Identification of ecogeographical gaps in the Spanish *Aegilops* collections with potential tolerance to drought and salinity

Rosa María GARCÍA¹, Mauricio PARRA-QUIJANO², Jose María IRIONDO³

¹Centro Nacional de Recursos Fitogenéticos (CRF-INIA), Alcalá de Henares, Madrid, Spain
²Fundación Agrobiodiversidad Colombiana, Ctra. 88 A 21-42 425, Bogotá D.C., Colombia
³Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain.

Corresponding Author:
Rosa María García¹
Email address: rosamaria.garcia@inia.es
Abstract

Drought, one of the most important abiotic stress factors limiting biomass, significantly reduces crop productivity. Salinization also affects the productivity of both irrigated and rain-fed wheat crops. Species of genus Aegilops can be considered crop wild relatives (CWR) of wheat and have been widely used as gene sources in wheat breeding, especially in providing resistance to pests and diseases. Five species (Ae. biuncialis, Ae. geniculata, Ae. neglecta, Ae. triuncialis and Ae. ventricosa) are included in the Spanish National Inventory of CWRs. This study aimed to identify ecogeographic gaps in the Spanish Network on Plant Genetic Resources for Food and Agriculture (PGRFA) with potential tolerance to drought and salinity. Data on the Spanish populations of the target species collected and conserved in genebanks of the Spanish Network on PGRFA and data on other population occurrences in Spain were compiled and assessed for their georeferencing quality. The records with the best geo-referencing quality values were used to identify the most important ecogeographical variables for Aegilops distribution in Spain. These variables were then used to produce ecogeographic land characterization maps for each species, allowing us to identify populations from low and non-represented ecogeographical categories in ex situ collections. Predictive characterization strategy was used to identify 45 Aegilops populations in these ecogeographical gaps with potential tolerance to drought and salinity conditions. Further efforts are being made to collect and evaluate these populations.

Introduction

Drought is one of the most important abiotic stress factors limiting biomass (Araus et al., 2002), and consequently, it significantly reduces crop productivity (Lambers et al., 2008; Noorka & Heslop-Harrison, 2014). Wheat (Triticum spp.), the second most-produced cereal in the world (FAO, 2013a), can be severely affected by this type of abiotic stress. For instance, in 2009, wheat yield in Kenya dropped by 45% due to drought, compared to 2010 production that took place under a good crop season. Australia, which suffered multi-year droughts between 2002 and 2010, experienced a 46% drop in wheat yield in 2006 (FAO, 2013b).

Another threat to both irrigated and rain-fed wheat crops is salinization (Mujeeb-Kazi & Diaz de Leon, 2002). Salt stress and drought have similar effects on structural and functional aspects of plants (Al-maskri et al., 2014), but salt stress also causes ion toxicity and ionic imbalance (Hameed et al., 2011). Drought stress decreases photosynthetic efficiency, greatly reducing growth and development (Al-maskri et al., 2014). Salt-affected soils occur in all continents and under almost all climatic conditions. However, they are more widely distributed in arid and semi-arid regions than in humid regions (Abrol et al., 1988). Many crops in these areas are grown under irrigation, but inadequate irrigation management may lead to secondary salinization (Glick et al., 2007). Large areas of naturally saline and alkaline soils account for 6% of...
the world’s land surface. These saline soils have never been cultivated because present major crops are salt-sensitive (Fita et al., 2015).

The species of genus *Aegilops* have been widely used as gene sources in wheat breeding, especially in providing resistance to pests and diseases such as leaf, stem and stripe rusts (*Puccinia recondita*, *P. graminis* and *P. stiiformis*) or hessian fly (*Mayetiola destructor*). Numerous studies have searched for drought and salt stress tolerant genotypes. For instance, Xing et al. (1993) studied the potential of some *Aegilops* species, including *Ae. ventricosa* Tausch, as gene donors in breeding for salt tolerance. Subsequent studies proposed *Ae. ovata* L. as a source of salt tolerance in wheat (Farooq, 2002) and determined that *Ae. ovata* and *Ae. biuncialis* Vis. have wide genetic variation for salt tolerance (Colmer et al., 2006). Molnár et al. (2004) compared the physiological and morphological responses to water stress in *Ae. biuncialis* and *Triticum aestivum* L. genotypes, and concluded that *A. biuncialis* genotypes from dry habitats have greater drought tolerance than wheat, making them good candidates for improving drought tolerance in this crop. Mondini et al. (2014) identified SNPs variants conferring salt tolerance in durum wheat. Due to their present and potential use as gene donors in wheat breeding, *Aegilops* species can be considered crop wild relatives (CWR) of wheat (Heywood et al., 2007). Five of these species (*A. biuncialis*, *A. geniculata* Roth, *A. neglecta* Req. ex Bertol., *A. triuncialis* L. and *A. ventricosa*) are included in the Spanish National Inventory of CWRs (Rubio et al., 2013).

