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

Rosa M Garcia Corresp., 1 , **Mauricio Parra-Quijano** 2 , **Jose M Iriondo** 3

1 Centro Nacional de Recursos Fitogenéticos (CRF-INIA), Alcalá de Henares, Madrid, Spain

2 Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá D.C., Colombia

3 Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain

Corresponding Author: Rosa M Garcia
Email address: rosamaria.garcia@inia.es

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 ecogeographical variables that might be important 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.

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

3 **Rosa María GARCÍA[1], Mauricio PARRA-QUIJANO[2], Jose María IRIONDO[3]**
4 [1]Centro Nacional de Recursos Fitogenéticos (CRF-INIA), Alcalá de Henares, Madrid, Spain
5 [2]Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá D.C., Colombia
6 [3]Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain.
7
8 Corresponding Author:
9
10 Rosa María García[1]
11 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 ecogeographical variables that might be important 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.

49

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

67 land surface. These saline soils have never been cultivated because present major crops are salt-
68 sensitive (Fita et al., 2015).

69 The species of genus *Aegilops* have been widely used as gene sources in wheat breeding,
70 especially in providing resistance to pests and diseases such as leaf, stem and stripe rusts
71 (*Puccinia recondita*, *P. graminis* and *P. stiiformis*) or hessian fly (*Mayetiola destructor*).
72 Numerous studies have searched for drought and salt stress tolerant genotypes. For instance,
73 Xing et al. (1993) studied the potential of some *Aegilops* species, including *Ae. ventricosa*
74 Tausch, as gene donors in breeding for salt tolerance. Subsequent studies proposed *Ae. ovata* L.
75 as a source of salt tolerance in wheat (Farooq, 2002) and determined that *Ae. ovata* and *Ae.*
76 *biuncialis* Vis. have wide genetic variation for salt tolerance (Colmer et al., 2006). Molnár et al.
77 (2004) compared the physiological and morphological responses to water stress in *Ae. biuncialis*
78 and *Triticum aestivum* L. genotypes, and concluded that *Ae. biuncialis* genotypes from dry
79 habitats have greater drought tolerance than wheat, making them good candidates for improving
80 drought tolerance in this crop. Mondini et al. (2014) identified SNPs variants conferring salt
81 tolerance in durum wheat. Due to their present and potential use as gene donors in wheat
82 breeding, *Aegilops* species can be considered crop wild relatives (CWR) of wheat (Heywood et
83 al., 2007). Five of these species (*Ae. biuncialis*, *Ae. geniculata* Roth, *Ae. neglecta* Req. ex
84 Bertol., *Ae. triuncialis* L. and *Ae. ventricosa*) are included in the Spanish National Inventory of
85 CWRs (Rubio et al., 2013).

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

91 Parra-Quijano et al. (2008) developed an ELC map for Peninsular Spain and the Balearic Islands,
92 using different sources of ecogeographical information. The ability of the ELC map to
93 discriminate different areas with different adaptive pressures was tested with eight crop and
94 CWR species. They found that the ELC map had an effective discriminatory capacity to
95 delineate adaptive scenarios. Because the efficacy in detecting plant adaptation was
96 heterogeneous for the species analyzed, it was concluded that the application of ecogeographical
97 maps for detecting plant adaptation may be better approached through an ad-hoc ELC map
98 generation for each target species. The ELC concept is detailed in depth in Parra-Quijano et al.
99 (2012a), and several applications related to the collection, conservation and efficient use of plant
100 genetic resources have been developed (e.g., Parra-Quijano et al., 2012a; Parra-Quijano et al.,
101 2012b; Thormann et al., 2016).

102 Gap analysis has been widely applied for conservation purposes (Maxted et al., 2008). For
103 instance, Ramírez-Villegas et al. (2010) applied a gap analysis methodology to collect the crop
104 genepool of *Phaseolus* beans and evaluated conservation deficiencies at three different levels

105   (taxonomic, geographic and environmental). Khoury et al. (2010) reviewed global crop and
106   regional conservation strategies and recognized the importance of filling gaps in genebanks of
107   plant genetic resources. Recently, Shehadeh et al. (2014) carried out a gap analysis of *Lathyrus*
108   L. species. In this study, predictive distribution maps for each *Lathyrus* taxon were produced
109   based on climatic data, and *ex situ* conservation gaps were identified as regions where the species
110   was predicted to occur but seed accessions had not been previously collected, or, alternatively,
111   the species was under-sampled.

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

123   Another interesting issue for genebank managers and stakeholders is the identification of
124   genotypes that can have a specific use in plant breeding. The Focused Identification of
125   Germplasm Strategy (FIGS) is a useful approach for screening large germplasm collections to
126   identify sets of accessions with a high probability of containing specific target traits based on the
127   ecogeographical information of the sites where the populations were collected (Mackay & Street,
128   2004). If we know where a set of *ex situ* accessions has evolved, or at least where they have
129   grown for a period long enough for adapted genotypes to have been selected, we can establish
130   relations or patterns between the environmental conditions of the site and the presence or
131   absence of the target trait. We can then make predictions on non-evaluated germplasm (Mackay
132   & Street, 2004).

133   In recent years, FIGS has been successfully used to identify sources of resistance to sunn pest in
134   wheat in Syria (El Bouhssini et al., 2009) and to Russian wheat aphid in bread wheat (El
135   Bouhssini et al., 2011). FIGS has also been used to identify traits related to abiotic stresses, such
136   as drought adaptation in *Vicia faba* L. (Khazaei et al., 2013). These studies selected accessions
137   from an ecogeographically-characterized collection that complied with certain values or ranges
138   for the characterized variables, set by the researchers based on their knowledge of the species.
139   Other FIGS approaches can be used to identify accessions of potential interest, when the trait
140   under consideration cannot be directly related to an ecogeographical variable. In this case, partial
141   evaluation data from the target collection are required to detect a reliable relationship between
142   the ecogeographical variables and the trait of interest. This approach has been successfully
143   applied by Thormann et al. (2016) and Endresen et al. (2012), who identified sources of

144 resistance to stem rust in bread and durum wheat, and by Bari et al. (2011, 2014), who predicted
145 resistance to stem rust and stripe rust in accessions of wheat landraces.

146 A broader concept has recently been developed to identify germplasm with a high probability of
147 containing specific target traits, named "predictive characterization". This term comprises a set
148 of approaches that use geographic and environmental data to search for particular traits in a
149 usually large set of populations, not only accessions and landraces but also CWR (Thormann et
150 al., 2014). This search can be carried out by means of the ecogeographical filtering method or the
151 calibration method (Thormann et al., 2014).

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

161

Materials & Methods

Target species

164 The targeted species were *Ae. biuncialis* Vis., *Ae. geniculata* Roth, *Ae. neglecta* Req. ex Bertol.,
165 *Ae. triuncialis* L. y *Ae. ventricosa* Tausch. Species names were standardized using van Slageren
166 (1994). Regarding their distribution in Spain, all of them are found in the Mediterranean
167 biogeographical region. *Ae. biuncialis* is also found in the Macaronesian region, whereas *Ae.*
168 *geniculata* can also be found in the Atlantic region. Table 1 provides a summary of information
169 on distribution, autecology and mating system of each species.

170 **Table 1** Autecology and mating system of the targeted *Aegilops* species. Information obtained
171 from Van Slageren (1994) unless otherwise stated

172

Species datasets

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

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

186   All accessions and external sources without geographic coordinates were removed. We also
187   removed external sources with geographic coordinates expressed in decimal degrees with less than
188   two decimals in both latitude and longitude or without textual description on the occurrence site,
189   and external sources with geographic coordinates expressed in UTM with lower resolution than
190   1x1 km. Passport and presence data were standardized to CAPFITOGEN data formats which is
191   basically the Multi-Crop Passport Descriptor (FAO-Bioversity, 2012) plus four additional
192   administrative fields for collecting or presence site description (Parra-Quijano et al., 2015).

