M. Parra-Quijano, D. Draper & J. M. Iriondo

GIS-based evaluation of the *in situ* conservation of a Crop Wild Relative: the case of Spanish lupins

Abstract

Parra-Quijano, M., Draper, D. & Iriondo, J. M.: GIS-based evaluation of the *in situ* conservation of a Crop Wild Relative: the case of Spanish lupins. — Bocconea 21: 105-116. 2007. — ISSN 1120-4060.

The six *Lupinus* species that grow naturally in Spain were used as a model of how GIS and gap analysis can be used in the assessment of the conservation status of a crop wild relative (CWR). Data on the geographic location of *Lupinus* populations was compiled along with relevant environmental data for Peninsular Spain. This information was used to generate predictive distribution models, to identify areas of high richness in *Lupinus* species and to assess whether the current network of protected areas holds sites of high richness in *Lupinus* species that might be utilized to establish genetic reserves.

Introduction

The genus *Lupinus* (*Fabaceae*) is widely distributed around the world with more than 200 species (Ainouche & al. 2004). Although there are several classifications of lupins, they are most commonly grouped into Old and New World species. New World lupins can be divided into North and Southeast American lupins, whereas Old World lupins are classified by their rough or smooth seeds. Old World lupins refer to North African and Mediterranean lupins, including a total of 12 species (Ainouche & Randall 1999). In the *Lupinus* genus there are four cultivated species of economic importance, three of which are Mediterranean species: *L. albus* L., *L. angustifolius* L. and *L. luteus* L. Wild forms of these three species and other lupin wild relatives can be found in the Mediterranean area (Gladstones 1974, 1998; Plitmann 1981). Six *Lupinus* species grow in the Iberian Peninsula - the three Mediterranean cultivated species and three wild relatives (*L. consentinii* Guss., *L. micranthus* Guss. and *L. hispanicus* Boiss. & Reut.). Spain and Portugal share *Lupinus* diversity with all six species occurring in both countries (Castroviejo & Pascual 1999).

Crop Wild Relatives (CWRs) are very important in plant breeding programs to improve agricultural quality and production. In consequence, their conservation must be a priority especially in areas where their habitats are under threat from alteration or loss. It is worthy to note that CWRs can be common or rare species, widely distributed or reduced and abundant or endangered. CWR diversity can be conserved *ex situ* (germplasm collections) and *in situ* (protected areas) (Jarvis & al. 2003). In the case of *ex situ* conservation there are several difficulties involved in collecting and storing CWRs. In fact, less than 15% of the six million PGR accessions conserved in *ex situ* collections worldwide are CWRs (FAO 1996). *In situ* conservation seems to be the most efficient and low-cost strategy for preserving CWR diversity. Protected areas conserve many species and their ecological framework at the same time (Parra-Quijano & al. 2003). Therefore, a good relationship between management and cost is achieved: species are conserved and evolution is allowed to continue. In this sense, Europe is considered an important centre for crop wild relatives and a thematic network called PGR forum (http://www.pgrforum.org) has been created to help conserve European CWRs *in situ*.

Geographical Information Systems (GIS) are a useful tool in the management and analysis of large amounts of data with a common geographical base. As a result, GIS have been used to assess the geographic distribution of individuals, populations and species in biology and ecology. Knowledge of the geographic distribution of a target species can provide additional information such as environmental conditions or human-relation aspects (political, social, economic, etc.). GIS have been used to assess the geographic distribution of many cultivated and wild species, including some CWRs. In the case of CWRs, GIS can be used not only to assess geographic distribution but also to detect species richness areas (Hijmans & Spooner 2001), to detect bias in *ex situ* collections (Hijmans & al. 2000), to collect germplasm for *ex situ* conservation (Greene & al. 1999) and to determine the coverage of protected areas for *in situ* conservation (Parra-Quijano & al. 2003). GIS is a very flexible tool that can be used jointly with other techniques like predictive distribution models (Guisan & Zimmermann 2000). These models are based on how environmental factors can determine species distribution (Johnston 1993). To predict species potential distributions, models relate known species distributions with spatial distribution of environmental variables (Guisan & Zimmermann 2000; Zaniewski & al. 2002). Predicted distributions have been used to create potential species richness maps of wild *Arachis* and to detect hotspots and suitable areas for protection between Bolivia and Brazil (Jarvis & al. 2003). In Portugal, Draper & al. (2003) used GIS-based modelling procedures to select protected areas according to habitat suitability for wild species.