Ecogeographical land characterization (ELC) map can be helpful in determining different adaptive scenarios of a species in a given territory. An ELC map represents the different ecogeographical conditions in which a particular species or group of species occur, using some variables of high importance in the species’ distribution likely to be determinant for the adaptive landscape (Parra-Quijano et al., 2012a).

Parra-Quijano et al. (2008) developed an ELC map for Peninsular Spain and the Balearic Islands, using different sources of ecogeographical information. The ability of the ELC map to discriminate different areas with different adaptive pressures was tested with eight crop and CWR species. They found that the ELC map had an effective discriminatory capacity to delineate adaptive scenarios. The ELC concept is detailed in depth in Parra-Quijano et al. (2012a), and several applications related to the collection, conservation and efficient use of plant genetic resources have been developed (e.g., Parra-Quijano et al., 2012a; Parra-Quijano et al., 2012b; Thormann, 2016).

Gap analysis has been widely applied for conservation purposes (Maxted et al., 2008). For instance, Ramírez-Villegas et al. (2010) applied a gap analysis methodology to collect the crop genepool of *Phaseolus* beans and evaluated conservation deficiencies at three different levels (taxonomic, geographic and environmental). Khoury et al. (2010) reviewed global crop and regional conservation strategies and recognized the importance of filling gaps in genebanks of plant genetic resources. Recently, Shehadeh et al. (2014) carried out a gap analysis of *Lathyrus* L. species. In this study, predictive distribution maps for each *Lathyrus* taxon were produced based on climatic data, and ex
situ conservation gaps were identified as regions where the species was predicted to occur but seed accessions had not been previously collected, or, alternatively, the species was under-sampled.

Optimized Collecting Design (OCD) is a technique described by Parra Quijano et al. (2012b) that involves the identification of ecogeographical gaps for a target species in a target ex situ genebank. Based on the premise that genetic adaptation is achieved through natural selection acting upon particular limiting environmental conditions, this technique aims to improve the genetic representativeness of genebank accessions by improving their ecogeographical representativeness. Like other gap analysis techniques, OCD compares the collecting locations of the target species accessions currently held in the genebanks and the species’ occurrence data from external sources (spatial gaps). It then uses ELC maps to detect adaptive scenarios not represented in the target ex situ genebank. Using ELC maps to design collecting strategies can help to include accessions from marginal or under-represented environments that may contain important traits related to adaptations to biotic and abiotic stress.

Another interesting issue for genebank managers and stakeholders is the identification of genotypes that can have a specific use in plant breeding. The Focused Identification of Germplasm Strategy (FIGS) is a useful approach for screening large germplasm collections to identify sets of accessions with a high probability of containing specific target traits based on the ecogeographical information of the sites where the populations were collected (Mackay & Street, 2004). If we know where a set of ex situ accessions has evolved, or at least where they have grown for a period long enough for adapted genotypes to have been selected, we can establish relations or patterns between the environmental conditions of the site and the presence or absence of the target trait. We can then make predictions on non-evaluated germplasm (Mackay & Street, 2004).

In recent years, FIGS has been successfully used to identify sources of resistance to sunn pest in Syria (El Bouhssini et al., 2009) and to Russian wheat aphid in bread wheat (El Bouhssini et al., 2011). FIGS has also been used to identify traits related to abiotic stresses, such as drought adaptation in Vicia faba L. (Khazaei et al., 2013). These studies selected accessions from an ecogeographically-characterized collection that complied with certain values or ranges for the characterized variables, set by the researchers based on their knowledge of the species. Other FIGS approaches can be used to identify accessions of potential interest, when the trait under consideration cannot be directly related to an ecogeographical variable. In this case, partial evaluation data from the target collection are required to detect a reliable relationship between the ecogeographic variables and the trait of interest. This approach has been successfully applied by Thormann et al. (2016) and Endresen et al. (2012), who identified sources of resistance to stem rust in bread and durum wheat, and by Bari et al. (2011, 2014), who predicted resistance to stem rust and stripe rust in accessions of wheat landraces.
A broader concept has recently been developed to identify germplasm with a high probability of containing specific target traits, named “predictive characterization”. This term comprises a set of approaches that use geographic and environmental data to search for particular traits in a usually large set of populations, not only accessions and landraces but also CWR (Thormann et al., 2014). This search can be carried out by means of the ecogeographical filtering method or the calibration method (Thormann et al., 2014).

