

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 geo-referencing 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.

Eliminado:

Eliminado: A.

Eliminado:

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).

Eliminado:

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).

Eliminado: our

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. striiformis*) 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).

Eliminado:

Comentario [u1]: Please change ll.

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).

Eliminado: s

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

1 *situ* conservation gaps were identified as regions where the species was predicted to
2 occur but seed accessions had not been previously collected, or, alternatively, the
3 species was under-sampled.

4 Optimized Collecting Design (OCD) is a technique described by Parra Quijano et al.
5 (2012b) that involves the identification of ecogeographical gaps for a target species in a
6 target *ex situ* genebank. Based on the premise that genetic adaptation is achieved
7 through natural selection acting upon particular limiting environmental conditions, this
8 technique aims to improve the genetic representativeness of genebank accessions by
9 improving their ecogeographical representativeness. Like other gap analysis techniques,
10 OCD compares the collecting locations of the target species accessions currently held in
11 the genebanks and the species' occurrence data from external sources (spatial gaps). It
12 then uses ELC maps to detect adaptive scenarios not represented in the target *ex situ*
13 genebank. Using ELC maps to design collecting strategies can help to include
14 accessions from marginal or under-represented environments that may contain
15 important traits related to adaptations to biotic and abiotic stress.

16 Another interesting issue for genebank managers and stakeholders is the identification
17 of genotypes that can have a specific use in plant breeding.

18 The Focused Identification of Germplasm Strategy (FIGS) is a useful approach for
19 screening large germplasm collections to identify sets of accessions with a high
20 probability of containing specific target traits based on the ecogeographical information
21 of the sites where the populations were collected (Mackay & Street, 2004). If we know
22 where a set of *ex situ* accessions has evolved, or at least where they have grown for a
23 period long enough for adapted genotypes to have been selected, we can establish
24 relations or patterns between the environmental conditions of the site and the presence
25 or absence of the target trait. We can then make predictions on non-evaluated
26 germplasm (Mackay & Street, 2004).

Eliminado:

27 In recent years, FIGS has been successfully used to identify sources of resistance to
28 sunn pest in Syria (El Bouhssini et al., 2009) and to Russian wheat aphid in bread wheat
29 (El Bouhssini et al., 2011). FIGS has also been used to identify traits related to abiotic
30 stresses, such as drought adaptation in *Vicia faba* L. (Khazaei et al., 2013). These
31 studies selected accessions from an ecogeographically-characterized collection that
32 complied with certain values or ranges for the characterized variables, set by the
33 researchers based on their knowledge of the species. Other FIGS approaches can be
34 used to identify accessions of potential interest, when the trait under consideration
35 cannot be directly related to an ecogeographical variable. In this case, partial evaluation
36 data from the target collection are required to detect a reliable relationship between the
37 ecogeographic variables and the trait of interest. This approach has been successfully
38 applied by Thormann et al. (2016) and Endresen et al. (2012), who identified sources of
39 resistance to stem rust in bread and durum wheat, and by Bari et al. (2011, 2014), who
40 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.

Eliminado: the territory of

Eliminado:

Eliminado: for the national genebank of Spain

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://wwwx.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://wwwx.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-Bioversity, 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 geophysics variables (4) (see Table S1).

Eliminado: 3

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. Bivariate 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 geophytic variables were considered in generating the ELC map for each species. The variables latitude and longitude were included as two additional geophytic 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.

Con formato: Fuente: 12 pto

Eliminado: –

Con formato: Fuente: 12 pto

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).

Comentario [u2]: Would be more logical consider a gap as a point on which there is no accessions. I would understand something like: For each species, location distant more than 1 km apart from accessions were considered spatial gaps.

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 (AI_L) and topsoil salinity. The Lang aridity index was calculated as:

$$AI_L = \text{Annual precipitation} / (\text{Annual mean temperature})$$

Following the ecogeographical filtering technique of predictive characterization (Thormann et al., 2016), we selected populations occurring in sites with $AI_L < 40$. 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 1 Number of *Aegilops* 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

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

Comentario [u3]: From where does this data come?

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, geophytic and edaphic variables) are shown in Table 2. The variables isothermality and altitude were selected in four of the five studied species.