The aim of this study is to compare the results from known distribution versus predicted distribution of the *Lupinus* species in Spain in terms of species richness and coverage of *in situ* conservation (all levels of Spanish protected areas). In this way, our approach improves the knowledge of the *in situ* conservation status of a CWR species in a certain region with the contribution of predictive distribution models.

Materials and methods

The area of study was peninsular Spain. Although *Lupinus* records were available for Portugal, some environmental layers (legends and projections) of Spanish and Portuguese thematic maps were not yet compatible.

Germplasm and Herbarium data were compiled from various sources. Geo-referenced data of germplasm accessions were obtained from INIA-CRF (Centro de Recursos Fitogenéticos, Instituto Nacional de Investigación y Tecnología Agraria, Spain) and ausPGRIS (Australian Plant Genetic Resources Information System). Geo-referenced herbarium data were obtained from the Real Jardín Botánico de Madrid (MA, Spain) and bibliographic data was available from the ANTHOS project database (http://www.programanthos.org.). We compiled a total of 1870 records (usually latitude/longitude points) transformed into 1870 UTM (30 grid zone) 1x1 km coordinates representing 3946 grid cells with 500x500 m resolution. The number of records for each *Lupinus* species was as follows: *L. albus* (292), *L. angustifolius* (1081), *L. consentinii* (5), *L. hispanicus* (303), *L. luteus* (159) and *L. micranthus* (30). Due to the low number of records available, *L. consentinii* was excluded from the study.

The environmental variables used to create a geo-referenced database (UTM 500x500 m resolution) may be classified into four data types: climatic, bioclimatic (indices), physical, and soil variables that were used as a mask to filter model predictions. A total of 44 environmental variables were used (Tab. 1). Details about the bioclimatic indices are explained in Tuhkanen (1980) and Draper & al. (2003).

In addition to our presence data, pseudo-absence data also had to be obtained in order to use GLM logistic models. The easiest way to obtain pseudo-absences is to choose the locations randomly over the study area (Hirzel & al. 2001; Zaniewski & al. 2002).

Table 1. Environmental variables included in the geo-referenced database.

Data Type	Variables	Units	Source
Climatic	Temperature: monthly mean, annual mean, maximum and minimum	ºC	Sánchez-Palomares & al., 1999
	Rainfall: monthly mean and annual mean	mm	Sánchez-Palomares & al., 1999
	Dry, cold and warm period	months	Tragsatec - Spanish Minist. of Agricult.
	Thermic amplitude	ºC	Sánchez-Palomares & al., 1999
Bioclimatic indices	Emberger		(Emberger 1932)
	Gorczynski		(Gorczynsky 2004)
	Dantin-Revenga		(Dantin & Revenga 1940)
	Lang		(Lang 1965)
	Angot		(Tuhkanen 1980)
	Thornwaite		(Thornwaite 1948)
Physical	Altitude	m	Draper & al. 2003
	Aspect	º	Draper & al. 2003
	Slope	º	Draper & al. 2003
	Longitude	º	Draper & al. 2003
	Latitude	º	Draper & al. 2003
Soil	Soil type (USDA classification)	8 classes	SEISnet (http://www.microleis.com), CSIC, Spain

However, this method bears the risk of generating pseudo-absences in locations that are in fact favourable to the species (Engler & al. 2004). These authors consider that choosing a wrong absence is not too relevant in common species because the numerous presence records will counteract its effect, but that in rare and threatened plant species this procedure may not be advisable due to their low number of records. Thus, in rare and threatened plant species it is better to select pseudo-absences with the help of specialized tools like ecological niche factor analysis (ENFA) models (Guisan & Zimmermann 2000). According to us the random selection of pseudo-absences is also problematic in common species because (a) presence data is not normally complete in chorological databases and therefore may not be wholly representative, and (b) being a common species increases the probability of obtaining a wrong absence through random sampling over the area. Taking into account the nature of our CWR species data, we used an intermediate procedure to obtain pseudo-absences. We first carried out a principal component analysis (Escofier & Pàges 1991) using the environmental data associated to the locations of each *Lupinus* species. The obtained first principal component (FPC) was used to create a new synthetic variable. We then obtained the mean and standard deviation of this new variable (FPC). The pseudo-absences were thus obtained by random selection from the area resulting from the following formula:

Pseudo-absence area = Total grid cells – Presence cells – (cells with FPC values within mean ± SD)

The environmental modeling procedure used multiple logistic regression (MLR) applied to UTM 500x500 m resolution layers. Only significant and non-correlated variables were used for modeling. The equation resulting from the MLR is:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_nx_n$$

where y is the occurrence of the species, a_0 is the intercept, $a_1 \ldots, a_n$ are the regression coefficients and $x_1 \ldots, x_n$ are the independent variables.