Using these ecogeographical approaches, we aimed to apply OCD and predictive characterization techniques to the optimization of the ex situ collection of crop wild relatives of wheat in the national reference genebank of Spain. The objectives of this study were: (1) to identify spatial and ecogeographical gaps within Spain land, which could be present in the Spanish genebank collections of Aegilops biuncialis, A. geniculata, A. neglecta, A. triuncialis and A. ventricosa, to design an optimized systematic collection strategy of crop wild relatives of wheat; and (2) to identify non-collected populations that might be of potential interest because of their tolerance to drought and salinity within the group of ecogeographical gaps for each species in order to prioritize their collection.

Materials & Methods

Species datasets

Data on Spanish populations of Aegilops spp. collected and conserved in the genebanks of the Spanish Network of Plant Genetic Resources for Food and Agriculture (PGRFA) (hereinafter “accessions”) were obtained from the Spanish Inventory of Plant Genetic Resources (available at http://www.inia.es/inventarionacional/, accessed 26 May 2015).

Data on population occurrences from additional sources (hereafter “external sources”) were obtained from the Global Biodiversity Information Facility (GBIF; available at http://www.gbif.org/, accessed 14 February 2014) and Anthos (available at http://www.anthos.es/, accessed 25 June 2015) databases. Populations conserved in ex situ genebanks that do not belong to the Spanish Network on PGRFA were also considered external sources. Accessions missing in the Spanish National Inventory, provided by the Plant Genetic Resources National Centre of the Spanish National Institute for Agricultural and Food Research and Technology (http://www.inia.es/coleccionescrf), were considered external sources as well.

Species names were standardized using GRIN Taxonomy (USDA, ARS, National Genetic Resources Program) except for Aegilops ovata L. According to GRIN, this species is synonymous to Aegilops neglecta Req. ex Bertol, but according to Anthos, the main supplier of information on Aegilops ovata L., this taxon is synonymous to Aegilops geniculata Roth. Thus, the latter was adopted in this study, and the targeted
species were *A. biuncialis* Vis., *A. geniculata* Roth, *A. neglecta* Req. ex Bertol., *A. triuncialis* L. y *A. ventricosa* Tausch.

All accessions and external sources without geographic coordinates were removed. We also removed accessions with geographic coordinates in sexagesimal system that did not include seconds and external sources with geographic coordinates expressed in decimal degrees with less than two decimals in both latitude and longitude or without textual information on the occurrence site. Passport and presence data were standardized to CAPFITOGEN data formats which is basically the Multi-Crop Passport Descriptor (FAO-Biodiversity, 2012) plus four additional administrative fields for collecting or presence site description (Parra-Quijano et al., 2015).

Georeferencing data of both accessions and external sources were cleared of spatial intraspecific duplicates. We considered that species occurrences less than 1 km apart belonged to the same population, following Iriondo et al. (2009). Therefore, these population occurrences were considered spatial duplicates and only one of them was considered.

Accessions and external sources free of spatial duplicates were subjected to a georeferencing quality evaluation using GEOQUAL from the CAPFITOGEN toolkit (Parra-Quijano et al., 2015). We set the quality threshold in TOTALQUAL100 = 80, so only records with quality values above this threshold were considered in subsequent analyses.

**Selection of ecogeographical variables**

Ecogeographical information was extracted for each occurrence site from raster layers with a 30 arc-second resolution and classified into three ecogeographical components: bioclimatic variables (17), edaphic variables (16) and geophysic variables (4) (see Table S1).