Comentario [u4]: This is M&M

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*.

Comentario [u5]: Why not a single one, see my comment

Comentario [u6]: opinión

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

Comentario [u7]: if there are population why you identify as a gap? Se Comment u2

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. Ramírez-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

Eliminado: l

Eliminado: ran

Comentario [u8]: so it is to avoid the random location

Comentario [u9]: best comparing with what. Probably "higher quality" would be better.

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-Álvarez 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

Comentario [u10]: Have you references about?

Comentario [u11]: Clarify, which Project with which aims.

Comentario [u12]: Seems to me on the same hand rather than an other.

Eliminado: ¶

Eliminado: to

Eliminado: y

1 is, the greater the representativeness of genebank collections is in relation to the
2 potential distribution of the taxon. Fielder et al. (2015, 2016) also used this
3 methodology to assess the geographical representativeness in their studies on the
4 conservation of CWR in England and Scotland. They established a threshold of five
5 accessions above which CWR are considered sufficiently represented in *ex situ*
6 genebanks. In these geographic coverage assessments, the use of the potential
7 distribution coverage of a taxon can lead to low GRS values due to the presence of false
8 positives in the model (i.e., predicted locations with no real population occurrence) even
9 though they have a good spatial representation in the *ex-situ* collections. In our study,
10 we adopted a more conservative approach based only on recorded occurrences to avoid
11 this problem.

Eliminado: ,

Comentario [u13]: But in your case you did not identify any CWR.

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

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

Comentario [u14]: It seems not in accord with the results reported see my comments on Figs 4 and 5.

Eliminado: s

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

Eliminado:

Eliminado:

Comentario [u15]: How do you know where the species occurs? if, you mentioned that the gaps in location is due to the gaps in collection rather than the gaps in species presence (see rows 6-9 page 12)

38 Including populations identified as priority ecogeographical gaps in the genebank
39 collections of the Spanish Network on PGRFA would qualitatively improve
40 ecogeographical representativeness, with increases in the percentages of ELC categories
41 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.

Eliminado: I

Eliminado: however

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.

Comentario [u16]: Conversely, wild species are not adapted to agricultural practices.

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 Bouhssini 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, Khazaei 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.

Eliminado: 2; Endresen et al

Eliminado: .

Eliminado: 1

Eliminado: El Bouhissini et al., 2009;

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

- Abrol IP, Yadav JSP, Massoud FI. 1988. Salt-affected soils and their management. FAO Soils Bulletin 39, FAO, Rome. Available at <http://www.fao.org/docrep/x5871e/x5871e03.htm> (accessed 06 June 2016).
- Al-maskri A, Hameed M, Ashraf M, Khan M, Fatima S, Nawaz T, Batool R. 2014. Structural features of some wheat (*Triticum* spp.) landraces/cultivars under drought and salt stress. *Arid Land Research and Management* 28:355-370.
- Araus JL, Slafer GA, Reynolds MP, Royo C. 2002. Plant breeding and drought in C₃ cereals: what should we breed for? *Annals of Botany* 89:925-940.
- Bari A, Amri, A Street K, Mackay M, De Pauw E, Sanders R, Nazari K, Humeid B, Konopka J, Alo F. 2014. Predicting resistance to stripe (yellow) rust (*Puccinia striiformis*) in wheat genetic resources using focused identification of germplasm strategy. *Journal of Agricultural Science* 152:906-916.
- Bari A, Street K, Mackay M, Endresen DTF, De Pauw E, Amri A. 2011. Focused identification of germplasm strategu (FIGS) detects wheat stem rust resistance linked to environmental variables. *Genetic Resources and Crop Evolution*. doi:10.1007/s10722-011-9775-5.
- Bhullar NK, Street K, Mackay M, Yahiaoui N, Keller B. 2009. Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. *Proceedings of the National Academy of Sciences* 106:9519-9524.