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

197   Accessions and external sources free of spatial duplicates were subjected to a geo-referencing
198   quality evaluation using GEOQUAL from the CAPFITOGEN toolkit (Parra-Quijano et al.,
199   2015). We set the quality threshold in TOTALQUAL100 = 80, so only records with quality
200   values above this threshold were considered in subsequent analyses.

201

202   **Selection of ecogeographical variables**

203   Ecogeographical information was extracted for each occurrence site from raster layers with a 30
204   arc-second resolution and classified into three ecogeographical components: bioclimatic
205   variables (37), edaphic variables (16) and geophysic variables (4) (see Supplemental Table 1).

206   The variables that might be the most relevant for each species in each ecogeographical
207   component were then identified using the SelecVar tool from CAPFITOGEN (Parra-Quijano et
208   al., 2015). SelecVar extracts information from the ecogeographical variables (layers) to the
209   occurrence sites and assesses the importance of each variable in generating different adaptive
210   scenarios for a species (Parra-Quijano et al., 2015). It estimates variable importance according to
211   the random forest classification (RFC) and detects redundant variables through bivariate
212   correlation analysis. The RFC analysis provides a ranking of the most important variables for
213   establishing ecogeographical categories, placing variables with a higher mean decrease in

214 accuracy in the first positions (Cutler et al., 2007). Rankings were obtained for each
215 ecogeographical component for each species. Bivariated correlation analysis detected correlated
216 variables in the top fifteen variables of the RFC ranking. Variables with Pearson correlation
217 coefficient > |0.50| and p-value < 0.05 in the same ecogeographical component were identified
218 and removed.

Generation of the ELC maps

220 An ecogeographical land characterizacion map or ELC map is a representation of the different
221 adaptive scenarios of a species (Parra-Quijano et al., 2012a). In this study, we generated an ad-
222 hoc ELC map for each target taxa assuming that each taxa may respond differently to the
223 environment. Thus, we considered that this approach would provide better results than those
224 obtained from the generation of a single ELC map for all the taxa.

225 The top three bioclimatic, three edaphic and two geophysic variables in the rankings of variable
226 importance (i.e., the ecogeographical variables suggested as relevant in the species distribution)
227 were considered in generating the ELC map for each species. The variables latitude and
228 longitude were included as two additional geophysic variables to obtain maps with spatially
229 aggregated categories. The "elbow" method was used to create the ecogeographical categories.
230 This is a simple system which uses K means as a clustering algorithm where the cut-off point is
231 determined on the basis of the decrease in the sum of the intra-group squares (Ketchen & Shook,
232 1996). The optimal number of categories is reached when the decrease in the intra-group sum of
233 squares in a range of n and n+1 groups is less than 50%. ELC maps were generated for each
234 species using the ELCmapas tool of the CAPFITOGEN toolkit (Parra-Quijano et al., 2015) and
235 the following parameters: 30-arc-second cell size resolution and 8 clusters as the maximum
236 number of categories allowed per ecogeographical component (bioclimatic, edaphic and
237 geophysic).

238

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

241 Spatial gaps were identified based on occurrence site coordinates. For each species, locations
242 cited by external sources more than 1 km apart from accessions were considered spatial gaps. An
243 ecogeographical representativeness analysis of the existing germplasm collections was then
244 carried out to identify ecogeographical gaps (ELC map categories not represented in *ex situ*
245 genebanks where the species occurs).

246 Areas with a high occurrence of external sources (listed as ecogeographical gaps) and a low
247 occurrence of accessions were considered priority collecting sites, as were areas corresponding
248 to the ELC map categories with a low frequency for the species and in the territory. Occurrence
249 data from external sources were then ranked according to their priority of collection based on the

250 frequency of each ELC map category in the study area, the frequency of each species in each
251 ELC map category and the differences between the external sources data set and the accessions
252 data set. Ecogeographical gap identification and prioritization of external sources occurrence
253 data for germplasm collection were performed using the Representa tool of CAPFITOGEN
254 (Parra-Quijano et al., 2015). External sources which occurred in ecogeographical categories not
255 represented by the corresponding species in the genebanks of the Spanish Network, *i.e.*, external
256 sources reclassified by the Representa tool within the range from 1 to 4, were considered high
257 priority ecogeographical gaps.

258

259 Selection of collection sites for traits of tolerance to drought and salinity

260 To search for populations with a higher probability of containing phenotypes with a high
261 tolerance to drought and salinity, the external sources considered high priority ecogeographical
262 gaps were filtered using the Lang aridity index (AI_L) and topsoil salinity. The Lang aridity index
263 was calculated as:

264 $$AI_L = \text{Annual precipitation / (Annual mean temperature)}$$

265 Following the ecogeographical filtering technique of predictive characterization (Thormann et
266 al., 2016), we selected populations occurring in sites with $AI_L < 40$. We considered the 20% of
267 the populations with the lowest Lang aridity index and the highest topsoil salinity (HWS
268 Database, 2012) values to be the fraction of interest for each of the five target species.

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

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

272

273 Results

274 Germplasm collection sites and presence data

275 Data pre-processing showed that the most frequently recorded species and those with the widest
276 distribution in Spain were *Ae. geniculata* and *Ae. triuncialis*. The least frequently recorded
277 species was *Ae. biuncialis*. Table 2 shows the number of accessions and occurrence data from
278 external sources for each species before and after clearing spatial duplicates and applying the
279 geo-referencing quality threshold. The percentage of spatial duplicates in the occurrence data of
280 external sources ranged from 47 to 65%, recorded for *Ae. biuncialis* and *Ae. geniculata*,
281 respectively. Applying the geo-referencing quality threshold decreased the number of non-
282 duplicated populations to 6 and 21 % in the occurrence data of external sources of *Ae. biuncialis*
283 and *Ae. geniculata,* respectively.

284 **Table 2** Number of *Aegilops* germplasm accessions and occurrence records from external
285 sources with geographical coordinates included in the study before and after clearing spatial
286 duplicates and filtering by the geo-referencing quality threshold

287 The species records for *Ae. geniculata* (accessions + external sources) remaining after the
288 clearing of spatial duplicates and the geo-referencing quality threshold are shown in Fig. 2. The
289 maps of the other *Aegilops* species are shown in Supplemental Figures 1-4.The populations
290 represented by accessions in the genebanks are not homogeneously distributed in the studied area
291 (see Fig. 2a), nor are they more frequently located in the areas where the presence of these
292 species was reported by external sources. For instance, 66% of the preserved populations of the
293 five target species (226 out of 345 populations) were collected in the autonomous communities
294 of Extremadura and Castilla-La Mancha, whereas only 10% of the external sources (272 out of
295 2614 populations) are located in these two communities. The populations reported by external
296 sources are abundant in the south of Spain, in the provinces on the eastern coast and in the
297 autonomous community of Navarra.

298 **Fig. 2** Location of (a) origin of germplasm accessions in Peninsular Spain, the Balearic Islands
299 and the Canary Islands and (b) population occurrence from external sources of *Ae. geniculata*

300

301 **Identification of variables of importance**

302 The variables selected for each studied species and ecogeographical component (bioclimatic,
303 geophysic and edaphic variables) are shown in Table 3. The variables isothermality and altitude
304 were selected in four of the five studied species.

305 **Table 3** Variables selected in each *Aegilops* species and ecogeographical component according
306 to the importance function of the random forest approach (see Supplemental Table 1 for variable
307 description)

308

309 **Generation of ELC maps**

310 The ELC map obtained for *Ae. geniculata* is shown in Fig. 3. The maps of the other *Aegilops*
311 species are shown in Supplemental Figures 5-8. The number of ecogeographical categories
312 generated in the maps ranged between 26 (in the ELC map for *Ae. biuncialis*) and 27 (in the
313 maps for the other species). The environmental characteristics of the different categories of each
314 ELC map are summarized in Supplemental Table 2.