The most important variables for each species in each ecogeographical component were then identified using the SelecVar tool from CAPFITOGEN (Parra-Quijano et al., 2015). SelecVar extracts information from the ecogeographical variables (layers) to the occurrence sites and assesses the importance of each variable in generating different adaptive scenarios for a species (Parra-Quijano et al., 2015). It estimates variable importance according to the random forest classification (RFC) and detects redundant variables through bivariate correlation analysis. The RFC analysis provides a ranking of the most important variables for establishing ecogeographical categories, placing variables with a higher mean decrease in accuracy in the first positions (Cutler et al., 2007). Rankings were obtained for each ecogeographical component for each species. Bivariated correlation analysis detected correlated variables in the top fifteen variables of the RFC ranking. Variables with Pearson correlation coefficient $>|0.50|$ and p-value $<0.05$ in the same ecogeographical component were identified and removed.
Generation of the ELC maps

An ecogeographical land characterization map or ELC map is a representation of the different adaptive scenarios of a species (Parra-Quijano et al., 2012a). The three most important bioclimatic and edaphic variables and the two most important geophysical variables were considered in generating the ELC map for each species. The variables latitude and longitude were included as two additional geophysical variables to obtain maps with spatially aggregated categories. The “elbow” method was used to create the ecogeographical categories. This is a simple system which uses K-means as a clustering algorithm where the cut-off point is determined on the basis of the decrease in the sum of the intra-group squares (Ketchen & Shook, 1996). The optimal number of categories is reached when the decrease in the intra-group sum of squares in a range of \( n \) and \( n+1 \) groups is less than 50%. ELC maps were generated for each species using the ELCmapas tool of the CAPFITOGEN toolkit (Parra-Quijano et al., 2015) and a 30-arc-second cell size, considering the area of Spain.

Identification of spatial and ecogeographical gaps and prioritization of occurrence sites for future collections in *Aegilops*

Spatial gaps were identified based on occurrence site coordinates. For each species, external sources that were more than 1 km apart from accessions were considered spatial gaps. An ecogeographical representativeness analysis of the existing germplasm collections was then carried out to identify ecogeographical gaps (ELC map categories not represented in ex situ genebanks where the species occurs). Areas with a high occurrence of external sources (listed as ecogeographical gaps) and a low occurrence of accessions were considered priority collecting sites, as were areas corresponding to the ELC map categories with a low frequency for the species and in the territory. Occurrence data from external sources were then ranked according to their priority of collection based on the frequency of each ELC map category in the study area, the frequency of each species in each ELC map category and the differences between the external sources data set and the accessions data set. Ecogeographical gap identification and prioritization of external sources occurrence data for germplasm collection were performed using the Representa tool of CAPFITOGEN (Parra-Quijano et al., 2015). External sources which occurred in ecogeographical categories not represented by the corresponding species in the genebanks of the Spanish Network, i.e., external sources reclassified by the Representa tool within the range from 1 to 4, were considered high priority ecogeographical gaps.

Selection of collection sites for traits of tolerance to drought and salinity
To search for populations with a higher probability of containing phenotypes with a high tolerance to drought and salinity, the external sources considered high priority ecogeographical gaps were filtered using the Lang aridity index ($A_{IL}$) and topsoil salinity. The Lang aridity index was calculated as:

$$A_{IL} = \frac{\text{Annual precipitation}}{\text{(Annual mean temperature)}}$$

Following the ecogeographical filtering technique of predictive characterization (Thormann et al., 2016), we selected populations occurring in sites with $A_{IL} < 40$. We considered the 20% of the populations with the lowest Lang aridity index and the highest topsoil salinity values to be the fraction of interest for each of the five target species.

The process followed to reach the objectives of the study is shown in Fig. 1.

**Fig. 1** Process carried out to identify ecogeographical gaps in the Spanish *Aegilops* germplasm collections with potential tolerance to drought and salinity

### Results

#### Germplasm collection sites and presence data

Data pre-processing showed that the most frequently recorded species and those with the widest distribution in Spain were *A. geniculata* and *A. triuncialis*. The least frequently recorded species was *A. biuncialis*. Table 1 shows the number of accessions and occurrence data from external sources for each species before and after clearing spatial duplicates and applying the geo-referencing quality threshold. The percentage of spatial duplicates in the occurrence data of external sources ranged from 47 to 65%, recorded for *A. biuncialis* and *A. geniculata*, respectively. Applying the geo-referencing quality threshold decreased the number of non-duplicated populations to 6 and 21% in the occurrence data of external sources of *A. biuncialis* and *A. geniculata*, respectively.

<table>
<thead>
<tr>
<th>Table 1 Number of <em>Aegilops</em> germplasm accessions and occurrence records from external sources with geographical coordinates included in the study before and after clearing spatial duplicates and filtering by the geo-referencing quality threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>The species records (accessions + external sources) remaining after the clearing of spatial duplicates and the geo-referencing quality threshold are shown in Fig. 2. The populations represented by accessions in the genebanks are not homogeneously distributed in the studied area (see Fig. 2a), nor are they more frequently located in the areas where the presence of these species was reported by external sources. For instance, 66% of the preserved populations of the five target species (226 out of 345 populations) were collected in the autonomous communities of Extremadura and Castilla-La Mancha, whereas only 10% of the external sources (272 out of 2614 populations) are located in these two communities. The populations reported by external</td>
</tr>
</tbody>
</table>
sources are abundant in the south of Spain, in the provinces on the eastern coast and in
the autonomous community of Navarra.