The probability (P) (Hill & Domínguez 1994) of *Lupinus* species occurrence was obtained by:

$$P = \exp^y / (\exp^y + 1)$$

ROC (Relative Operating Characteristic) statistic was used to validate the models (Fielding & Bell 1997).

A value of $P = 0.8$ was considered the threshold above which the species is more likely to be present than absent. Predicted distributions were filtered by soil type variables, eliminating all categories where *Lupinus* is unlikely to be found (frequency < 5%). Thus, the soil classes eliminated were: aridisol (0.72%), histosol (0.32%), spodosol (0.16%), ultisol (0.48%) and vertisol (1.77%).

We also used the Sites of Community Importance (SCI) from the Natura 2000 Network (Spanish Ministry of Environment). The SCI layer was used to detect matches between habitat suitability levels and protected areas.

Idrisi Kilimanjaro and MapInfo 4.1 were used as GIS software while SPSS 10.0 and Statgraphics 5 were used for statistics and regressions.

Thematic maps of *Lupinus* distributions, *Lupinus s*pecies richness, and matches with protected areas were generated with GIS software to compare known versus potential distributions.

Results

The equations resulting from the modeling process (GLM-MLR) are shown in Table 2. Dantin index was a common factor in all equations and had a negative related effect. Lang index had a high negative weight in the models for the distribution of *L. luteus* and *L. micranthus*, showing affinity for the areas with lower rainfalls. Similarly, Emberger index was also highly negative in *L. angustifolius* and *L. hispanicus* models, keeping the species between the temperate and humid Mediterranean zone of the Iberian Peninsula.

With regard to climatic variables, rainfall affected all species models. In *L. albus* annual rainfall produced a negative effect on species occurrence, whereas in *L. angustifolius* the effect was positive. Species occurrence was positively affected by January rainfall in *L. luteus* and *L. micranthus*, March rainfall in *L. hispanicus* and *L. albus*, and November rainfall in *L. albus*. These positive correlations are coherent with the annual life form of *Lupinus* species (autumn germination and flowering in spring). Temperature variables were selected in the models for *L. albus*, *L. angustifolius*, *L. hispanicus* and *L. luteus*. *L. luteus* distribution was positively related to the variable cold period, but negatively related to December mean temperature, indicating that this species tolerates a long cold period but not extremely low temperatures.

The physical variables altitude and longitude positively affected the distribution of *L. albus* and *L. angustifolius*, respectively, while *L. hispanicus* distribution was positively affected by both.

ROC statistics ranged between 0.7 and 0.9. These values fall within the expected values for adjusting models according to Fielding and Bell (1997).

Maps with known and predicted distributions are shown in Figure 1 for the five *Lupinus* species modeled. Known and predictive distributions can be compared to identify new areas with a high probability of *Lupinus* species occurrence. The models detected potential areas for exploration for all species except *L. angustifolius*.

Table 2. Equations used for modeling *Lupinus* species distributions and their ROC statistics: R - Rainfall, T - Temperature, 01, 02,…12 - January, February,…December. Max - Maximum, Min - Minimum, Dantin - Danting-Revenga index.

Lupinus	Equation	ROC
albus	Y= -0.145+1.33*[Altitude]-5.35*[Dantin]+0.657*[R03]+2.52*[R 11]-4.78*[R anual]+4.28*[T05]	0.78
angustifolius	Y= 0.82-2.94*[Dantin]-3.49*[Emberger]-2.8*[P07]+0.7*[R anual]-0.26*[T min]+1.34*[Longitude]	0.75
hispanicus	Y=1.03+1.21*[Altitude]-4.15*[Dantin]-4.84*[Emberger]+1.02*[R 03]+0.71*[Tmax]+1.1*[Longitude]	0.81
luteus	Y=0.96-2.94*[Dantin]-7.83*[Lang]+4.262*[R 01]-2.9*[T 12]+2.13*[ColdPeriod]	0.9
micranthus	Y = -1.1-1.82*[Dantin]-7.76*[Lang]+4.06*[R01]	0.7