315 **Fig. 3** ELC map of *Ae. geniculata* for Peninsular Spain, the Balearic Islands and the Canary
316 Islands. The environmental characteristics of the different categories are described in
317 Supplemental Table 2

318

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

320 Among the populations reported by external sources, 2571 were identified as spatial gaps. *Ae.*
321 *geniculata* was identified as the species with the largest number of spatial gaps. As seen in Table
322 4, 393 populations were identified as priority ecogeographical gaps in the registered genebank
323 collections of *Aegilops* in the Spanish National Inventory of Plant Genetic Resources. These 393
324 populations occur in ecogeographical categories that are not represented by the corresponding
325 species in the Spanish Network. *Ae. geniculata*, the first-ranking species in number of spatial
326 gaps, was also identified as the species with the largest number of ecogeographical gaps. *Ae.*
327 *biuncialis*, the species with the lowest number of preserved accessions (Table 2), is also the
328 species whose *ex situ* ecogeographical representativeness needs the most improvement, as 80%
329 of the available external sources were identified as high priority gaps. On the contrary, only 10%
330 of the analyzed external sources of *Ae. geniculata* and *Ae. triuncialis*, the two species with the
331 highest number of preserved accessions (Table 2), were identified as high priority
332 ecogeographical gaps.

333 **Table 4** Number of population occurrences from external sources of *Aegilops* subjected to
334 representativeness analysis and number of spatial gaps and priority ecogeographical gaps
335 identified in Spain

336 The geographic distribution of the populations identified as priority ecogeographical gaps of *Ae.*
337 *geniculata* is shown in Fig. 4. The maps of the other *Aegilops* species are shown in Supplemental
338 Figures 9-12.

339 **Fig. 4** Location of the *Ae. geniculata* populations identified as priority ecogeographical gaps in
340 Spain

341 Including germplasm from priority ecogeographical gaps in the genebanks of the Spanish
342 Network on PGRFA would significantly improve their ecogeographical representativeness
343 (number of ecogeographical categories not currently represented in the Spanish Network, Table
344 5). The obtained percentage of ELC categories represented in the Spanish Network ranged from
345 8% to 48% in *Ae. biuncialis* and *Ae. geniculata*, respectively. As the priority ecogeographical
346 gaps belong to categories not yet represented, their collection and conservation would contribute
347 to increasing the percentage of ELC categories represented in the Spanish Network to values
348 ranging from 27% to 70% in *Ae. biuncialis* and *Ae. neglecta,* respectively.

349 **Table 5** Number of ELC categories for *Aegilops* currently represented in the Spanish Network
350 and potential increase (%) in representativeness after collecting priority ecogeographical gaps

351

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

353   Among the 393 populations identified as priority ecogeographical gaps, 223 populations inhabit
354   sites with a Lang index value < 40, and thus are potentially adapted to arid environments (Table
355   6). The geographic location of these accessions is shown in Fig. 5a. The 20% of these 223
356   populations with the highest values of topsoil salinity for each species (20%) and their
357   geographical distribution are shown in Table 6 and Fig. 5b, respectively. These 45 populations
358   (Table 6) constitute the predictive characterization (PC) subset of *Aegilops* populations of
359   potential interest due to their potential tolerance to drought and salinity, in addition to
360   representing ecogeographical gaps. Table 7 contains complete information on the geographic
361   location, the Lang aridity index and topsoil salinity of the populations included in this PC subset.

362   **Table 6** Priority ecogeographical gaps of targeted *Aegilops* species selected for drought and
363   salinity tolerance in Spain

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

367   **Table 7** Geographic description of the Spanish *Aegilops* populations selected as potentially
368   tolerant to drought and salinity

369   The inclusion of the predictive characterization subset populations in the genebanks of the
370   Spanish Network on PGRFA would improve their ecogeographical representativeness between
371   7% and 11% (Table 8).

372   **Table 8** Number of ELC categories of *Aegilops* species currently represented in the Spanish
373   Network and potential increase (%) in representativeness by collecting populations of the
374   predictive characterization (PC) subset

375

376   **Discussion**

377   **The preprocessing of presence data**

378   The quality of geo-referencing in spatial analysis is very important to obtain reliable results.
379   Maldonado et al. (2015), who studied the role of natural history collections in unveiling correct
380   patterns of biodiversity and distribution, concluded that geographic inaccuracy affects diversity
381   patterns more than taxonomic uncertainties. Similarly, Graham et al. (2008) evaluated how
382   uncertainty in geo-references and associated location errors in occurrences influence species
383   distribution modeling and found that models run with data subject to random location errors
384   resulted in less accurate models in many species. However, few studies on genetic diversity or
385   taxonomic spatial distribution describe robust methods to ensure the quality of geo-referenced
386   data. Some authors, such as Fielder et al. (2015, 2016), excluded records dated from before 1970,
387   records lacking both coordinates and location descriptions and records with a precision lower

388   than 4 km². Ramírez-Villegas et al. (2010) carried out a process to verify and correct the
389   coordinates using BioGeomancer, Google Earth and highly detailed maps. Khoury et al. (2015)
390   cross-checked the coordinates to country and verified that they occurred on land. After that,
391   occurrence data were evaluated for correctness with experts on the target species. In our study,
392   apart from removing accessions and external sources with low accuracy according to the
393   established criteria, the assessment of the quality of the georeferenced data allowed us to identify
394   the records with the highest quality. In this sense, the final number of records included in the
395   analysis (considering both accessions and external sources) ranged between 30% and 53% of the
396   initial number of records (for *Ae. geniculata* and *Ae. biuncialis*, respectively). This selective use
397   of records, which complies with the minimum standards of georeferencing quality, reduces the
398   probability of generating erroneous results in the analysis.

399

400   **Identification of spatial and ecogeographical gaps**

401   Spatial bias in collecting activities and chorological studies often affects the spatial distribution
402   of the species, as shown in Maldonado et al. (2015). In our study, the spatial distribution of the
403   accessions could reflect the intensity of collecting activities rather than the real distribution of the
404   species (see Fig. 2a). A recent project focused on collecting *Aegilops* germplasm may have
405   contributed to the difference observed in the numbers of accessions between areas because it was
406   focused on the collection of *Ae. geniculata, Ae. neglecta, Ae. triuncialis* and *Ae. ventricosa* in
407   the west, center and south of the country.

408   In a similar way, the distribution of the external sources of the targeted species (Fig. 2 b) shows a
409   higher number of populations in some areas of the country, such as the autonomous communities
410   of Navarra and Valencia. This may reflect a higher intensity of chorological studies rather than a
411   higher presence of the species in these areas. The fact that the distribution of external sources is
412   different from accessions indicates lack of collection missions in particular areas, for example
413   the Canary Islands.

414   It is worthy to note that the real distribution of the species could be more extensive that the
415   distribution shown by the combination of external sources and accessions. For instance, the
416   target species might occur in places completely outside of the parameters used that are based on
417   historical information. In this sense, it should be clearly understood that the efficacy of this
418   collecting methodology, like any other, heavily depends on how well currently available
419   chorological data represents the real distribution of the species.

420   Several chorological studies involving the spatial gap analysis of different species have been
421   used as a guide for germplasm collecting. For instance, Maxted et al. (2008) identified *ex situ*
422   conservation gaps in *Aegilops* germplasm collections as regions where the species were predicted
423   to occur according to species distribution models, but previous collection had not taken place. In
424   this study, the priority of germplasm collecting for each of *Aegilops* species was ranked high,

425   medium or low, according to the number of germplasm accessions already conserved *ex situ* and
426   the number of predicted under-sampled regions. Shehadeh et al. (2013) followed this
427   methodology to carry out a gap analysis of *Lathyrus* L. species. Another recent methodology for
428   gap analysis was described in Ramírez-Villegas et al. (2010) and applied to wild taxa of the
429   *Phaseolus* genepool. It involves an eight-step process to evaluate conservation deficiencies at
430   three different levels (taxonomic, geographic and environmental) through the calculation of
431   sampling, geographic and environmental representativeness scores. This methodology was also
432   applied in Castañeda-Álvares et al. (2015) to identify *ex situ* conservation priorities for the wild
433   relatives of potato and in Khoury et al. (2015) in their study of the CWR pigeonpea.