**Fig. 2** Location of (a) origin of germplasm accessions in Peninsular Spain, the Balearic
Islands and the Canary Islands and (b) population occurrence from external sources of
the targeted *Aegilops* species

**Identification of variables of importance**

The variables selected for each studied species and ecogeographical component
(bioclimatic, geophysic and edaphic variables) are shown in Table 2. The variables
isothermality and altitude were selected in four of the five studied species.

**Table 2** Variables selected in each *Aegilops* species and ecogeographical component
according to the importance function of the random forest approach (see Table S1 for
variable description)

**Generation of ELC maps**

The ELC maps obtained for each species are shown in Fig. 3. The number of
ecogeographical categories generated in the maps ranged between 26 (in the ELC map
for *A. biuncialis*) and 27 (in the maps for the other species). There are some similarities
among the five maps. All of them contain one large ecogeographical category that
spreads out on the northern and southern sub-mesetas. The distribution of the categories
in the southwest of the country, as well as in the Pyrenees, is also quite similar in the
five maps. Finally, several categories coexist in the relatively small area of Sierra
Nevada. However, the similarities between the maps are less apparent in the case of *A.
triuncialis*.

The environmental characteristics of the different categories of each ELC map are
summarized in Table S2.

**Fig. 3** ELC maps of (a) *Aegilops biuncialis*, (b) *A. geniculata*, (c) *A. neglecta*, (d) *A.
triuncialis* and (e) *A. ventricosa* for Peninsular Spain, the Balearic Islands and the
Canary Islands

**Identification of spatial and ecogeographical gaps in the *Aegilops* germplasm
collections**

Among the populations reported by external sources, 2571 were identified as spatial
gaps. *A. geniculata* was identified as the species with the largest number of spatial gaps.
As seen in Table 3, 393 populations were identified as priority ecogeographical gaps in
the registered genebank collections of *Aegilops* in the Spanish National Inventory of
Plant Genetic Resources. These 393 populations occur in ecogeographical categories that are not represented by the corresponding species in the Spanish Network. *A. geniculata*, the first-ranking species in number of spatial gaps, was also identified as the species with the largest number of ecogeographical gaps. *A. biuncialis*, the species with the lowest number of preserved accessions (Table 1), is also the species whose *ex situ* ecogeographical representativeness needs the most improvement, as 80% of the available external sources were identified as high priority gaps. On the contrary, only 10% of the analyzed external sources of *A. geniculata* and *A. triuncialis*, the two species with the highest number of preserved accessions (Table 1), were identified as high priority ecogeographical gaps.

**Table 3** Number of population occurrences from external sources of *Aegilops* subjected to representativeness analysis and number of spatial gaps and priority ecogeographical gaps identified in Spain.

The geographic distribution of the populations identified as priority ecogeographical gaps is shown in Fig. 4.

**Fig. 4** Location of the *Aegilops* populations identified as priority ecogeographical gaps in Spain.

Including germplasm from priority ecogeographical gaps in the genebanks of the Spanish Network on PGRFA would significantly improve their ecogeographical representativeness (number of ecogeographical categories not currently represented in the Spanish Network, Table 4). The obtained percentage of ELC categories represented in the Spanish Network ranged from 8% to 48% in *A. biuncialis* and *A. geniculata*, respectively. As the priority ecogeographical gaps belong to categories not yet represented, their collection and conservation would contribute to increasing the percentage of ELC categories represented in the Spanish Network to values ranging from 27% to 70% in *A. biuncialis* and *A. neglecta*, respectively.

**Table 4** Number of ELC categories for *Aegilops* currently represented in the Spanish Network and potential increase (%) in representativeness after collecting priority ecogeographical gaps.

**Selection of priority collecting sites for traits of tolerance to drought and salinity**

Among the 393 populations identified as priority ecogeographical gaps, 223 populations inhabit sites with a Lang index value < 40, and thus are potentially adapted to arid environments (Table 5). The geographic location of these accessions is shown in Fig. 5a. The 20% of these 223 populations with the highest values of topsoil salinity for each species (20%) and their geographical distribution are shown in Table 5 and Fig. 5b, respectively. These 45 populations (Table 5) constitute the predictive characterization (PC) subset of *Aegilops* populations of potential interest due to their potential tolerance to drought and salinity, in addition to representing ecogeographical gaps. Table 6
contains complete information on the geographic location, the Lang aridity index and
topsoil salinity of the populations included in this PC subset.

