Robustness and accuracy of Maxent niche modelling for *Lactuca* species distributions in light of collecting expeditions

M. M. P. Cobben[1,2]*, R. van Treuren[1], N. P. Castañeda-Álvarez[3],
C. K. Khoury[3,4], C. Kik[1] and T. J. L. van Hintum[1]

[1]*Centre for Genetic Resources, The Netherlands (CGN), PO Box 16, 6700 AA Wageningen, The Netherlands,* [2]*Netherlands Institute of Ecology (NIOO-KNAW), PO Box 50, 6700 AB Wageningen, The Netherlands,* [3]*International Center for Tropical Agriculture (CIAT), Apartado Aéreo 6713, Calí, Colombia and* [4]*Centre for Crop Systems Analysis, Wageningen University, PO Box 16, 6700 AA Wageningen, The Netherlands*

Abstract

Niche modelling software can be used to assess the probability of detecting a population of a plant species at a certain location. In this study, we used the distribution of the wild relatives of lettuce (*Lactuca* spp.) to investigate the applicability of Maxent species distribution models for collecting missions. Geographic origin data of genebank and herbarium specimens and climatic data of the origin locations were used as input. For *Lactuca saligna*, we varied the input data by omitting the specimens from different parts of the known distribution area to assess the robustness of the predicted distributions. Furthermore, we examined the accuracy of the modelling by comparing the predicted probabilities of population presence against recent expedition data for the endemic *Lactuca georgica* and the cosmopolitan *Lactuca serriola*. We found Maxent to be quite robust in its predictions, although its usefulness was higher for endemic taxa than for more widespread species. The exclusion of occurrence data from the perceived range margins of the species can result in important information about local adaptation to distinct climatic conditions. We discuss the potential for enhanced use of Maxent in germplasm collecting planning.

Keywords: collecting expedition; genebank; *Lactuca*; Maxent; niche modelling; plant genetic resources; species distribution modelling

Introduction

Species distribution models have been used to predict the distributions of a wide range of taxa (e.g. Guisan and Thuiller, 2005; Araújo and Guisan, 2006), and are increasingly employed for crop wild relatives (Guarino, 1995; Afonin and Greene, 1999; Greene *et al.*, 1999; Jarvis *et al.*, 2005; Maxted *et al.*, 2008; Parra-Quijano *et al.*, 2012). The Crop Wild Relatives and Climate Change project of the Global Crop Diversity Trust and Millennium Seed Bank Partnership, Royal Botanic Garden, Kew (http://www.cwrdiversity.org) uses a gap analysis methodology developed by Ramírez-Villegas *et al.* (2010), which incorporates the use of the maximum entropy model 'Maxent' (Phillips *et al.*, 2006) to support collection planning for crop wild relatives. The input data for Maxent include the geographic origins of both genebank and herbarium specimens, and statistics representing the current climate, i.e. a set of temperature and precipitation parameters. The Maxent output distribution maps intend to give an indication of locations where the species may be present.

---

*Corresponding author. E-mail: m.cobben@nioo.knaw.nl

Collectors of plant genetic resources (PGR) are interested in material with new genetic diversity, preferably from species that can be crossed with the cultivated taxa for crop improvement. Such PGR can typically be collected from regions that have not been sampled previously. The gap analysis method for crop gene pools (Ramírez-Villegas *et al.*, 2010) assesses the priority with which a certain crop wild relative should be collected using the sampling representativeness score (SRS), estimating gross representation in genebanks; the geographic representativeness score (GRS), estimating the comprehensiveness of genebank collections regarding the geographic range of the species; and the environmental representativeness score (ERS), estimating comprehensiveness based on a principal component analysis (PCA) of the full environmental range of the modelled species. As such, it is possible to get an overview of populations that are under-represented in genebanks and which may contain novel genetic diversity. The GRS and ERS depend upon Maxent distribution models. In this method, it is implicitly assumed that the herbarium data provide the full coverage of the distribution area of the species. In practice, herbarium data are incomplete and sampling bias can result in the systematic exclusion of certain regions of the species distribution from the presence data. As a result, the distribution deduced with an associative species distribution model such as Maxent is not necessarily the complete distribution of the species. Maxent improves its models by excluding part of the presence data from the training sample to subsequently use it as a test sample. However, this test sample is selected randomly, so it does not systematically exclude a specific area, which mimics the detection of an undiscovered region of the species distribution.

Here we use the distribution of the wild relatives of lettuce (*Lactuca* spp.) as a case study to investigate how the predicted distribution of *Lactuca saligna* depends on the input occurrence data. In addition, we compare the Maxent distribution predictions against the expedition data of the endemic *Lactuca georgica* and the cosmopolitan *Lactuca serriola*. The results are utilized to discuss the applicability of Maxent to support PGR collecting missions.