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

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

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

455   On the other hand, predictive species distribution models based on bioclimatic information only
456   offer a partial view of plant abiotic adaptation. In our study, we generated ecogeographical land
457   characterization maps using the values of bioclimatic, geophysic and edaphic variables at species
458   occurrence sites to suggest the potential adaptive scenarios of the species. The representativeness
459   analysis, based on the comparison of the ELC categories of the accessions and the external
460   sources considered spatial gaps, helped us to select priority collecting sites, avoid over-
461   representation and identify populations from low and non-represented ecogeographical
462   categories. This allowed us to identify populations in marginal environments in the species'

463 range, where interesting traits related to abiotic stress tolerance may be found. The easy-to-use
464 tools employed in this study (i.e. CAPFITOGEN tools, Parra-Quijano et al., 2015) allow
465 genebank curators and technicians in charge of collecting activities to develop their own
466 germplasm collecting design based on spatial and ecogeographical analyses.

467 A sound ecogeographical gap analysis requires that the resolution of the ecogeographical layers
468 used is in a comparable scale to the resolution used for the chorological data. That is why, in our
469 study, chorological data and ecogeographical data were filtered and/or chosen to accommodate
470 to a similar resolution.

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

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

487

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

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

493 One of the results of this study is the identification of 45 *Aegilops* populations of high interest
494 due to their potential tolerance to drought and salinity, in addition to being ecogeographical gaps
495 in the existing Spanish germplasm collections. Although predictive characterization is an
496 inexpensive and effective approach to maximizing the likelihood of capturing a desirable level of
497 trait expression among accessions of landraces and traditional varieties, this is the first time that
498 predictive characterization has been applied to identifying wheat wild relatives with potential

499   tolerance to drought and salinity. Interestingly, the validity of predictive characterization is
500   presumably greater when used with wild relatives rather than when used with landraces, because
501   wild relatives are more greatly affected by natural selection (no artificial selection). Thus, the
502   link between existing environmental conditions and genotypes with local genetic adaptation is
503   likely to be stronger than the link between environmental conditions and landraces or modern
504   varieties, which is explored by FIGS.

505   In spite of having different geographic distributions, all studied *Aegilops* species occur in a
506   variety of soil bedrock types (although predominantly in limestone), have a considerable
507   precipitation range and are largely autogamous. The latter implies that genetic diversity is likely
508   to be small within populations while local adaptation can generate large genetic differentiation
509   among populations subject to different environmental conditions. Priority ecogeographical gaps
510   of *Aegilops* that occur in sites where the Lang aridity index is < 40 (Fig. 5a) show a good
511   representation of all five species, including *Ae. neglecta* and *Ae. triuncialis*, which can also been
512   found in habitats more humid than those found in the other three species. The selection for the
513   highest topsoil salinity values (Fig. 5b) generated a more heterogenous representation of the five
514   species. Although *Ae. ventricosa* is the only species explicitly characterized in saline locations,
515   just two populations of this species where selected in the PC subset for salinity tolerance. On the
516   other hand, *Ae. geniculata* and *Ae. neglecta*, were the most represented in this subset (47 and
517   33%, respectively). This is largely explained by their large distribution area and great number of
518   recorded populations. This implies that both species with more narrowly specialized habitats and
519   those with larger distribution and wider ecological range can positively contribute genetic
520   adaptations to particular environmental conditions.

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

529   According to predictive characterization methods (Thormann et al., 2016), the probability of
530   capturing phenotypes tolerant to drought and salinity in the predictive characterization subset
531   would be higher than in a randomly chosen set. The 45 priority populations identified by the
532   predictive characterization approach in this study are now considered priority collection
533   populations for further optimized collecting activities of *Aegilops* germplasm. They will be
534   shortly looked for and collected for *ex situ* conservation. When planning the collecting mission,
535   the combined skills of a good field botanist, genetic resources expert and ecologist taking part in
536   the collecting team should be used to critically assess the predictions obtained with this analysis
537   and make further predictions for sites with potentially salt and drought tolerant germplasm.

538 Quality limitations of currently available species distribution and ecogeographic data may
539 generate priority collection sites where the species no longer exist or the estimated
540 ecogeographic conditions do not actually take place. Therefore, it is essential that the aridity and
541 topsoil salinity conditions are verified at the site at the same time seeds are collected in the
542 population. Then, the collected populations will be assessed for their tolerance to drought and
543 salinity in order to validate the methodology.

544 **Conclusions**

545 This study aimed to identify spatial and ecogeographical gaps in the Spanish germplasm
546 collection of *Aegilops* and priority populations of potential interest due to their possible tolerance
547 to drought and salinity. The methodology employed allowed us to establish an optimized
548 collecting strategy by filtering potential collecting sites, thereby avoiding over-representation
549 and identifying populations from low and non-represented ecogeographical categories. It also
550 provided a subset of 45 populations of potential interest in terms of tolerance to drought and
551 salinity. Because current knowledge of the real distribution of this species is still incomplete, this
552 methodological approach should be observed in dynamic terms. As the quantity and quality of
553 species distribution data improves, this type of study should be carried out again and the
554 collection priorities should be updated accordingly.

555 Subsequent collections and evaluations of these populations will provide essential feedback on
556 the efficacy of these approaches to improving the genetic representativity of genebank
557 collections and identify genotypes with desired traits.

558

559 **Acknowledgments**

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

562

563 **References**

564 - Abrol IP, Yadav JSP, Massoud FI. 1988. Salt-affected soils and their management. FAO
565   Soils Bulletin 39, FAO, Rome. Available at
566   http://www.fao.org/docrep/x5871e/x5871e03.htm (accessed 06 June 2016).
567 - Al-maskri A, Hameed M, Ashraf M, Khan M, Fatima S, Nawaz T, Batool R. 2014.
568   Structural features of some wheat (*Triticum* spp.) landraces/cultivars under drought and
569   salt stress. Arid Land Research and Management 28:355-370.
570 - Araus JL, Slafer GA, Reynolds MP, Royo C. 2002. Plant breeding and drought in C_3
571   cereals: what should we breed for? Annals of Botany 89:925-940.