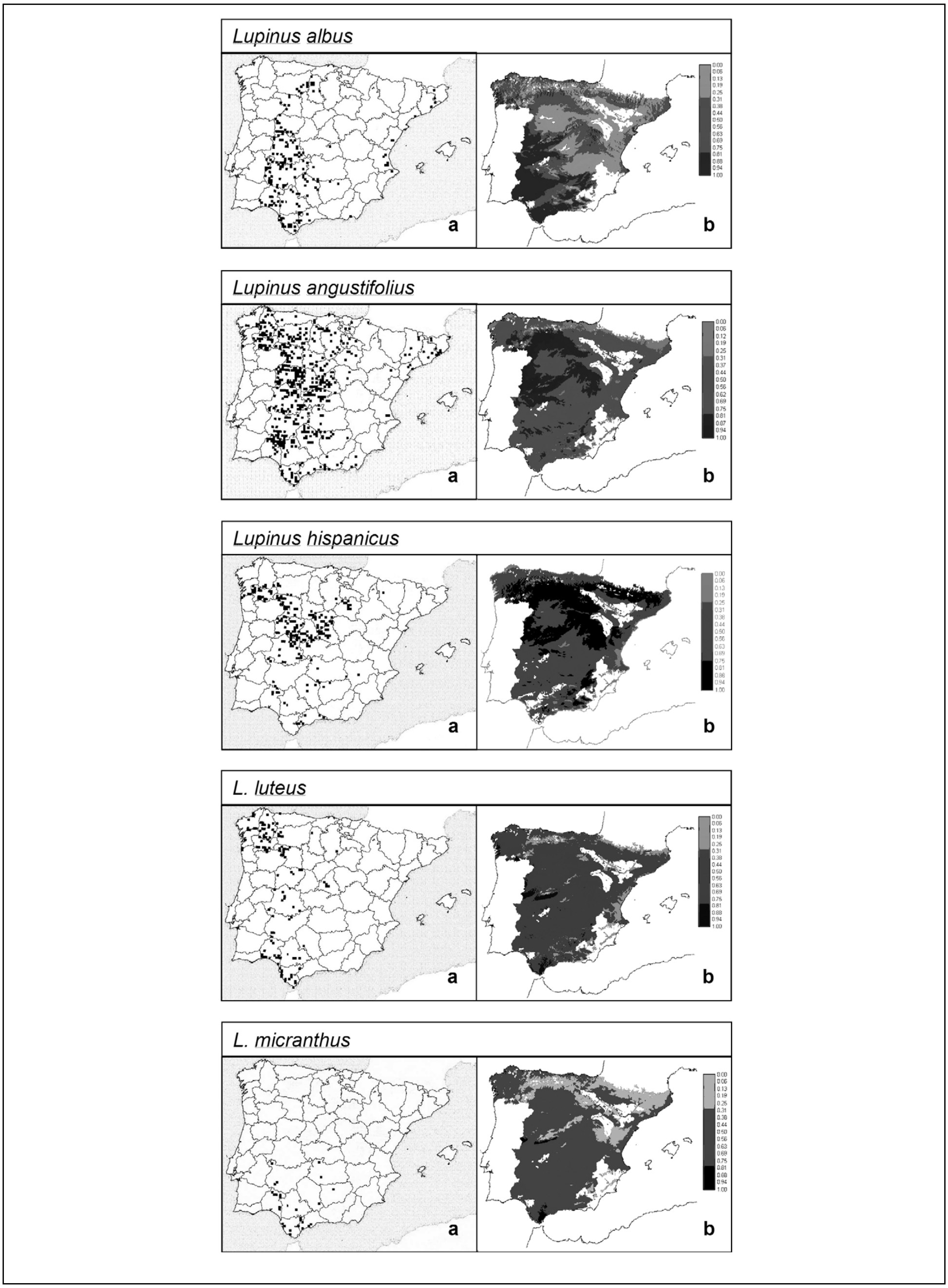

Fig. 1. Known vs. predicted distributions for five *Lupinus* species in Spain: a - known *Lupinus* populations in 10x10 km grids, b - predicted distributions generated by GLM models. The legend shows the intervals of probability according to the model: black areas correspond to a probability of 0.8 or higher, white areas have no data or have been filtered by soil type.