Materials and methods

Study material

The wild relatives of lettuce were chosen as the focus of the case study due to high available knowledge and data of this crop, and the coincidence with a *Lactuca* collecting mission in the Trans-Caucasus organized by the Centre for Genetic Resources, The Netherlands (CGN) in 2013. Moreover, the wild relatives of lettuce represent a wide variety of both niche and distribution sizes, ranging from pan-temperate distributions for *L. serriola* (D'Andrea *et al.*, 2009; Alexander, 2013) and *L. saligna*, to the narrow endemism of *L. georgica* in the Caucasus region (Zohary, 1991), and thus permit an evaluation of modelling methods for a variety of species types.

We collected information about known *Lactuca* populations from both herbarium and genebank databases (Table 1). The International *Lactuca* Database, Eurisco and the National Plant Germplasm System were used as the sources of information about genebank accessions. For herbarium samples, we consulted the Global Biodiversity Information Facility as well as obtained occurrence records directly from herbaria and researchers (see online supplementary Table S1). We used the inventory by Van Treuren *et al.* (2012) to select the species known to belong to the gene pool of cultivated lettuce (*Lactuca sativa*) (Table 1) and to check for synonyms (see online supplementary Table S2) and reassignments from other genera. Van Treuren *et al.* (2012) conducted a survey of the International Plant Names Index, which revealed a total number of 538 *Lactuca* spp., of which 357 referred to synonyms and basionyms, whereas for another 51, taxonomic status and their belonging to the genus *Lactuca* was questionable. Of the remaining 130 species, 20 are generally considered to be part of the lettuce gene pool (Table 1).

The data were cleaned; records without and with only coarse geographic information were removed. In addition, we removed the duplicate species-specific locations. For the remaining locations, we used DIVA-GIS (vs 7.5, www.diva-gis.org) (Hijmans *et al.*, 2012) to cross-check

Table 1. Occurrence data of the species from the primary (I), secondary (II) and tertiary (III) gene pools of *Lactuca sativa* in herbaria and genebanks

Species	Gene pool	Herbarium samples	Genebank accessions
Lactuca aculeata	I	3	4
Lactuca altaica	I	3	2
Lactuca azerbaijanica	I	0	0
Lactuca dregeana	I	4	0
Lactuca georgica	I	17	1
Lactuca scarioloides	I	1	0
Lactuca serriola	I	23,520	1177
Lactuca saligna	II	1451	102
Lactuca virosa	II	3318	102
Lactuca acanthifolia	III	34	0
Lactuca aurea	III	0	0
Lactuca longidentata	III	0	0
Lactuca orientalis	III	141	0
Lactuca quercina	III	106	6
Lactuca sibirica	III	854	2
Lactuca taraxacifolia	III	2	0
Lactuca tatarica	III	861	12
Lactuca viminea	III	728	12
Lactuca watsoniana	III	0	0

the match between longitude/latitude combination and the stated country.

Species distribution modelling

Current climatic conditions were downloaded from http://www.worldclim.org at a scale of 2.5 arcminutes, including 19 bioclimatic variables (Table 2). Maxent (Phillips *et al.*, 2006) uses a presence-only dataset as inputs and background points as pseudo-absences. The required 10,000 background points were selected randomly in each of the continents with species-specific occurrence data to ameliorate sample bias (VanDerWal *et al.*, 2009). Populations that were located outside the grid of climate cells, e.g. in the sea, were removed from the dataset.

Robustness of predicted distributions

L. saligna was chosen to assess the robustness of the Maxent projections in the marginal areas of the distribution range, since this species has a pan-temperate distribution with sufficient data points to exclude specific regions. Together with *L. serriola* and *L. virosa*, *L. saligna* serves as an important source of novel diversity for exploitation in the development of novel lettuce varieties by breeding companies (Lebeda *et al.*, 2009).

Occurrence locations of *L. saligna* included North America, Australia and Europe, and from there extending

Table 2. Nineteen bioclimatic variables used as input for the model, downloaded at a scale of 2.5 arcminutes from http://www.worldclim.org