572   -   Arrigo N, Guadagnuolo R, Lappe S, Pasche S, Parisod C, Felber F. 2011. Gene flow
573   between wheat and wild relatives: empirical evidence from Aegilops geniculata, Ae.
574   neglecta and Ae. triuncialis. Evolutionary Applications. DOI:10.1111/j.1752-
575   4571.2011.00191.x
576   -   Bari A, Amri, A Street K, Mackay M, De Pauw E, Sanders R, Nazari K, Humeid B,
577   Konopka J, Alo F. 2014. Predicting resistance to stripe (yellow) rust (*Puccinia*
578   *striiformis*) in wheat genetic resources using focused identification of germplasm
579   strategy. Journal of Agricultural Science 152:906-916.
580   -   Bari A, Street K, Mackay M, Endresen DTF, De Pauw E, Amri A. 2011. Focused
581   identification of germplasm strategu (FIGS) detects wheat stem rust resistance linked to
582   environmental variables. Genetic Resources and Crop Evolution. doi:10.1007/s10722-
583   011-9775-5.
584   -   Bhullar NK, Street K, Mackay M, Yahiaoui N, Keller B. 2009. Unlocking wheat genetic
585   resources for the molecular identification of previously undescribed functional alleles at
586   the Pm3 resistance locus. Proceedings of the National Academy of Sciences 106:9519-
587   9524.
588   -   Castañeda-Álvares NP, de Haan S, Juárez H, Khoury CK, Achicanoy HA, Sosa CC,
589   Vernay V, Salas A, Heider B, Simon R, Maxted N, Spooner DM. 2015. Ex situ
590   conservation priorities for the wild relatives of potato (*Solanum* L. section *petota*). Plos
591   One. doi:101371/journalpone0122599.
592   -   Colmer TD, Flowers TJ, Munns R. 2006. Use of wild relatives to improve salt tolerance
593   in wheat. Journal of Experimental Botany 57 (5):1059-1078.
594   -   Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007.
595   Random Forests for classification in ecology. Ecology 88 (11):2783-2792.
596   -   El Bouhssini M, Street K, Amri A, Mackay M, Ogbonnaya FC, Omran A, Abdalla O,
597   Baum M, Dabbous A, Rihawi F. 2011. Sources of resistance in bread wheat to Russian
598   wheat aphid (*Diuraphis noxia*) in Syria identified using the Focused Identification of
599   Germplasm Strategy (FIGS). Plant Breeding 130:96-97.
600   -   El Bouhssini M, Street K, Joubi A, Ibrahim Z, Rihawi F. 2009. Sources of wheat
601   resistance to Sunn pest, *Eurygaster integriceps* Puton, in Syria. Genetic Resources and
602   Crop Evolution 56:1065-1069.
603   -   Endresen DTF, Street K, Mackay M, Bari A, Amri A, De Pauw E, Nazari K, Yahyaoui
604   A. 2012. Sources of resistance to stem rust (Ug99) in bread wheat and durum wheat
605   identified using focused identification of germplasm strategy. Crop Science 52:764-773.
606   -   Endresen DTP. 2011. Predictive association between biotic stress traits and eco-
607   geographic data for wheat and barley landraces. Crops Science 51:2036-2055.
608   -   FAO/Bioversity Multi-crop Passport Descriptors V.2. 2012. Available at
609   http://www.bioversityinternational.org/fileadmin/user_upload/online_library/publications
610   /pdfs/FAO-Bioversity_multi_crop_passport_descriptors_V_2_Final_rev_1526.pdf
611   (accessed 30 June 2016).
612   -   FAO. 2013a. FAOSTAT. Available at http://faostat3.fao.org/home/E (07 January 2016).
613   -   FAO. 2013b. Drought. Available at http://www.fao.org/docrep/017/aq191e/aq191e.pdf
614   (accessed 07 January 2016).

615 - Farroq S. 2002. *A. ovata*: a potential gene source for improvement of salt tolerance of
616   wheat. In: Ahmad R, Malik KA, ed. Prospects for saline agriculture, vol 37. Dordrecht:
617   Kluwer Academic Publishers, 123-130.
618 - Fielder H, Brotherton P, Hosking J, Hopkins JJ, Ford-Lloyd B, Maxted N. 2015.
619   Enhancing the conservation of crop wild relatives in England. PLoS ONE 10(6):
620   e0130804. doi:10.1371/journal.pone.0130804.
621 - Fielder H, Sith C, Ford-Lloyd B, Maxted N. 2016. Enhancing the conservation of crop
622   wild relatives in Scotland. Journal for Nature Conservation 29:51-61.
623 - Fita A, Rodríguez-Burruezo A, Boscaiu M, Prohens J, Vicente O. 2015. Breeding and
624   domesticating crops adapted to drought and salinity: a new paradigm for increasing food
625   production. Frontiers in Plant Science volumen 6 article 978.
626 - Glick BR, Cheng Z, Czarny J, Duan J. 2007. Promotion of plant growth by ACC
627   deaminase-producing soil bacteria. European Journal of Plant Pathology 119:329-339.
628 - Graham CH, Elith J, Hijmans RJ, Guisan A, Peterson AT, Loisells BA, The Nceas
629   Predicting Species Distributions Working Group. 2008. Journal of Applied Ecology
630   45:239-247.
631 - Hameed M, Ashraf M, Naz N. 2011. Anatomical and physiological characteristics
632   relating to ionic relations in some salt tolerant grasses from the Salt Range. Pakistan Acta
633   Physiologiae Plantarum 33:1399-1409.
634 - Hammer K. 1980. Vorarbeiten zur Monographischen Darstellung von
635   Wildpflanzensortimenten: Aegilops L. Kulturpflanze 28:33–180.
636 - Harmonized world soil (HWS) Database (version 12). 2012. Available at
637   http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/. FAO,
638   Rome, Italy and IIASA, Laxenburg, Austria (accessed 1 April 2013).
639 - Heywood V, Casas A, Ford-Lloyd B, Kell S, Maxted N. 2007. Conservation and
640   sustainable use of crop wild relatives. Agriculture, Ecosystems and Environment 121:
641   245-255.
642 - Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution
643   interpolated climate surfaces for global land areas International Journal of Climatology
644   25:1965-1978 (http://www.worldclim.org).
645 - Iriondo JM, Albert MJ, Giménez L, Domínguez F, Escudero A. 2009. Poblaciones en
646   peligro: viabilidad demográfica de la flora vascular amenazada de España. Dirección
647   General de Medio Natural y Política Forestal (Ministerio de Medio Ambiente, y Medio
648   Rural y Marino), Madrid.
649 - Jarvis A, Reuter HI, Nelson A, Guevara E. 2008. Digital Elevation Models (DEM) of the
650   Shuttle Radar Topography Mission (SRTM). Hole-filled SRTM for the globe Version 4.
651   Available from the CGIAR-CSI SRTM 90m Database at http://srtm.csi.cgiar.org/
652   (accessed 07 January 2016).
653 - Ketchen DJ, Shook CL. 1996. The application of cluster analysis in strategic
654   management research: an analysis and critique. Strategic Management Journal 17
655   (6):441-458.

656   -   Khazaei H, Strett K, Bari A, Mackay M, Stoddard F. 2013. The FIGS (Focused
657   Identification of Germplasm Strategy (FIGS) approach identifies traits related to drought
658   adaptation in *Vicia faba* genetic resources. Plos One, October, Volume 8, Issue 3.
659   -   Khoury CK, Castañeda-Álvarez NP, Achicanoy HS, Sosa CC, Bernau V, Kassa MT,
660   Norton SL, van der Maesen JG, Upadhyaya HD, Ramírez-Villegas J, Jarvis A, Struik PC.
661   2015. Crop wild relatives of pigeonpea (Cajanus cajan (L.) Millsp.): distributions, ex situ
662   conservation status, and potential genetic resources for abiotic stress tolerance. Biological
663   Conservation 184:259-270.
664   -   Khoury C, Laliberté B, Guarino L. 2010. Trends in ex situ conservation of plant genetic
665   resources: a review of global crop and regional conservation strategies. Genetic
666   Resources and Crop Evolution 57: 625-639.
667   -   Lambers H, Chapin FS, Pons TL. 2008. Plant Physiological Ecology, 2nd Ed. Springer,
668   New York.
669   -   Mackay MC, Street K. 2004. Focused identification of germplasm strategy – FIGS. In:
670   Black CK, Panozzo JF, Rebetzke GJ, ed. Cereals 2004, Proceedings of the 54th Australian
671   Cereal Chemistry Conference and the 11th Wheat Breeders' Assembly Cereal Chemistry
672   Division. Royal Australian Chemical Institute, Melbourne, 138-141.
673   -   Maldonado C, Molina CI, Zizka A, Persson C, Taylor CM, Albán J, Chilquillo E,
674   Ronsted N, Antonelli A. 2015. Estimating species diversity and distribution in the era of
675   Big Data: to what extent can we trust public databases? Global Ecology and
676   Biogeography. doi:101111/geb12326.
677   -   Maxted N, White K, Valkoun J, Konopka J, Hargreaves S. 2008. Towards a conservation
678   strategy for *Aegilops* species. Plant Genetic Resources: Characterization and Utilization 6
679   (2):126-141.
680   -   Mólnar I, Gáspár L Sárvári E, Dulai S, Hoffmann B, Molnár-Láng M, Galiba G. 2004.
681   Physiological and morphological responses to water stress in *Aegilops biuncialis* and
682   *Triticum aestivum* genotypes with differing tolerance to drought. Functional Plan Biology
683   31:1149-1159.
684   -   Mondini L, Nachit MM, Pagnotta MA. 2015. Allelic variants in durum wheat (*Triticum
685   turgidum* L. var. *durum*) DREB genes conferring tolerance to abiotic stresses. Molecular
686   Genetics and Genomics 290:531-544.
687   -   Mujeeb-Kazi A, Diaz de Leon JL. 2002. Conventional and alien genetic diversity for salt
688   tolerant wheats: focus on current status and new germplasm development. In: Ahmad R,
689   Malik KA, ed. Prospects for saline agriculture 37. Dordrecht: Kluwer Academic
690   Publishers, 69-82.
691   -   Noorka IR, Pat Heslop-Harrison JS. 2014. Water and Crops: Molecular biologists,
692   physiologists, and plant breeders approach in the context of evergreen revolution. In:
693   Taylor, F, ed. Hand book of Plant and Crop Physiology. CRC Press, 967-978.
694   -   Parra-Quijano M, Torres E, Iriondo JM, López F. 2015. CAPFITOGEN Tools. User
695   Manual version 2.0. International Treaty on Plant Genetic Resources for Food and
696   Agriculture, FAO, Rome.