- 1 - Castañeda-Álvarez NP, de Haan S, Juárez H, Khoury CK, Achicanoy HA, Sosa
2 CC, Vernay V, Salas A, Heider B, Simon R, Maxted N, Spooner DM. 2015. Ex
3 situ conservation priorities for the wild relatives of potato (*Solanum* L. section
4 *petota*). Plos One. doi:101371/journal.pone.0122599.
- 5 - Colmer TD, Flowers TJ, Munns R. 2006. Use of wild relatives to improve salt
6 tolerance in wheat. Journal of Experimental Botany 57 (5):1059-1078.
- 7 - Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ.
8 2007. Random Forests for classification in ecology. Ecology 88 (11):2783-2792.
- 9 - El Bouhssini M, Street K, Amri A, Mackay M, Ogbonnaya FC, Omran A,
10 Abdalla O, Baum M, Dabbous A, Rihawi F. 2011. Sources of resistance in bread
11 wheat to Russian wheat aphid (*Diuraphis noxia*) in Syria identified using the
12 Focused Identification of Germplasm Strategy (FIGS). Plant Breeding 130:96-
13 97.
- 14 - El Bouhssini M, Street K, Joubi A, Ibrahim Z, Rihawi F. 2009. Sources of wheat
15 resistance to Sunn pest, *Eurygaster integriceps* Puton, in Syria. Genetic
16 Resources and Crop Evolution 56:1065-1069.
- 17 - Endresen DTF, Street K, Mackay M, Bari A, Amri A, De Pauw E, Nazari K,
18 Yahyaoui A. 2012. Sources of resistance to stem rust (Ug99) in bread wheat and
19 durum wheat identified using focused identification of germplasm strategy. Crop
20 Science 52:764-773.
- 21 - Endresen DTP. 2011. Predictive association between biotic stress traits and eco-
22 geographic data for wheat and barley landraces. Crops Science 51:2036-2055.
- 23 - FAO/Bioversity Multi-crop Passport Descriptors V.2. 2012. Available at
24 [http://www.bioversityinternational.org/fileadmin/user_upload/online_library/publications/pdfs/FAO-](http://www.bioversityinternational.org/fileadmin/user_upload/online_library/publications/pdfs/FAO-Bioversity_multi_crop_passport_descriptors_V_2_Final_rev_1526.pdf)
25 [Bioversity_multi_crop_passport_descriptors_V_2_Final_rev_1526.pdf](http://www.bioversityinternational.org/fileadmin/user_upload/online_library/publications/pdfs/FAO-Bioversity_multi_crop_passport_descriptors_V_2_Final_rev_1526.pdf)
26 (accessed 30 June 2016).
- 27 - FAO. 2013a. FAOSTAT. Available at <http://faostat3.fao.org/home/E> (07
28 January 2016).
- 29 - FAO. 2013b. Drought. Available at
30 <http://www.fao.org/docrep/017/aq191e/aq191e.pdf> (accessed 07 January 2016).
- 31 - Farroq S. 2002. *A. ovata*: a potential gene source for improvement of salt
32 tolerance of wheat. In: Ahmad R, Malik KA, ed. Prospects for saline agriculture,
33 vol 37. Dordrecht: Kluwer Academic Publishers, 123-130.
- 34 - Fielder H, Brotherton P, Hosking J, Hopkins JJ, Ford-Lloyd B, Maxted N. 2015.
35 Enhancing the conservation of crop wild relatives in England. PLoS ONE 10(6):
36 e0130804. doi:10.1371/journal.pone.0130804.
- 37 - Fielder H, Sith C, Ford-Lloyd B, Maxted N. 2016. Enhancing the conservation
38 of crop wild relatives in Scotland. Journal for Nature Conservation 29:51-61.
- 39 - Fita A, Rodríguez-Burruezo A, Boscaiu M, Prohens J, Vicente O. 2015.
40 Breeding and domesticating crops adapted to drought and salinity: a new
41 paradigm for increasing food production. Frontiers in Plant Science volumen 6
42 article 978.
- 43 - Glick BR, Cheng Z, Czarny J, Duan J. 2007. Promotion of plant growth by ACC
44 deaminase-producing soil bacteria. European Journal of Plant Pathology
45 119:329-339.
- 46