BIO1	Annual mean temperature
BIO2	Mean diurnal range (mean of monthly (maximum temperature − minimum temperature))
BIO3	Isothermality (BIO2/BIO7) (×100)
BIO4	Temperature seasonality (standard deviation × 100)
BIO5	Maximum temperature of the warmest month
BIO6	Minimum temperature of the coldest month
BIO7	Temperature annual range (BIO5−BIO6)
BIO8	Mean temperature of the wettest quarter
BIO9	Mean temperature of the driest quarter
BIO10	Mean temperature of the warmest quarter
BIO11	Mean temperature of the coldest quarter
BIO12	Annual precipitation
BIO13	Precipitation of the wettest month
BIO14	Precipitation of the driest month
BIO15	Precipitation seasonality (coefficient of variation)
BIO16	Precipitation of the wettest quarter
BIO17	Precipitation of the driest quarter
BIO18	Precipitation of the warmest quarter
BIO19	Precipitation of the coldest quarter

into the eastern Mediterranean and Caucasus (see online supplementary Fig. S1). To assess the robustness of the projections, we excluded the *L. saligna* occurrence samples in, respectively, the Greek region (GRC, 88 occurrences), the Israeli region (ISR, 32 occurrences) and the Eurasian region (EUR, 684 occurrences) (see online supplementary Fig. S1). The Eurasian region was excluded to serve as a benchmark for the other predictions. The Greek and Israeli regions were very distinct marginal areas within the Eurasian distribution, from a geographic point of view (see online supplementary Fig. S1). Maxent was run for each of the three new datasets and the output was compared with the model based on the original, complete dataset. We used a tenfold division of the input data, each fold replicating the model using the consecutive parts as the test sample, while the remaining 90% of the input data were used as the training sample. For visual comparison, we used the Maxent projections, based on this tenfold cross-validation for each dataset. In addition, we correlated the Maxent estimated probabilities of occurrence of each 2.5-arcminute cell in each of the excluded regions that resulted from the models with the reduced and the original dataset. This indicates how the estimated probability of occurrence in a single cell changes when the occurrence data are excluded from the dataset on which the probability model was based.

Relationship with expedition data

A *Lactuca* expedition to the Trans-Caucasus was organized by CGN in 2013, which resulted in 94 unique collection locations. We compared the presence and absence locations of *L. georgica* and *L. serriola* with the model predictions for these locations. The model predictions were made without including the data of the sites visited during the expedition. So, while the modelling was done after the completion of the expedition, we mimicked the data availability prior to the collecting expedition, as if the modelling was done in its preparation. *L. georgica* and *L. serriola* were chosen because these two species were collected in a fair number of populations (32 and 55, respectively). In addition, they represent two opposites of the endemism spectrum, with *L. georgica* being endemic to the Trans-Caucasus, while *L. serriola* having a pan-temperate distribution, with very many data points (Table 1). To investigate the effect of zooming in on the target area, a new *L. serriola* model was made using only the 133 known occurrence samples from the expedition region, again without including the data from the newly sampled populations. The predicted probability of occurrence was determined for the 2.5-arcminute grid cell in which a population was

found. All probabilities were classified in categories ranging from 0.1 to 1.0, and the number of presence and absence locations were summed per class.

Results

Quality of the models

The four Maxent models for *L. saligna* (the full dataset and those excluding, respectively, GRC, ISR and EUR) and the model for *L. georgica* were classified as valid models (see online supplementary Table S3) according to the gap analysis protocol (Ramírez-Villegas *et al.*, 2010), for which the average test area under the curve (AUC) should be larger than 0.7, the standard deviation of the test AUC smaller than 0.15 and the proportion of the potential distribution area with a standard deviation of the estimated probability of occurrence >0.15 should be smaller than 10%. The *L. serriola* model including all occurrence data was not considered valid, as a result of the low average test AUC of 0.65. However, when the analysis was limited to the expedition region, this resulted in a valid model for *L. serriola*.

Robustness of model predictions for Lactuca saligna

For *L. saligna*, excluding the occurrence data points in the GRC resulted in small changes in the estimated probabilities of occurrence in this region. The patterns in probability distributions were very similar between the two models (Fig. 1, maps) with a good correlation between the two models (Fig. 1, scatter plot) with a small decrease in estimated probabilities from the model where the occurrences in the GRC were omitted when compared with the model that included all occurrences.

Excluding the occurrence data points in the ISR led to large changes in the estimated probabilities of the occurrence of *L. saligna* in this region (Fig. 2). The similarity in the patterns of probability distributions was recognizable (Fig. 2, maps), but the decrease in estimated probabilities from the model in which the local occurrence data were omitted was substantial (Fig. 2, scatter plot).