697   -   Parra-Quijano M, Iriondo JM, Torres E. 2012a. Ecogeographical land characterization
698   maps as a tool for assessing plant adaptation and their implications in agrobiodiversity
699   studies. Genetic Resources and Crop Evolution 59:205-217.
700   -   Parra-Quijano M, Iriondo JM, Torres ME. 2012b. Improving representativeness of
701   genebank collections through species distribution models, gap analysis and
702   ecogeographical maps. Biodiversity and Conservation 21:79-96.
703   -   Parra-Quijano M, Draper D, Torres E, Iriondo JM. 2008. Ecogeographical
704   representativeness in crop wild relative ex-situ collections. In: Maxted N, Ford-Lloyd
705   BV, Kell SP, Iriondo JM, Dullo ME, Turok J, ed. Crop wild relative conservation and
706   use. CAB International, Wallingford, 249-273.
707   -   Ramírez-Villegas J, Khoury C, Javis A, Debouck DG, Guarino L. 2010. A gap analysis
708   methodology for collecting crop genepools: a case study with *Phaseolus* beans. PLoS
709   ONE 5(10): e13497. doi:10.1371/journal.pone.0013497.
710   -   Rubio Teso ML, Iriondo JM, Parra-Quijano M, Torres E. 2013. National Strategy for the
711   Conservation of Crop Wild Relatives of Spain PGR Secure. Available at
712   http://www.pgrsecure.bham.ac.uk/sites/default/files/documents/public/National_CWR_C
713   onservation_Strategy_Spain.pdf (accessed 30 June 2016).
714   -   Shehadeh A, Amri A, Maxted N. 2014. Ecogeographic survey and gap analysis of
715   *Lathyrus* L. species. Genetic Resources and Crop Evolution 60:2101-2113.
716   -   Thormann I, Parra-Quijano M, Rubio Teso ML, Endresen DTF, Dias S, Iriondo JM,
717   Maxted N. 2016. Predictive characterization methods for accessing and using CWR
718   diversity. In: Maxted N, Dulloo ME, Ford-Lloyd BV, ed. Capturing wild relative and
719   landrace diversity for crop improvement, 64-77.
720   -   Tóth G, Montanarella L, Rusco E. 2008. Updated map of salt affected soils in the
721   European Union. In: Tóth G, Montanarella L, Rusco E, ed. Threats to soil quality in
722   Europe. EUR 23438 – Scientific and Technical Research series Luxembourg: Office for
723   Official Publications of the European Communities, 61-74.
724   -   Van Slageren MW. 1994. Wild wheats: A monograph of *Aegilops* L. and *Amblyopyrum*
725   (Jaub. & Spach) Eig. Wageningen Agricultural University, Veenman Drukkers,
726   Wageningen. The Nehterlands.
727   -   Xing X, Monneveux P, Damania AB, Zarahieva M. 1993. FAO/IPGRI Plant Genetic
728   Resources Newsletter 96:11-1.