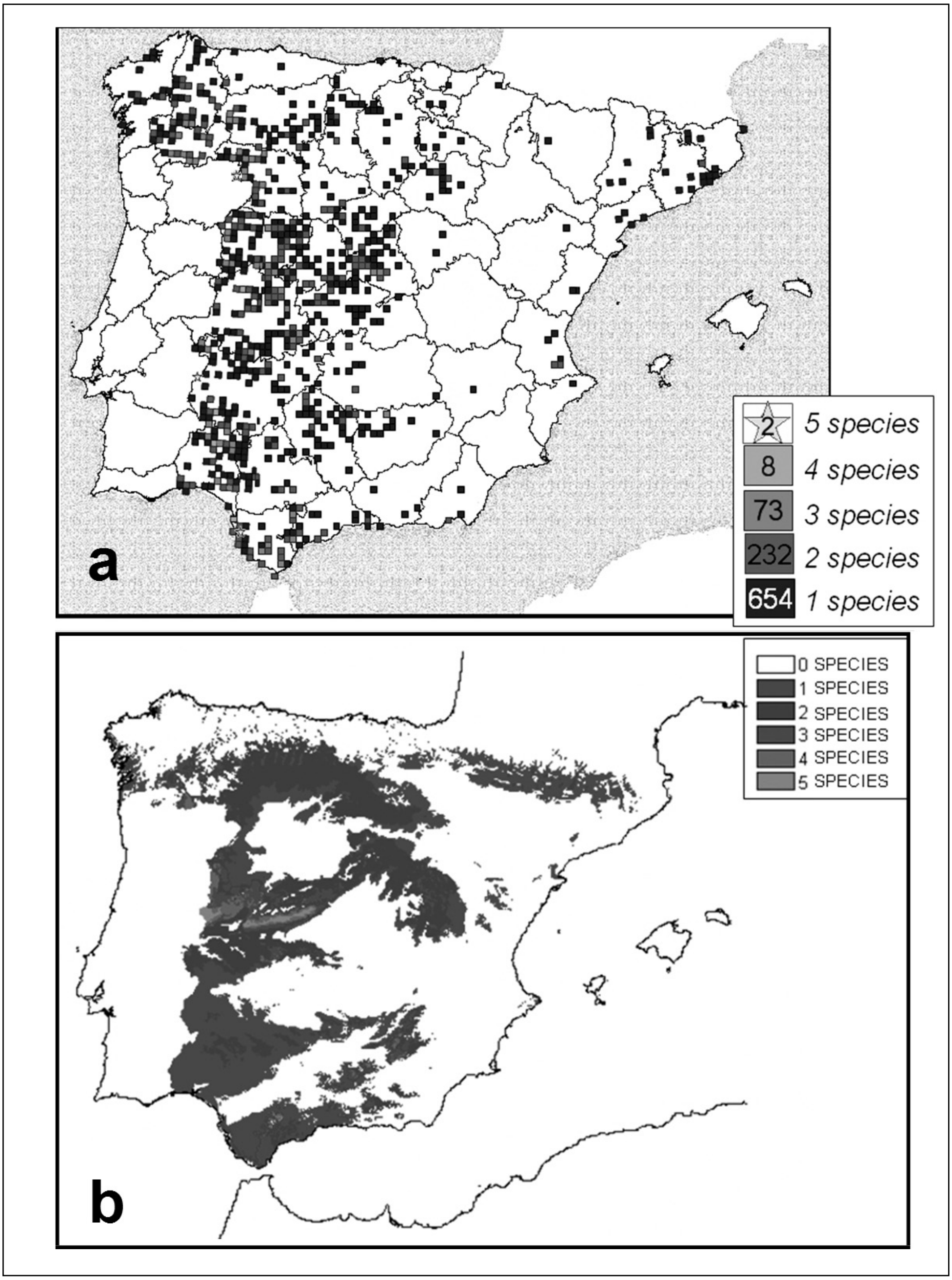

Fig. 2. Richness maps obtained from known and predicted *Lupinus* distributions: a - species richness from currently known distributions, the number in the shaded square indicates the number of 1x1 km grids that contain five, four, three, two or one species; b - species richness based on predicted distributions with p≥0.8, grid size is 500x500 m.