When all the occurrences in the EUR were excluded, the probabilities of the occurrence of *L. saligna* changed drastically (Fig. 3, scatter plot), showing increases and decreases depending on the location. The pattern of the potential distribution changed from an emphasis on Western Europe to a most probable occurrence in the Middle East and Central Asia (Fig. 3, maps). The
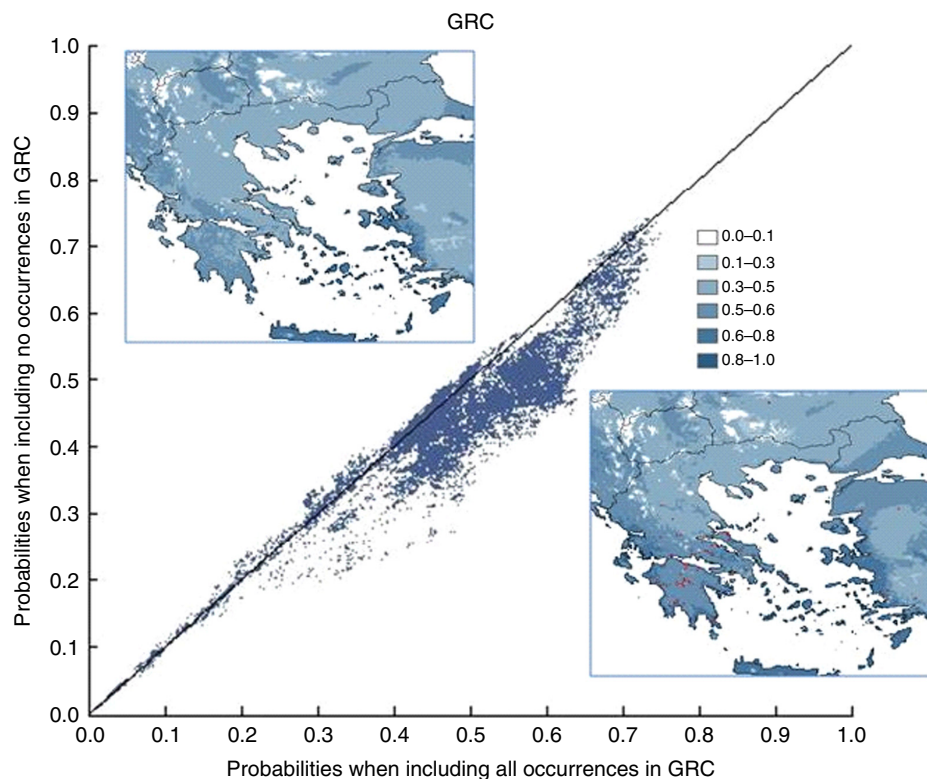

Fig. 1. Projected probabilities of occurrence for *Lactuca saligna* in the Greek region (GRC) when including and excluding the occurrence data of this region on a scatter plot showing the changes in projected probabilities for the GRC when omitting the occurrence data.
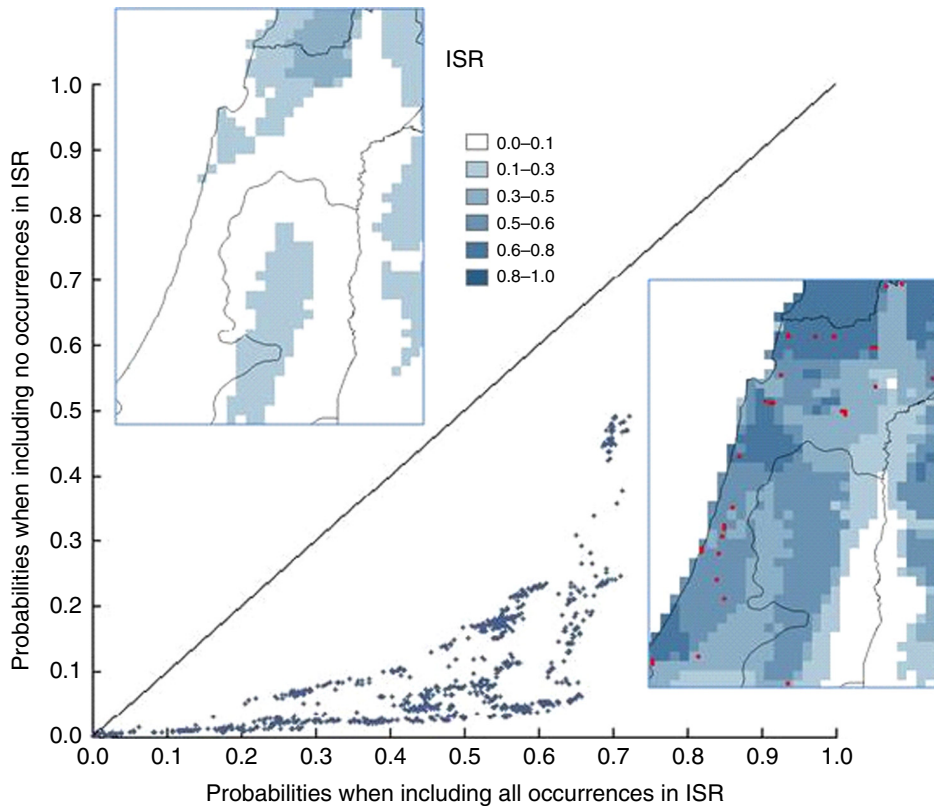
Fig. 2. Projected probabilities of occurrence for *Lactuca saligna* in the Israeli region (ISR) when including and excluding the occurrence data in this region on a scatter plot showing the changes in projected probabilities for the ISR when omitting the occurrence data.

maximum estimated probability of occurrence increased from 0.75 to 0.97. However, the margins of the potential distribution of *L. saligna* were very similar between the two models.