Figure 1

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

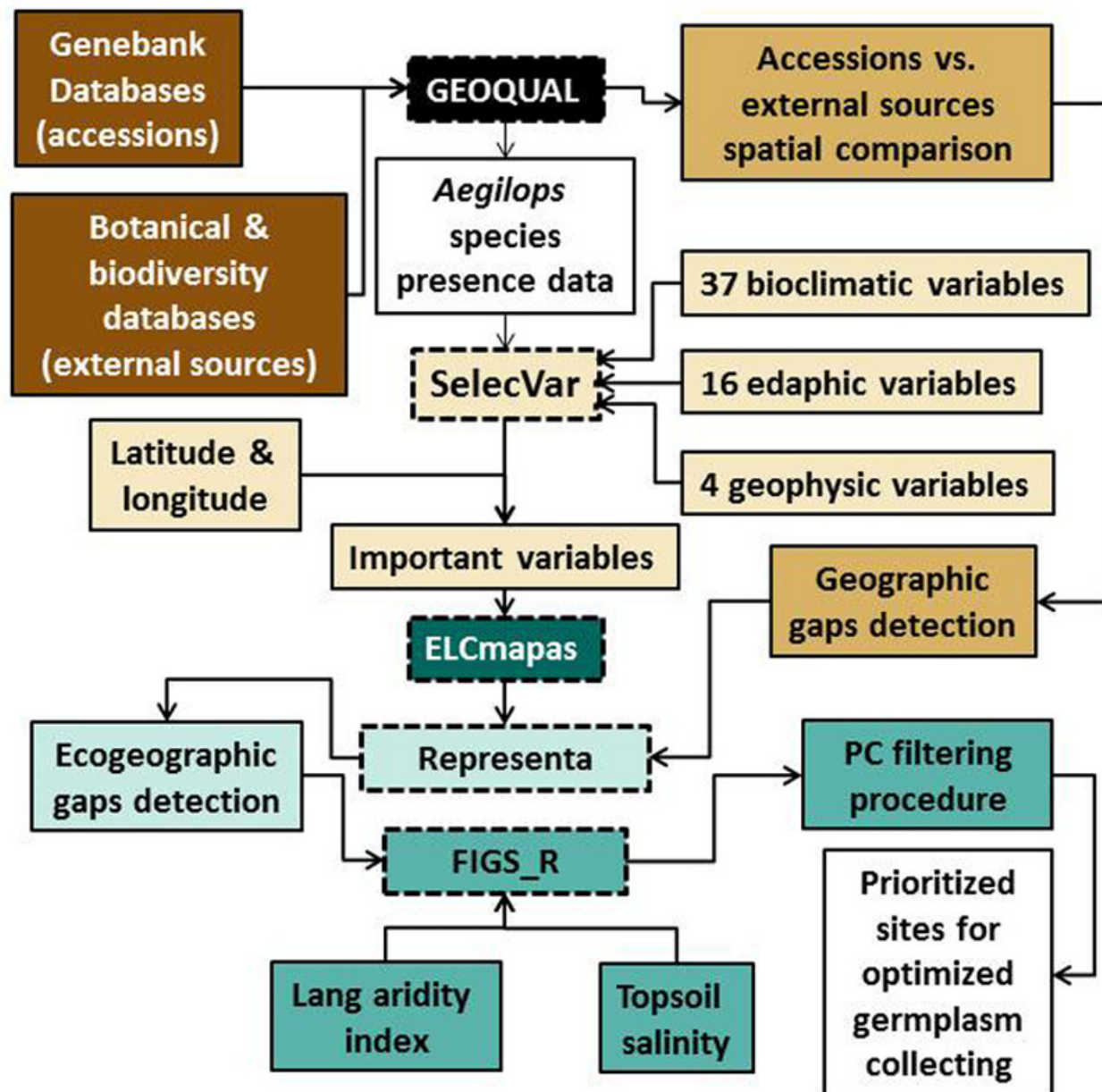



Figure 2

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 *Ae. geniculata.*

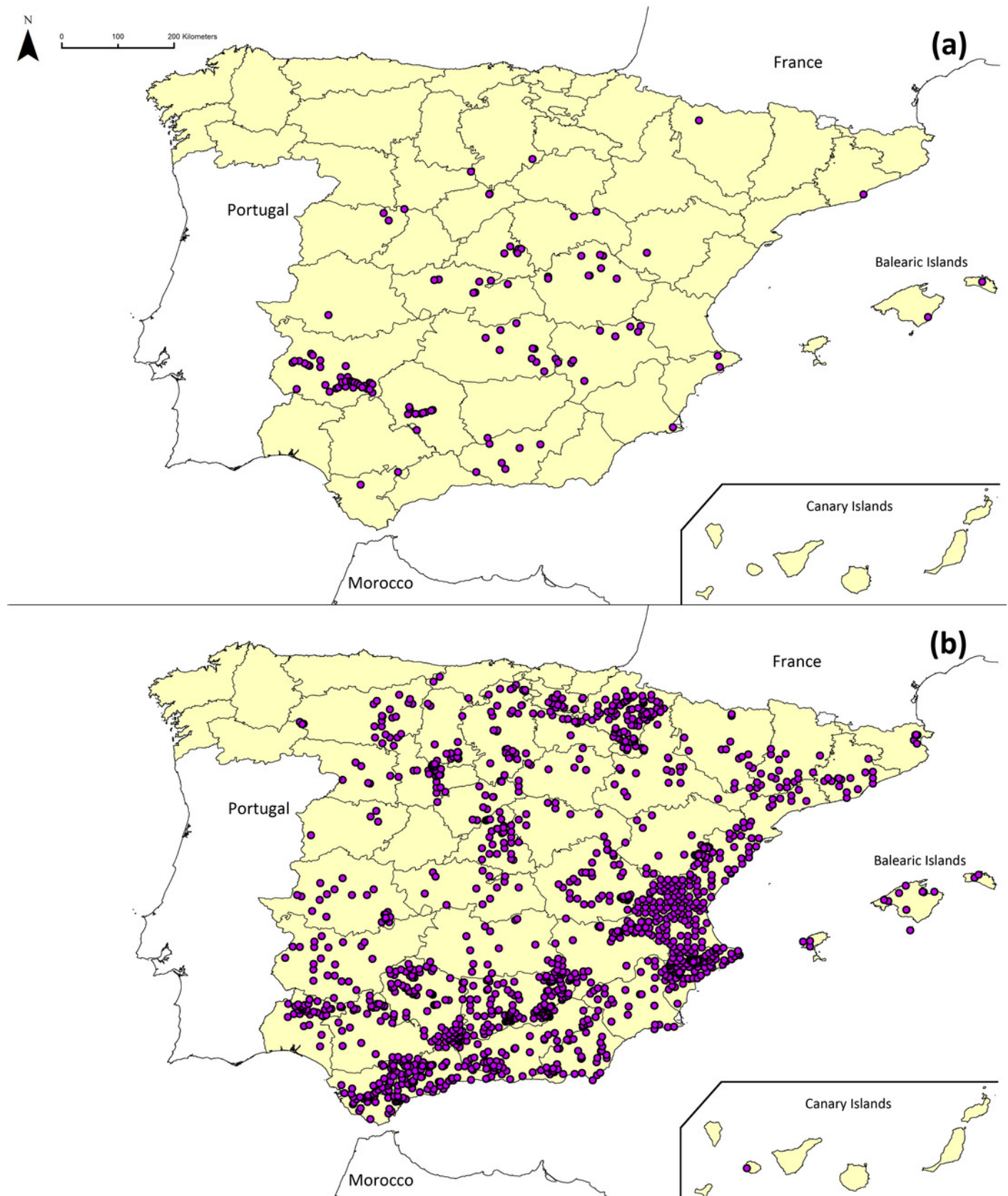



Figure 3

ELC map of *Ae. geniculata* for Peninsular Spain, the Balearic Islands and the Canary Islands. The environmental characteristics of the different categories are described in Table S10.

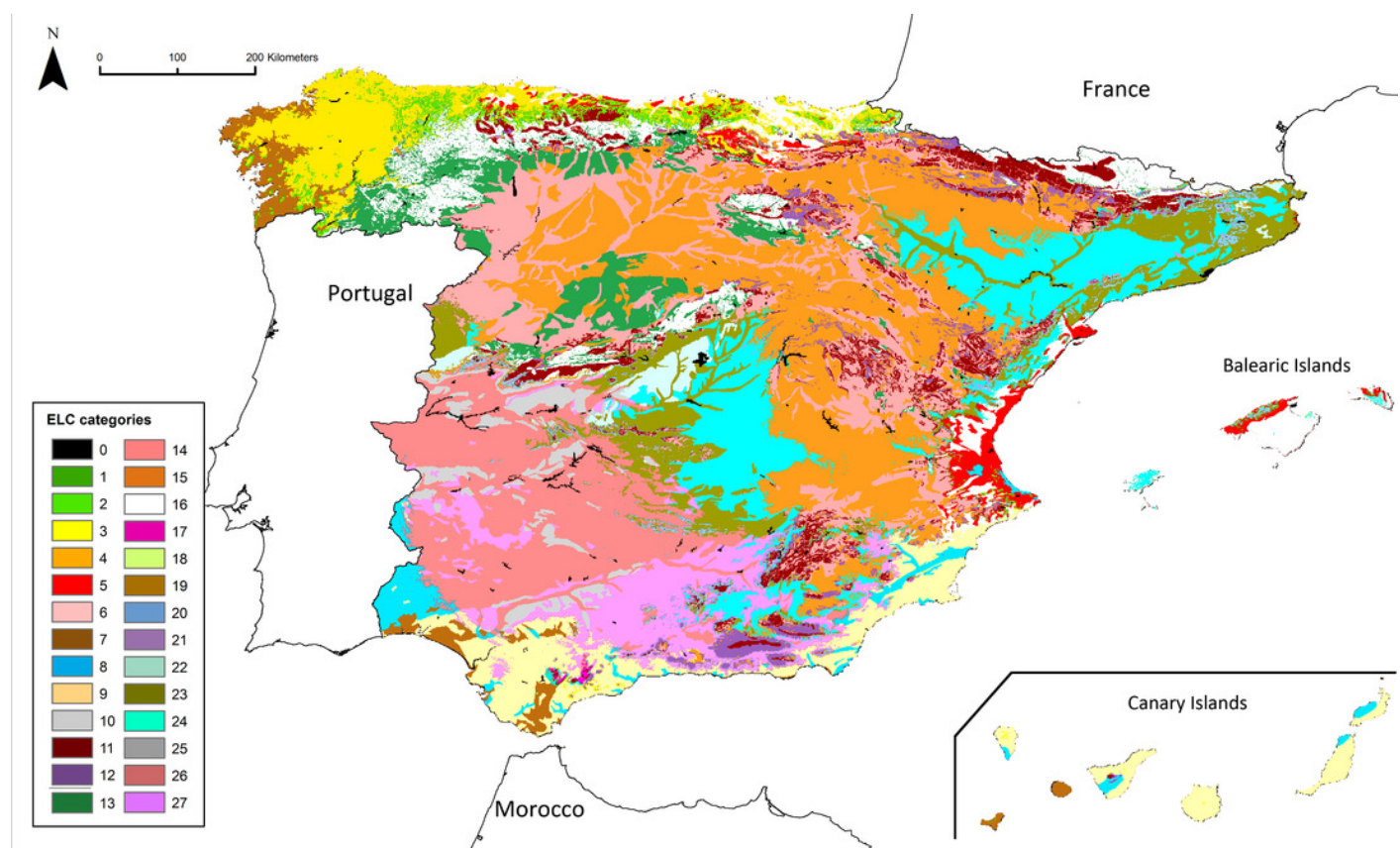



Figure 4

Location of the *Ae. geniculata* populations identified as priority ecogeographical gaps in Spain.