**Table 5** Priority ecogeographical gaps of targeted *Aegilops* species selected for drought
and salinity tolerance in Spain

**Fig. 5** Location of (a) priority ecogeographical gaps of *Aegilops* in Spain that occur in
sites where the Lang aridity index is < 40, and (b) priority ecogeographical gaps that
occur in sites where the highest salinity values are also found

**Table 6** Geographic description of the Spanish *Aegilops* populations selected as
potentially tolerant to drought and salinity

The inclusion of the predictive characterization subset populations in the genebanks of
the Spanish Network on PGRFA would improve their ecogeographical
representativeness between 7% and 11% (Table 7).

**Table 7** Number of ELC categories of *Aegilops* species currently represented in the
Spanish Network and potential increase (%) in representativeness by collecting
populations of the predictive characterization (PC) subset

**Discussion**

The preprocessing of presence data

The quality of geo-referencing in spatial analysis is very important to obtain reliable
results. Maldonado et al. (2015), who studied the role of natural history collections in
unveiling correct patterns of biodiversity and distribution, concluded that geographic
inaccuracy affects diversity patterns more than taxonomic uncertainties. Similarly,
Graham et al. (2008) evaluated how uncertainty in geo-references and associated
location errors in occurrences influence species distribution modeling and found that
models run with data subject to random location errors resulted in less accurate models
in many species. However, few studies on genetic diversity or taxonomic spatial
distribution describe robust methods to ensure the quality of geo-referenced data. Some
authors, such as Fielder et al. (2015, 2016), excluded records dated from before 1970,
records lacking both coordinates and location descriptions and records with a precision
lower than 4 km². Ramirez-Villegas et al. (2010) carried out a process to verify and
correct the coordinates using BioGeomancer, Google Earth and highly detailed maps.
Khoury et al. (2015) cross-checked the coordinates to country and verified that they
occurred on land. After that, occurrence data were evaluated for correctness with
experts on the target species. In our study, apart from removing accessions and external
sources with low accuracy according to the established criteria, the assessment of the
quality of the georeferenced data allowed us to identify the records with the best quality.
In this sense, the final number of records included in the analysis (considering both
accessions and external sources) ranged between 30% and 53% of the initial number of
records (for *A. geniculata* and *A. biuncialis*, respectively). This selective use of records, which complies with the minimum standards of georeferencing quality, guaranteed that the results of the analysis were as accurate as possible.

**Identification of spatial and ecogeographical gaps**

Spatial bias in collecting activities and chorological studies often affects the spatial distribution of the species, as shown in Maldonado et al. (2015). In our study, the spatial distribution of the accessions could reflect the intensity of collecting activities rather than the real distribution of the species (see Fig. 2a). Some recent projects focused on collecting *Aegilops* germplasm may have contributed to the difference observed in the numbers of accessions between areas. On the other hand, the distribution of the external sources of the targeted species (Fig. 2 B) shows a higher number of populations in some areas of the country, such as the autonomous communities of Navarra and Valencia. However, this may reflect a higher intensity of chorological studies rather than a higher presence of the species in these areas.

Several chorological studies involving the spatial gap analysis of different species have been used as a guide for germplasm collecting. For instance, Maxted et al. (2008) identified *ex situ* conservation gaps in *Aegilops* germplasm collections as regions where the species were predicted to occur according to species distribution models, but previous collection had not taken place. In this study, the priority of germplasm collecting for each of *Aegilops* species was ranked high, medium or low, according to the number of germplasm accessions already conserved *ex situ* and the number of predicted under-sampled regions. Shehadeh et al. (2013) followed this methodology carried out a gap analysis of *Lathyrus* L. species. Another recent methodology for gap analysis was described in Ramírez-Villegas et al. (2010) and applied to wild taxa of the *Phaseolus* genepool. It involves an eight-step process to evaluate conservation deficiencies at three different levels (taxonomic, geographic and environmental) through the calculation of sampling, geographic and environmental representativeness scores. This methodology was also applied in Castañeda-Álvares et al. (2015) to identify *ex situ* conservation priorities for the wild relatives of potato and in Khoury et al. (2015) in their study of the CWR pigeonpea.

