winter. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation* 2:667-680.
- Greene, S.L., T. Hart and A. Afonin. 1999b. Using geographic information to acquire wild crop germplasm: II. Post collection analysis. *Crop Science* 39:843-849.
- Guarino, L., A. Jarvis, R.J. Hijmans and N. Maxted. 2001. Geographic Information Systems (GIS) and the conservation and use of plant genetic resources”

- International Conference on Science and Technology for Managing Plant Genetic Diversity in the 21st Century (SAT21). Kuala Lumpur, Malaysia. 12-16 June, 2000.
- Hijmans R.J., K.A. Garrett, Z. Huaman, D.P. Zhang, M. Schreuder, M. Bonierbale. 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology*. 14(6):1755-1765.
- Jones, P.G., and A. Gladkov. 1999. FloraMap Version 1. A computer tool for predicting the distribution of plants and other organisms in the wild. CD-ROM and Manual 99 p. Centro Internacional de Agricultura Tropical, Cali Colombia.
- Jones, P.G., N. Galwey, S.E. Beebe and J. Tohme. 1997a. The use of geographical information systems in biodiversity exploration and conservation. *Biodiversity and Conservation* 6:947-958.
- Jones, P.G., M.C. Sawkins, B.L. Maass and P.C. Kerridge. 1997b. GIS and genetic diversity case studies in *Stylosanthes*. Poster presented at XVIII International Grassland Congress, June 8-19, Winnipeg, Canada.
- Jones, P.G., S. Segura, L. Guarino and M. Peters. 2000. FloraMap: a computer tool for predicting the distribution of plants and other organisms in the wild. Poster presented at the International Conference on Science and Technology for Managing Plant Genetic Diversity in the 21st Century (SAT21). Kuala Lumpur, Malaysia. 12-16 June, 2000.
- Libreros, D.F., S. Segura and L. Guarino. 2000. El potencial de algunos ruitales Sur Americanos en México: Una nueva herramienta para una antigua intención.
- Segura, S.D., G. Coppens d'Eeckenbrugge and P. Ollitrault. 1998. Isozyme variation in five species of *Passiflora* subgenus *Tacsonia* and *Passiflora manicata*. In: XLIV Annual Meeting of the Interamer. Soc. Trop., September 28-October 2 1998, Barquisimeto, Venezuela.
- Segura, S.D., L. Guarino, G. Coppens d'Eeckenbrugge, M. Grum and P. Ollitrault. 1999. Mapping the distribution and regions climatically suitable for four species in *Passiflora* subgenus *Tacsonia* (Passifloraceae) and *P. manicata*. Poster presented at II Simposio de Recursos Genéticos para América Latina e Caribe - SIRGEALC, 21-26 November 1999, Brasília, Brasil.

Steiner, J.J. and S.L. Greene. 1996. Proposed ecological descriptors and their utility for plant germplasm collections. *Crop Science* 36:439-451.