For *L. albus* the model detected some suitable areas to the east and north of its currently known distribution. In the case of *L. angustifolius*, the predicted distribution seems to be well represented by known populations, while for *L. hispanicus* a high probability of species occurrence was detected in several areas with no known records, especially in the Northeast. *L. luteus* and *L. micranthus* predicted distributions showed a similar pattern. In *L. luteus*, some new locations with high probability of species occurrence were found, whereas in Galicia and Andalusia some areas with known populations showed low probability of species occurrence. In *L. micranthus* a new area with high probability of occurrence far from its currently known distribution was identified in the Sierra de Gredos (Central Spain). This region showed a high probability of occurrence for all modeled species.

Known and predicted richness areas are shown in Figure 2. The richness map of currently known localities is represented by a 1x1 km grid. The richness map based on predicted localities is a product of the combination of all species predicted distributions. Shade legend is the same for both maps.

Figure 3 shows maps with the matches found between SCIs and currently known populations or predicted distributions. These maps indicate the coverage degree of Spanish protected areas over *Lupinus* populations.

The number of matches between distributions and SCI notoriously increased from 95 matches with currently known populations to 914 with predicted distribution (960%). This increase was most evident in SCIs containing 4 and 5 species.

Conclusions

The equations used by the models reflect the most influential variables affecting each *Lupinus* species distribution. A relevant relationship between the models for *L. angustifolius* and *L. hispanicus* and for *L. luteus* and *L. micranthus* is found since their most influential equation components are the same. According to equation components, *L. albus* prefers dry areas with high temperatures in April, which coincides with the period of pod filling and the end of flowering, and high rainfall in November at germination. These results are in accordance with a study on growth and yield of *L. albus* in the south of Spain (Lopez-Bellido & al. 1994). Great similarities can be found between the known and predicted distribution for *L. albus* and few new areas with a high probability of *L. albus* species occurrence were found. It is also interesting to verify how the models can also be fitted in the case of cultivated species like *L. albus*, the only *Lupinus* species cultivated on a commercial scale in Spain. *L. luteus* is cultivated in the Iberian Peninsula but mainly in Portugal. *L. angustifolius*, the third cultivated species, is found in Spain and Portugal but only in its wild form.

In the case of *L. angustifolius* and *L. hispanicus*, both species prefer semi-arid or subhumid regions according to the negative relationship with Emberger and Dantin-Revenga indices. For *L. angustifolius* the predicted distribution is very similar to the known distribution. In fact, predicted distribution does not show a high probability of finding populations in some areas where there are known populations. For this case we think that the modeling approach has been very conservative. One factor that could explain this situation is the great amount of data of known populations for *L. angustifolius* (1081 records). On

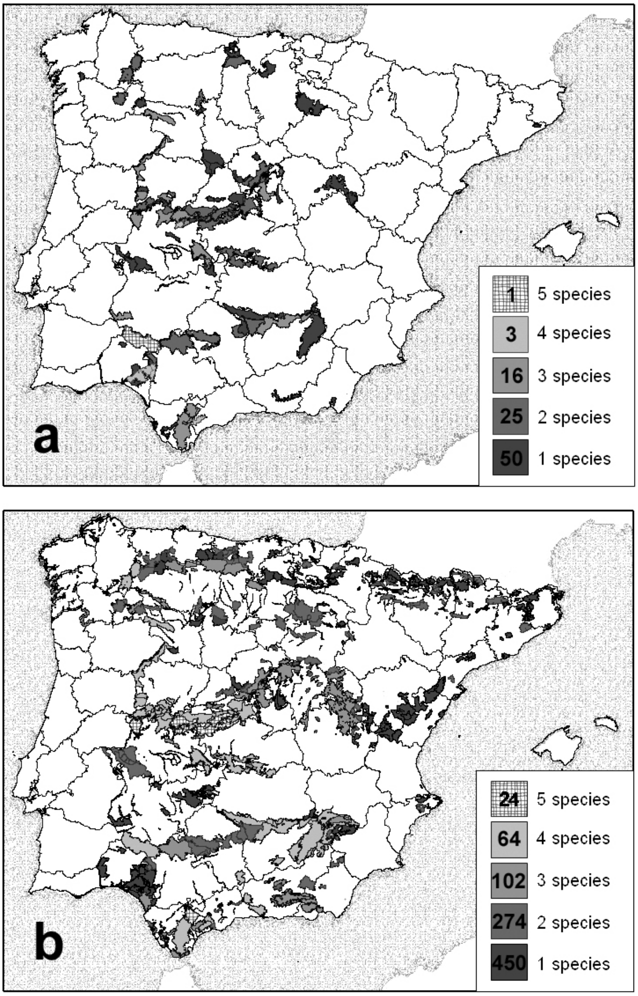

Fig. 3. Matches between SCIs and a) currently known locations or b) predicted distribution (p≥0.8) for five *Lupinus* species in Spain. The number in the shaded square indicates the number of SCI that contain five, four, three, two or one species.