Model predictions in relation to expedition data

Figure 4 shows the presence and absence of both *L. georgica* and *L. serriola* in 94 unique locations sampled during the CGN *Lactuca* expedition in 2013 in relation to the Maxent estimated probabilities of occurrence in the matching 2.5-arcminute grid cells. The model performed quite well for *L. georgica*, as at locations with low estimated probabilities of occurrence, only absence of the species was observed, while at the locations with high estimated probabilities of occurrence, observed presence was considerably higher than absence of the species. In contrast, the *L. serriola* projection showed very little differentiation across the expedition region, restricting the estimated probability of occurrence to only a few classes in the middle of the potential range. The majority of locations fell within the 0.4–0.5 probability class, at which more or less equal numbers of absences and presences were observed. The new

L. serriola model, excluding all occurrence data but the ones in the expedition region, showed much more differentiation. However, the estimated probability of occurrence appeared to be a poor predictor for the presence and absence of the species.

Discussion

Species distribution modelling

Species distribution models have been used for a number of decades and for many purposes (see the review by Elith and Leathwick, 2009), such as the consequences of climate change on species distributions and the assessment of the distribution of an invasive species. With the growing availability of digital records from natural history museums, herbaria and genebanks coupled with the demand for mapped predictions, the incentive to put this source of presence-only data to use has been increasing (Elith and Leathwick, 2009). Many different modelling techniques exist, and the discussion on how to best model presence-only data continues. It is now common to compare presence data with background or pseudo-absence data, by e.g. using regression methods such as
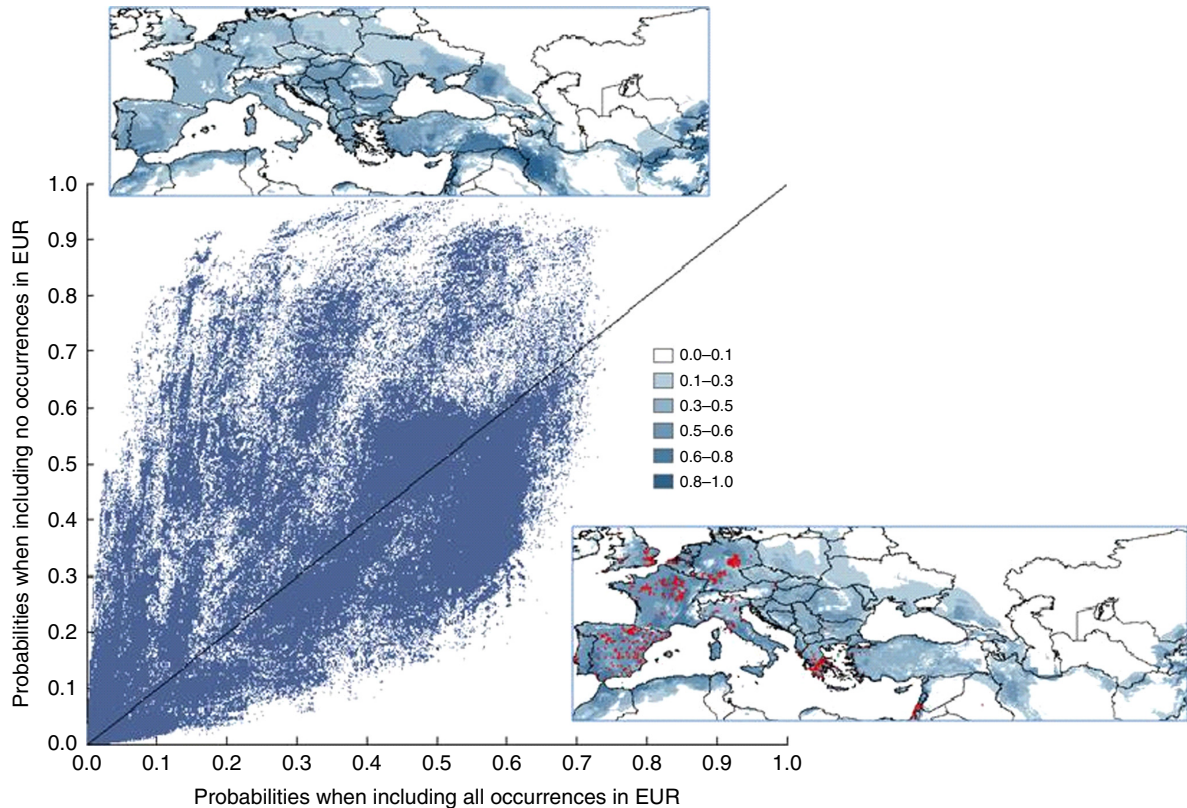
Fig. 3. Projected probabilities of occurrence for *Lactuca saligna* in the Eurasian region (EUR) when including and excluding the occurrence data in this region on a scatter plot showing the changes in projected probabilities for the EUR when omitting the occurrence data.