- 1 - Graham CH, Elith J, Hijmans RJ, Guisan A, Peterson AT, Loisells BA, The
2 Nceas Predicting Species Distributions Working Group. 2008. Journal of
3 Applied Ecology 45:239-247.
- 4 - Hameed M, Ashraf M, Naz N. 2011. Anatomical and physiological
5 characteristics relating to ionic relations in some salt tolerant grasses from the
6 Salt Range. Pakistan Acta Physiologiae Plantarum 33:1399-1409.
- 7 - Harmonized world soil (HWS) database (version 12). 2012. Available at
8 [http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-](http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/)
9 [database/HTML/](http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/)). FAO/IIASA/ISRIC/ISSCAS/JRC. FAO, Rome, Italy and
10 IIASA, Laxenburg, Austria. (accessed 07 January 2016).
- 11 - Heywood V, Casas A, Ford-Lloyd B, Kell S, Maxted N. 2007. Conservation and
12 sustainable use of crop wild relatives. Agriculture, Ecosystems and Environment
13 121: 245-255.
- 14 - Iriondo JM, Albert MJ, Giménez L, Domínguez F, Escudero A. 2009.
15 Poblaciones en peligro: viabilidad demográfica de la flora vascular amenazada
16 de España. Dirección General de Medio Natural y Política Forestal (Ministerio
17 de Medio Ambiente, y Medio Rural y Marino), Madrid.
- 18 - Ketchen DJ, Shook CL. 1996. The application of cluster analysis in strategic
19 management research: an analysis and critique. Strategic Management Journal
20 17 (6):441-458.
- 21 - Khazaei H, Strett K, Bari A, Mackay M, Stoddard F. 2013. The FIGS (Focused
22 Identification of Germplasm Strategy (FIGS) approach identifies traits related to
23 drought adaptation in *Vicia faba* genetic resources. Plos One, October, Volume
24 8, Issue 3.
- 25 - Khoury CK, Castañeda-Álvarez NP, Achicanoy HS, Sosa CC, Bernau V, Kassa
26 MT, Norton SL, van der Maesen JG, Upadhyaya HD, Ramírez-Villegas J, Jarvis
27 A, Struik PC. 2015. Crop wild relatives of pigeonpea (*Cajanus cajan* (L.)
28 Millsp.): distributions, ex situ conservation status, and potential genetic
29 resources for abiotic stress tolerance. Biological Conservation 184:259-270.
- 30 - Lambers H, Chapin FS, Pons TL. 2008. Plant Physiological Ecology, 2nd Ed.
31 Springer, New York.
- 32 - Mackay MC, Street K. 2004. Focused identification of germplasm strategy –
33 FIGS. In: Black CK, Panozzo JF, Rebetzke GJ, ed. Cereals 2004, Proceedings of
34 the 54th Australian Cereal Chemistry Conference and the 11th Wheat Breeders’
35 Assembly Cereal Chemistry Division. Royal Australian Chemical Institute,
36 Melbourne, 138-141.
- 37 - Maldonado C, Molina CI, Zizka A, Persson C, Taylor CM, Albán J, Chilquillo
38 E, Ronsted N, Antonelli A. 2015. Estimating species diversity and distribution in
39 the era of Big Data: to what extent can we trust public databases? Global
40 Ecology and Biogeography. doi:101111/geb12326.
- 41 - Maxted N, White K, Valkoun J, Konopka J, Hargreaves S. 2008. Towards a
42 conservation strategy for *Aegilops* species. Plant Genetic Resources:
43 Characterization and Utilization 6 (2):126-141.
- 44 - Mólnar I, Gáspár L Sárvári E, Dulai S, Hoffmann B, Molnár-Láng M, Galiba G.
45 2004. Physiological and morphological responses to water stress in *Aegilops*

biuncialis and *Triticum aestivum* genotypes with differing tolerance to drought. Functional Plant Biology 31:1149-1159.