The *ex situ* conservation gap analyses in these studies were obtained by overlapping the distribution maps of germplasm accession data and predictive distribution maps generated from the climatic envelope data of the accessions, or of both accessions and external sources.

With regard to identifying spatial gaps, Ramírez-Villegas et al. (2010) assessed the adequacy of geographic coverage of genebank accessions by means of a geographical representativeness score (GRS). This score is the geographic coverage of germplasm collections (modeled using the circular area statistic with a 50 km radius value) divided by the potential distribution coverage of the taxon under analysis. The higher the GRS...
is, the greater the representativeness of genebank collections is in relation to the potential distribution of the taxon. Fielder et al. (2015, 2016) also used this methodology to assess the geographical representativeness in their studies on the conservation of CWR in England and Scotland. They established a threshold of five accessions above which CWR are considered sufficiently represented in ex situ genebanks. In these geographic coverage assessments, the use of the potential distribution coverage of a taxon can lead to low GRS values due to the presence of false positives in the model (i.e., predicted locations with no real population occurrence) even though they have a good spatial representation in the ex-situ collections. In our study, we adopted a more conservative approach based only on recorded occurrences to avoid this problem.

With regard to identifying ecogeographical gaps, the exclusive use of species distribution models to develop strategies for germplasm collecting may also lead to the over-representation of some adaptive scenarios because these models guide collectors to the species’ most preferred habitats (Parra-Quijano et al., 2012b).

On the other hand, predictive species distribution models based on bioclimatic information only offer a partial view of plant abiotic adaptation. In our study, we generated ecogeographical land characterization maps using the values of bioclimatic, geophysical and edaphic variables at species occurrence sites to describe as accurately as possible the different adaptive scenarios of the species. The representativeness analysis, based on the comparison of the ELC categories of the accessions and the external sources considered spatial gaps, helped us to select priority collecting sites, avoid over-representation and identify populations from low and non-represented ecogeographical categories. This allowed us to identify populations in marginal environments in the species’ range, where interesting traits related to abiotic stress tolerance may be found. The easy-to-use tools employed in this study (i.e. CAPFITOGEN tools, Parra-Quijano et al., 2015) allow genebank curators and technicians in charge of collecting activities to develop their own germplasm collecting design based on spatial and ecogeographical analyses.

When trying to improve the ecogeographical representativeness of an ex situ germplasm collection, we would expect that the lower the number of accessions of a species in a genebank is, the higher the probability of improving its ecogeographical representativeness by collecting seeds from new populations. However, this may not be the case. For instance, A. biuncialis, the target species with the lowest number of accessions in this study, is also the species with the lowest potential percentage of representativeness improvement. This is probably because the existing accessions are a good representation of the few ecogeographical categories where the species occurs.

Including populations identified as priority ecogeographical gaps in the genebank collections of the Spanish Network on PGRFA would qualitatively improve ecogeographical representativeness, with increases in the percentages of ELC categories represented in the Network between 27% and 70%. Such increases are higher than those
obtained in the *Lupinus* species collecting activities described in Parra-Quijano et al. (2012b), which ranged between 7% and 11%. However, it should be noted that the representativeness of *Lupinus* species in the reference germplasm collection before the optimized collecting activities were carried out, was higher than in the case of the targeted *Aegilops* species.

Selection of priority collecting sites for traits of tolerance to drought and salinity

The search for the 20% of the populations adapted to arid environments and with the highest topsoil salinity values allowed us to identify populations that occur in sites in the western Mediterranean distribution of the targeted species with the highest salinity values, according to the work on saline and sodic soils in the European Union by Tóth et al. (2008).

One of the results of this study is the identification of 45 *Aegilops* populations of high interest due to their potential tolerance to drought and salinity, in addition to being ecogeographical gaps in the existing Spanish germplasm collections. Although predictive characterization is an inexpensive and effective approach to maximizing the likelihood of capturing a desirable level of trait expression among accessions of landraces and traditional varieties, this is the first time that predictive characterization has been applied to identifying wheat wild relatives with potential tolerance to drought and salinity. Interestingly, the validity of predictive characterization is presumably greater when used with wild relatives rather than when used with landraces, because wild relatives are more greatly affected by natural selection (no artificial selection). Thus, the link between existing environmental conditions and genotypes with local genetic adaptation is likely to be stronger than the link between environmental conditions and landraces or modern varieties, which is explored by FIGS.