the other hand, the model for *L. hispanicus* is the least conservative. *L. hispanicus* with only 303 records has a predicted distribution that shows many areas where there are no currently known populations, such as the Pyrenee Mountains. It is even possible to find areas with basic soils with a high probability of *L. hispanicus* occurrence when it is known that there is a strong relationship between *L. hispanicus* and acid soils (Castroviejo & Pascual 1999). Thus, the model for *L. hispanicus* could be improved if the final soil filter were more selective.

For *L. luteus* and *L. micranthus* equation models, the most imperative condition was the Lang rainfall index. These two species occur in arid regions, but they need rainfall in January. In the case of *L. luteus*, a cold period without extreme low temperature is important, likely for the vernalization process. For *L. micranthus* the predicted distribution is clearly oriented to the south of the Iberian Peninsula. However, the predicted distributions of both of these species detected an area with a high probability of finding new populations around the Sierra de Gredos in the center of the Iberian Peninsula where there are few *L. luteus* records and no *L. micranthus* records. The models for the other *Lupinus* species also indicated a high probability of finding new populations in this area.

It should be noted that the models applied are only based on abiotic factors. *Lupinus* species normally occur in perturbed habitats at the first stages of succession and are poor competitors. Therefore, some of the sites predicted by these models may not be suitable due to biotic factors. It is also important to take into account that *L. albus* is a cultivated species with no wild forms in Spain. Therefore, in this case, the models show potential sites for cultivation rather than natural occurrence.

On comparing predicted and known richness maps, it is evident that a larger number of SCI areas than what is currently known must contain a relevant number of different species of *Lupinus*. The predictive richness map provides relevant clues on candidate sites for a high concentration of *Lupinus* species. Thus, modeling species distribution is a tested statistical tool that allows us to identify potential high species richness areas that have not been detected previously and to select them for *in situ* conservation. In the case of *Lupinus*, efforts are currently under way in Spain to select upon the identified candidates sites that may be turned into genetic reserves for the *in situ* conservation and management of these crop wild relatives.

References

Ainouche, A. & Randall, J. B. 1999: Phylogenetic relationships in *Lupinus* (Fabaceae: Papilionoideae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA. – Amer. J. Bot. **86(4):** 590-607.

—, Bayer, R. J. & Misset, M. T. 2004: Molecular phylogeny, diversification and character evolution in *Lupinus* (Fabaceae) with special attention to Mediterranean and African lupines. – Pl. Syst. Evol. **246:** 211-222.

Castroviejo, S. & Pascual, H. 1999: *Lupinus* L. Pp. – 253-260 in: Talavera, S., Aedo, C., Castroviejo, S., Romero Zarco, C., Sáez, L., Salgueiro, F. J. & Velayos, M. (eds), Flora Ibérica: plantas vasculares de la Península Ibérica e Islas Baleares. – Madrid.

Dantin, J. & Revenga, A. 1940: Una nueva relación climatológica: El índice termopluviométrico. Avance al estudio de la aridez en España. – Zaragoza.

Draper, D., Roselló-Graell, A., García, C., Tauleigne, C. & Sérgio, C. 2003: Application of GIS in plant conservation programmes in Portugal. – Biol. Conserv. **113:** 337-349.

Emberger, L. 1932: Sur une formule climatique et ses applications en botanique. – La Météorologie: 423-432.

Engler, R., Guisan, A. & Rechsteiner, L. 2004: An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. – J. Appl. Ecol. **41:** 263-274.

Escofier, B. & Pàges, J. 1991: Presentation of correspondence analysis and multiple correspondence analysis with the help of examples. – Pp. 1-32 in: Devillers, J. & Karcher, W. (eds), Applied multivariate analysis in SAR and environmental studies. – Boston.

FAO 1996: Informe sobre el estado de los recursos fitogenéticos en el mundo. – Roma.