- [Mondini, L., Nachit, M. M., Pagnotta, M. A. 2015. Allelic variants in durum wheat \(*Triticum turgidum* L. var. *durum*\) DREB genes conferring tolerance to abiotic stresses. Molecular Genetics and Genomics 290: 531-544.](#)
- Mujeeb-Kazi A, Diaz de Leon JL. 2002. Conventional and alien genetic diversity for salt tolerant wheats: focus on current status and new germplasm development. In: Ahmad R, Malik KA, ed. Prospects for saline agriculture 37. Dordrecht: Kluwer Academic Publishers, 69-82.
- Noorka IR, Pat Heslop-Harrison JS. 2014. Water and Crops: Molecular biologists, physiologists, and plant breeders approach in the context of evergreen revolution. In: Taylor, F, ed. Hand book of Plant and Crop Physiology. CRC Press, 967-978.
- Parra-Quijano M, Draper D, Torres E, Iriondo JM. 2008. Ecogeographical representativeness in crop wild relative ex-situ collections. In: Maxted N, Ford-Lloyd BV, Kell SP, Iriondo JM, Dulloo ME, Turok J, ed. Crop wild relative conservation and use. CAB International, Wallingford, 249-273.
- Parra-Quijano M, Iriondo JM, Torres E. 2012a. Ecogeographical land characterization maps as a tool for assessing plant adaptation and their implications in agrobiodiversity studies. Genetic Resources and Crop Evolution 59:205-217.
- Parra-Quijano M, Iriondo JM, Torres ME. 2012b. Improving representativeness of genebank collections through species distribution models, gap analysis and ecogeographical maps. Biodiversity and Conservation 21:79-96.
- Parra-Quijano M, Torres E, Iriondo JM, López F. 2015. CAPFITOGEN Tools. User Manual version 2.0. International Treaty on Plant Genetic Resources for Food and Agriculture, FAO, Rome.
- Ramírez-Villegas J, Khoury C, Jarvis A, Debouck DG, Guarino L. 2010. A gap analysis methodology for collecting crop genepools: a case study with *Phaseolus* beans. PLoS ONE 5(10): e13497. doi:10.1371/journal.pone.0013497.
- Rubio Teso ML, Iriondo JM, Parra-Quijano M, Torres E. 2013. National Strategy for the Conservation of Crop Wild Relatives of Spain PGR Secure. Available at http://www.pgrsecure.bham.ac.uk/sites/default/files/documents/public/National_CWR_Conservation_Strategy_Spain.pdf (accessed 30 June 2016).
- Shehadeh A, Amri A, Maxted N. 2014. Ecogeographic survey and gap analysis of *Lathyrus* L. species. Genetic Resources and Crop Evolution 60:2101-2113.
- Jarvis A, Reuter HI, Nelson A, Guevara E. 2008. Digital Elevation Models (DEM) of the Shuttle Radar Topography Mission (SRTM). Hole-filled SRTM for the globe Version 4. Available from the CGIAR-CSI SRTM 90m Database at <http://srtm.csi.cgiar.org/> (accessed 07 January 2016).
- Thormann I, Parra-Quijano M, Rubio Teso ML, Endresen DTF, Dias S, Iriondo JM, Maxted N. 2016. Predictive characterization methods for accessing and using CWR diversity. In: Maxted N, Dulloo ME, Ford-Lloyd BV, ed. Capturing wild relative and landrace diversity for crop improvement, 64-77.

- 1 - Tóth G, Montanarella L, Rusco E. 2008. Updated map of salt affected soils in
2 the European Union. In: Tóth G, Montanarella L, Rusco E, ed. Threats to soil
3 quality in Europe. EUR 23438 – Scientific and Technical Research series
4 Luxembourg: Office for Official Publications of the European Communities, 61-
5 74.
- 6 - USDA, ARS, National Genetic Resources Program Germplasm Resources
7 Information Network – (GRIN). National Germplasm Resources Laboratory,
8 Beltsville, Maryland. [http://wwwars-gringov4/cgi-](http://wwwars-gringov4/cgi-bin/npgs/html/queriespl?language=es)
9 [bin/npgs/html/queriespl?language=es](http://wwwars-gringov4/cgi-bin/npgs/html/queriespl?language=es) (accessed 25 June 2015)
- 10 - Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high
11 resolution interpolated climate surfaces for global land areas International
12 Journal of Climatology 25:1965-1978 (<http://www.worldclim.org>).
- 13 - Xing X, Monneveux P, Damania AB, Zarahieva M. 1993. FAO/IPGRI Plant
14 Genetic Resources Newsletter 96:11-1.