generalized linear models, generalized additive models or multivariate adaptive regression splines, but also GARP (Genetic Algorithm for Rule Set Production, Stockwell and Peters, 1999), ENFA (Ecological Niche Factor Analysis, Hirzel *et al.*, 2002) and Maxent (Phillips *et al.*, 2006), in chronological order. Elith *et al.* (2006) reviewed and compared these methods. More recently, a platform was developed to combine different techniques for ensemble forecasting of species distributions (Araújo and New, 2007), called BIOMOD (R-package, http://R-Forge.R-project.org) (Thuiller *et al.*, 2009). This software is able to fit and compare different model classes in an attempt to reach more robust forecasts by treating the methodological uncertainties in different models. The latest version of BIOMOD, biomod2, also includes Maxent as one of the techniques.

In this study, we chose to use Maxent as the modelling technique to align our results with the gap analysis method (Ramírez-Villegas *et al.*, 2010), which is used as input for many planned collecting expeditions in the Crop Wild Relatives and Climate Change project. With this project in mind, we aimed to provide information about the robustness and accuracy of Maxent with regard to such collecting expeditions. Apart from these considerations, the accessibility and relative ease of use of the Maxent software

compared with the others mentioned are valuable assets for application by non-experts in the planning of collecting expeditions. However, for increasing the strength of the predictions, using and comparing different modelling techniques as is done in BIOMOD is probably a valuable contribution. Compared with Maxent, this does require increased need for processing power, technical knowledge and time, which might limit the usability of BIOMOD for the non-expert.

Robustness of Lactuca saligna models

From the comparison of the different *L. saligna* models, we conclude that the Greek region is not climatically distinct within the known *L. saligna* distribution area (Fig. 1). In contrast, the Israeli region appeared to be very distinct (Fig. 2). Comparing the response curves of the Maxent predictions with the different environmental variables (see online supplementary Fig. S2), we found the Israeli region to differ from other areas within the distribution range, especially regarding the mean annual temperature (BIO1), the maximum temperature of the warmest month (BIO5), the variation in the precipitation over the seasons (BIO15), and the total precipitation of the warmest
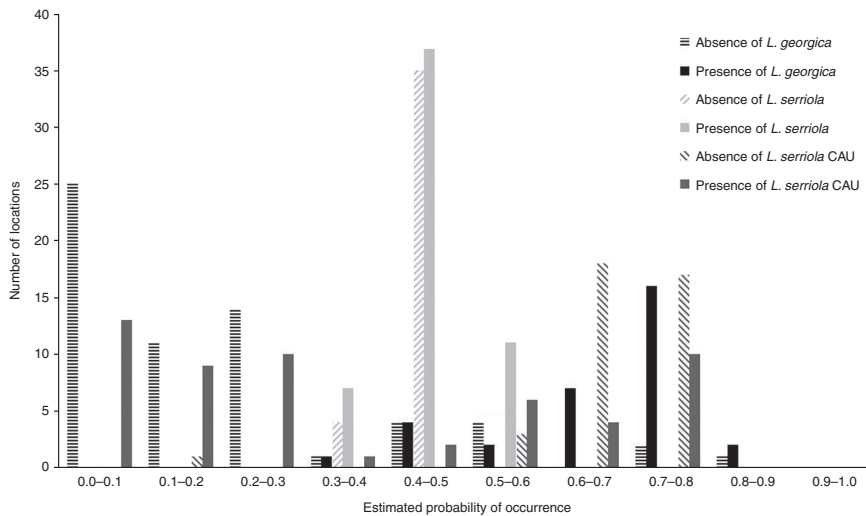
Fig. 4. Number of locations in the region under study where *Lactuca georgica* and *Lactuca serriola* were present or absent plotted against the estimated probabilities of occurrence at these locations, based on the model including all occurrence data (*L. georgica* and *L. serriola*) and only occurrence data in the region under study (*L. serriola* Caucasus (CAU)).

quarter (BIO18). Figure 2 indicates that the model is not capable of predicting this latter region as a potential distribution area when data from this region are excluded, while the unique climatic conditions may indicate the presence of potentially interesting diversity. For the Eurasian region of investigation, although the estimated local probabilities differ substantially between the two models (Fig. 3), the distribution borders are very similar. The changes in distribution pattern indicate that the southeastern region, where the probabilities in occurrence increase when the Eurasian occurrences are excluded, is climatically more similar to the other regions in the world where *L. saligna* is sampled. From this it follows that, in addition to the Israeli region, Western Europe may be considered a relatively exceptional climatic region in the *L. saligna* distribution area.