Fielding, A. H. & Bell, J. F. 1997: A review of methods for the assessment of prediction errors in conservation presence/absence models. – Environm. Conserv. **24:** 38-49.

Gladstones, J. S. 1974: Lupins of the Mediterranean region and Africa. – Tech. Bull., Dept. Agric., Western Australia **26:** 1-48.

Gladstones, J. S. 1998: Distribution, origin, taxonomy, history, importance. In: Gladstones, J. S., Atkins, C. A. & Hamblin, J. (eds), Lupin as crop plants: biology, production and utilization. – Oxon.

Gorczynsky, L. 2004: Sur le calcul du degré du continentalisme et son application dans la climatologie. – Geogr. Ann. **2:** 324-331.

Greene, S. L., Hart, T. C. & Afonin, A. 1999: Using geographical information to acquire wild crop germplasm for *ex situ* collections: I. Map development and field use. – Crop Sci. **39:** 836-842.

Guisan, A. & Zimmermann, N. E. 2000: Predictive habitat distribution models in ecology. – Ecol. Modeling **135:** 147-186.

Hijmans, R., Garret, K. A., Huamán, Z., Zhang, D., Schreuder, M. & Bonierbale, M. W. 2000: Assessing the geographic representativeness of genebank collections: the case of Bolivian potatoes wild potatoes. – Conserv. Biol. **14:** 1755-1765.

— & Spooner, D. M. 2001: Geographic distribution of wild potato species. – Amer. J. Bot. **88(11):** 2101-2112.

Hill, M. O. & Domínguez, F. 1994: A numerical analysis of the distribution of Liverworts in Great Britain. – Pp. 11-20 in: Hill, M. O., Preston, C. D. & Smith, A. J. E. (eds), Atlas of the Bryophytes of Britain and Ireland. – Colchester.

Hirzel, A. H., Helfer, V. & Métral, F. 2001: Assessing habitat-suitability models with a virtual species. – Ecol. Modeling **145:** 111-121.

Jarvis, A., Ferguson, M. E., Williams, D. E., Guarino, L., Jones, P. G., Stalker, H. T., Valls, J. F. M., Pittman, R. N., Simpson, C. E. & Bramel, P. 2003: Biogeography of wild *Arachis*: Assessing conservation status and setting future priorities. – Crop Sci. **43:** 1100-1108.

Johnston, C. A. 1993: Introduction to quantitative methods and modeling in community, population, and landscape ecology. – Pp. 276-283 in: Goodchild, M. F., Parks, B. O. & Steyaert, L. T. (eds), Environmental Modeling with GIS. – New York.

Lang, A. 1965: Physiology of flower initiation. – Pp. 1380-1536 in: Ruhland, W. (ed.), Encyclopedia of plant physiology. – Berlin.

López-Bellido, L., Fuentes, M., Lhamby, J. C. B. & Castillo, J. E. 1994: Growth and yield of white lupin (*Lupinus albus*) under Mediterranean conditions: effect of sowing date. – Field Crops Res. **36:** 87-94.

Parra-Quijano, M., Draper, D. & Iriondo, J. M. 2003: Assessing *in situ* conservation of *Lupinus* spp. in Spain through GIS. – Crop Wild Relative **1:** 8-9.

Plitmann, U. 1981: Evolutionary history of the old world lupines. – Taxon **30:** 430-437.

Thornwaite, C. W. 1948: An approach toward a rational classification of climate. – Geogr. Rev. **38(1):** 55-94 (reprint).

Tuhkanen, S. 1980: Climatic parameters and indices in plant geography. – Acta Phytogeogr. Suecica **67:** 1-109.

USDA & NRCS 2003: Keys to soil taxonomy. – Lincoln (Nebraska).

Zaniewski, A. E., Lehmann, A. & Overton, J. M. 2002: Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. – Ecol. Modeling **157:** 261-280.

Addresses of the authors:
Mauricio Parra-Quijano, Facultad de Agronomía, Universidad Nacional de Colombia sede Bogotá, Ciudad Universitaria, Bogotá D.C., Colombia.
David Draper,Universidade de Lisboa, Museu Nacional de História Natural, Jardim Botánico, Rua da Escola Politécnica nº 58. 1200-102 Lisboa, Portugal.
José Maria Iriondo, Departamento de Biología Vegetal, Universidad Politécnica de Madrid, 28040 Madrid, Spain.