Until now, the search for tolerance to drought and salinity in wheat has been carried out using approaches such as those in Colmer et al. (2006); Molnár et al. (2004); Farooq (2002); and Xing et al. (1993). FIGS approaches in cultivated wheat have focused on resistance to biotic factors such as plagues and diseases (Endresen et al., 2011, 2012; El Bouhissini et al., 2009, 2011; Bhullar et al., 2009). Nevertheless, in other crops FIGS has been successfully applied in the search for abiotic stress tolerance. For example, Khazaiei et al. (2013) tested the effectiveness of FIGS to search for traits related to drought adaptation in a large faba bean (*Vicia faba* L.) collection.

According to predictive characterization methods (Thormann et al., 2016), the probability of capturing phenotypes tolerant to drought and salinity in the predictive characterization subset would be higher than in a randomly chosen set. The 45 priority populations identified by the predictive characterization approach in this study are now considered priority collection populations for further optimized collecting activities of *Aegilops* germplasm. Furthermore, these populations will be assessed for their tolerance to drought and salinity in order to validate the methodology applied.
Conclusions

This study aimed to identify spatial and ecogeographical gaps in the Spanish germplasm collection of *Aegilops* and priority populations of potential interest due to their possible tolerance to drought and salinity. The methodology employed allowed us to establish an optimized collecting strategy by filtering potential collecting sites, thereby avoiding over-representation and identifying populations from low and non-represented ecogeographical categories. It also provided a subset of 45 populations of potential interest in terms of tolerance to drought and salinity. Subsequent collections and evaluations of these populations will provide essential feedback on the efficacy of these approaches to improving the genetic representativity of genebank collections and identify genotypes with desired traits.

Acknowledgments

We would like to thank the personnel at CRF-INIA. We are also grateful to Lori De Hond for her linguistic assistance.

References

- Graham CH, Elith J, Hijmans RJ, Guisan A, Peterson AT, Loisells BA, The
  Nceas Predicting Species Distributions Working Group. 2008. Journal of
  characteristics relating to ionic relations in some salt tolerant grasses from the
- Harmonized world soil (HWS) database (version 12). 2012. Available at
  http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-
  database/HTML/). FAO/IASA/ISRIC/ISSCAS/JRC, FAO, Rome, Italy and
  IIASA, Laxenburg, Austria. (accessed 07 January 2016).
  sustainable use of crop wild relatives. Agriculture, Ecosystems and Environment
  121: 245-255.
  Poblaciones en peligro: viabilidad demográfica de la flora vascular amenazada
  de España. Dirección General de Medio Natural y Política Forestal (Ministerio
  de Medio Ambiente, y Medio Rural y Marino), Madrid.
- Ketchen DJ, Shook CL. 1996. The application of cluster analysis in strategic
  management research: an analysis and critique. Strategic Management Journal
  17 (6):441-458.
- Khazaei H, Strett K, Bari A, Mackay M, Stoddard F. 2013. The FIGS (Focused
  Identification of Germplasm Strategy (FIGS) approach identifies traits related to
  drought adaptation in Vicia faba genetic resources. Plos One, October, Volume
  8, Issue 3.
- Khoury CK, Castañeda-Álvarez NP, Achicanoy HS, Sosa CC, Bernau V, Kassa
  MT, Norton SL., van der Maesen JG, Upadhyaya HD, Ramirez-Villegas J, Jarvis
  A, Struik PC. 2015. Crop wild relatives of pigeonpea (Cajanus cajan (L.)
  Millsp.): distributions, ex situ conservation status, and potential genetic
  resources for abiotic stress tolerance. Biological Conservation 184:259-270.
  FIGS. In: Black CK, Panozzo JF, Rebetzke GJ, ed. Cereals 2004, Proceedings of
  the 54th Australian Cereal Chemistry Conference and the 11th Wheat Breeders’
  Assembly Cereal Chemistry Division. Royal Australian Chemical Institute,
  Melbourne, 138-141.
  E, Ronsted N, Antonelli A. 2015. Estimating species diversity and distribution in
  the era of Big Data: to what extent can we trust public databases? Global
  Ecology and Biogeography. doi:101111/geb12326.
  conservation strategy for Aegilops species. Plant Genetic Resources:
  Characterization and Utilization 6 (2):126-141.
- Molnár I, Gáspár L Sárvari E, Dulai S, Hoffmann B, Molnár-Láng M, Galiba G.
  2004. Physiological and morphological responses to water stress in Aegilops
biuncialis and Triticum aestivum genotypes with differing tolerance to drought.

Functional Plan Biology 31:1149-1159.