From this analysis, we conclude that the outputs of Maxent are generally robust, yet on a local scale, on which a collecting mission is typically planned, the estimated probabilities of occurrence can differ to a larger extent. This depends on whether populations in the specific region have been sampled before (and the data used as input for the model), and on the climatic relatedness to other sampled regions within the distribution. It is important to note that all species distribution models project suitable habitat by climate association, and thus that none can predict potential distribution in regions that are climatically unique compared with regions where the species has been sampled. Particularly for the purpose investigated here, in search of unique PGR, this is an important limitation. However, excluding occurrence data from the perceived range margins of the species may result in important information about local adaptation to distinct climatic conditions. A PCA of the

climatic data from the occurrence locations can provide the same information and may be a good starting point for such an analysis. There are other possibilities to obtain information about local adaptation, e.g. using SNP (single nucleotide polymorphism) data to confirm that a population is genetically different from similar populations, or checking phenotypic characteristics. Such methods tend to be used at a later stage, while the method that we suggest here can be done with currently available information about population locations to get an indication about where such useful locally adapted populations may be present.

L. saligna is native to Eurasia and North Africa (Lebeda *et al.*, 2004a, b; GRIN, 2014), and when omitting the Eurasian occurrence locations, we have essentially predicted the native distribution of the species from its non-native distribution. While this is not a logical procedure for collection planning, it gives indication of robustness of the predictions and the possibility to predict species' presence in a region where the species has not been sampled. Omitting many data points from the native area of the species then makes an interesting benchmark study with which to compare the other results from the Greek and Israeli regions. Interestingly, this procedure could be used to indicate the potential origin of populations in the non-native distribution area (Alexander, 2013).

Relationship with expedition data

L. georgica is an endemic (Zohary, 1991) species that lives in an equilibrium environment where natural competition determines the distribution of species. *L. serriola* is a ruderal species, thus living in disturbed environments

(Grime, 1977), with a cosmopolitan distribution (D'Andrea *et al.*, 2009; Alexander, 2013). The widespread distribution of the latter and the very many data points that are included in the initial model (Table 1) probably account for the rather undifferentiated model projections observed for *L. serriola*. It is interesting to note that, although not very informative, this model does provide a good prediction of its occurrence, with a similar number of presence and absence locations across the 50% probability region. When we limited the model to the expedition region, the number of probability classes increased substantially. However, actual presence and absence locations appeared to correlate poorly with the corresponding probability of occurrence according to the modelling (Fig. 4). The *L. serriola* absence locations are mostly *L. georgica* presence locations, representing fairly undisturbed habitats. In addition, the ruderal nature of *L. serriola*, combined with its global distribution, explains its relative insensitivity to climatic conditions. Thus, the Maxent model for this global, ruderal species would not have been informative or otherwise useful for the Trans-Caucasus expedition, not even when the projections would have been restricted to the region of interest. This is in line with the gap analysis protocol as suggested by Ramírez-Villegas *et al.* (2010), who excluded weedy species from the analysis. Here it needs to be noted that the collection of widespread and ruderal species such as *L. serriola* does not require any modelling support, since such species can be easily located. In contrast, expected and actual presence data correlated well for the endemic *L. georgica*, living in pristine habitat. In fact, a population was sampled in a region where the local experts had not expected it, while the model predicted a high probability of occurrence at this location. In the case that the projection for *L. georgica* would have been available prior to the collecting mission, the expedition route would have been slightly adjusted to explore a nearby region where also high probabilities of occurrence were estimated. Thus, Maxent distribution models may be useful to support collecting missions and based on our findings, this may particularly apply to endemic species growing in relatively undisturbed habitats.

Concluding remarks

Based on our results, we recommend organizers of collecting missions to run Maxent or similar Species distribution models for their species of interest prior to the expedition in complement to expert knowledge on species distributions. Given sufficient input data, particular faith may be given to the model results for endemic species among the range of relevant crop wild relatives. The resulting maps should be combined with the knowledge of local authorities to identify potential new populations of these species. In addition, excluding occurrence data from the perceived range margins of the species may result in important information about local adaptation to distinct climatic conditions.

To increase access to the methodology, avoiding the necessity of installing and operating the software, a web-based version of Maxent, including the world-clim.org climatic dataset and standard model settings, would greatly facilitate the application of species distribution modelling in the preparation phase of collecting missions, and would be particularly useful for plant genetic resources conservation efforts with limited resources.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262114000847

References

Afonin A and Greene SL (1999) Germplasm collecting using modern geographic information technologies: directions explored by the N.I. Vavilov Institute of Plant Industry. In: Greene SL and Guarino L (eds) *Linking Genetic Resources and Geography: Emerging Strategies for Conserving and Using Crop Biodiversity*. CSSA Special Publication 27. Madison, WI: American Society of Agronomy, pp. 75–85.
Alexander JM (2013) Evolution under changing climates: climatic niche stasis despite rapid evolution in a non-

native plant. *Proceedings of the Royal Society B* 280: 20131446.

Araújo MB and Guisan A (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33: 1677–1688.

Araújo MB and New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22: 42–47.

D'Andrea L, Broennimann O, Kozlowski G, Guisan A, Morin X, Keller-Senften J and Felber F (2009) Climate change, anthropogenic disturbance and the northward range expansion of *Lactuca serriola* (Asteraceae). *Journal of Biogeography* 36: 1573–1587.

Elith J and Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40: 677–697.

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JMM, Townsend Peterson A, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS and Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.

Greene SL, Hart TC and Afonin A (1999) Using geographic information to acquire wild crop germplasm for *ex situ* collections: I. map development and field use. *Crop Science* 39: 836–842.

Grime JP (1977) Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *The American Naturalist* 111: 1169–1194.

GRIN (2014) http://www.ars-grin.gov/cgi-bin/npgs/html/taxon.pl?21359

Guarino L (1995) Mapping the ecogeographic distribution of biodiversity. In: Guarino L, Ramanatha Rao V and Reid R (eds) *Collecting Plant Genetic Diversity Technical Guidelines*. Wallingford, UK: CAB International, pp. 287–327.

Guisan A and Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993–1009.

Hijmans RJ, Guarino L and Mathur P (2012) DIVA-GIS vs 7.5. A geographic information system for the analysis of species distribution data Software and manual available at www.diva-gis.org

Hirzel AH, Hausser J, Chessel D and Perrin N (2002) Ecological-niche factor analysis: how to compute habitat-suitability map without absence data. *Ecology* 83: 2027–2036.

Jarvis A, Williams K, Williams D, Guarino L, Caballero PJ and Mottram G (2005) Use of GIS for optimizing a collecting

mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. *Genetic Resources and Crop Evolution* 52: 671–682.

Lebeda A, Dolezalová I, Feráková V and Astley D (2004a) Geographical distribution of wild *Lactuca* species (Asteraceae, Lactuceae). *The Botanical Review* 70: 328–356.

Lebeda A, Dolezalová I and Astley D (2004b) Representation of wild *Lactuca* spp. (Asteraceae, Lactuceae) in world genebank collections. *Genetic Resources and Crop Evolution* 51: 167–174.

Lebeda A, Dolezalová I, Křístková E, Kitner M, Petrželová I, Mieslerová B and Novotná A (2009b) Wild Lactuca germplasm for lettuce breeding: current status, gaps and challenges. *Euphytica* 170: 15–34.

Maxted N, Dulloo E, Ford-Lloyd BV, Iriondo JM and Jarvis A (2008) Gap analysis: a tool for complementary genetic conservation assessment. *Diversity and Distributions* 14: 1018–1030.

Parra-Quijano M, Iriondo JM and Torres E (2012) Improving representativeness of genebank collections through species distribution models, gap analysis and ecogeographical maps. *Biodiversity and Conservation* 21: 79–96.

Phillips SJ, Anderson RP and Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.

Ramírez-Villegas J, Khoury C, Jarvis A, Debouck DG and Guarino L (2010) A gap analysis methodology for collecting crop gene pools: a case study with *Phaseolus* beans. *PLoS ONE* 5: e13497.

Stockwell D and Peters D (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* 13: 143–158.

Thuiller W, Lafourcade B, Engler R and Araújo MB (2009) BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32: 369–373.

Van Treuren R, Coquin P and Lohwasser U (2012) Genetic resources collections of leafy vegetables (lettuce, spinach, chicory, artichoke, asparagus, lamb's lettuce, rhubarb and rocket salad): composition and gaps. *Genetic Resources and Crop Evolution* 59: 981–997.

VanDerWal J, Shoo LP, Graham C and Williams SE (2009) Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling* 220: 589–594.

Zohary D (1991) The wild genetic resources of cultivated lettuce (*Lactuca sativa* L.). *Euphytica* 53: 31–35.