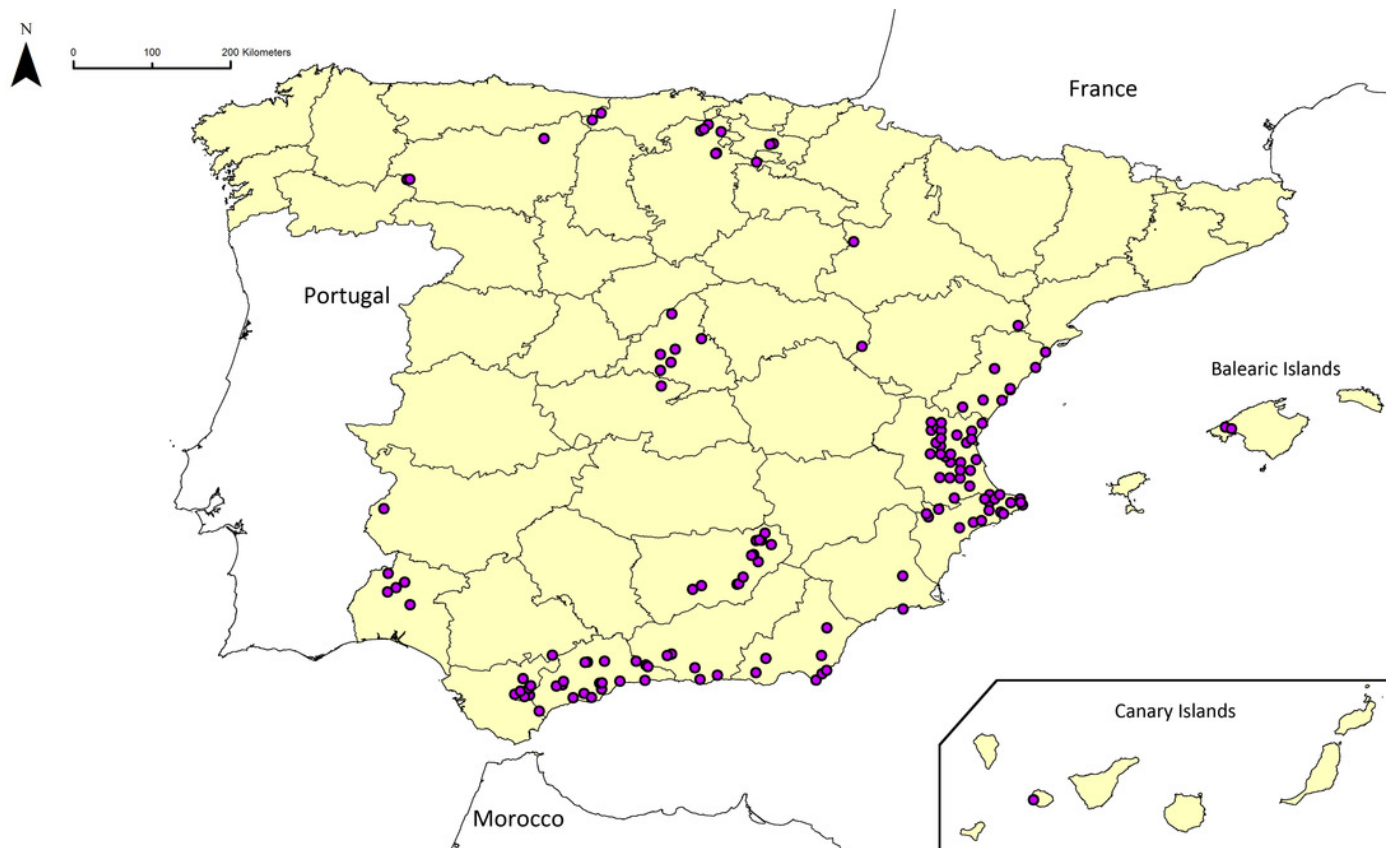
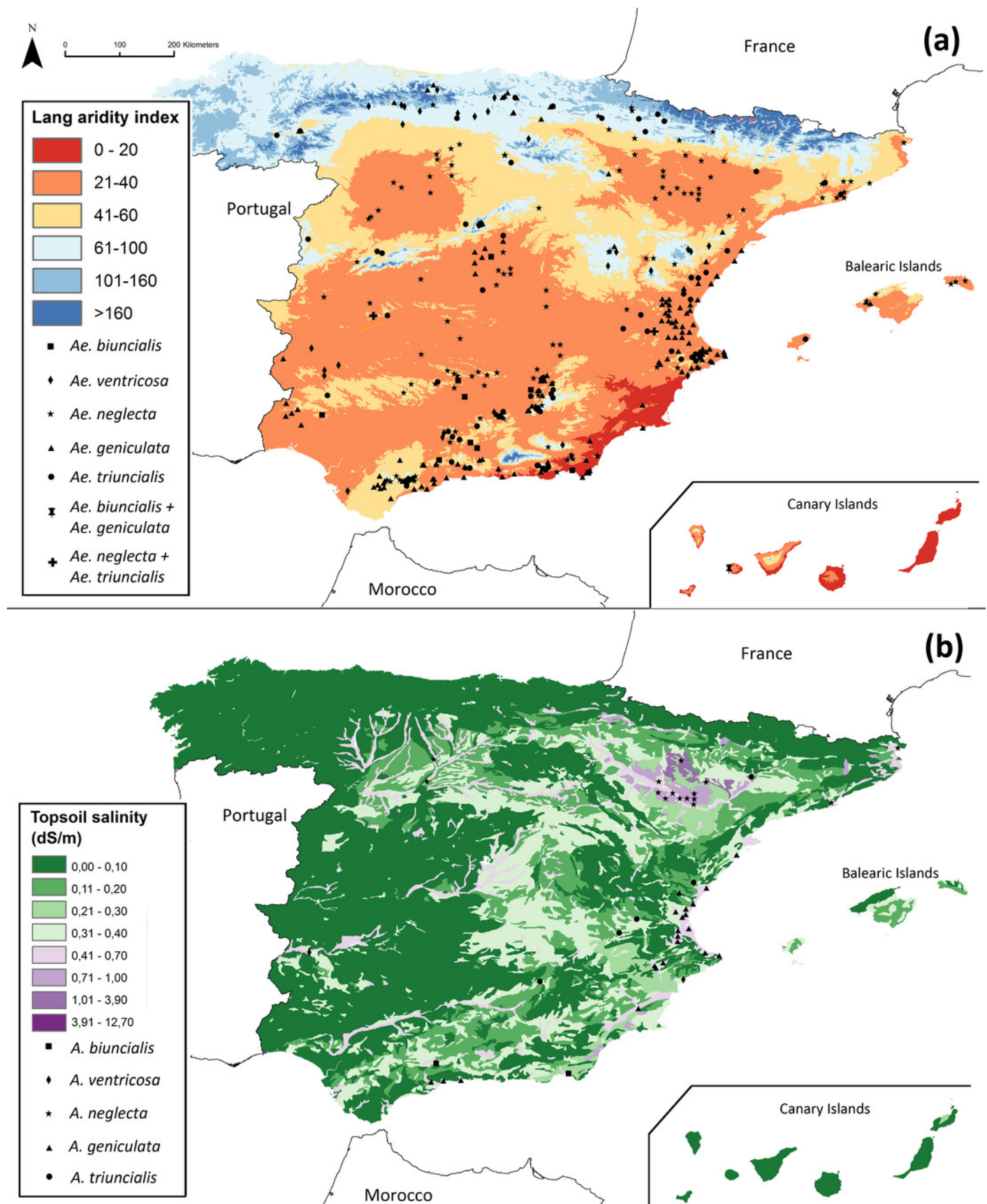



Figure 5

Location of (a) priority ecogeographical gaps (PEG) of *Aegilops* in Spain that occur in sites where $AI_L < 40$, and (b) PEG that occur in sites where the highest salinity values are also found.

(a)

N

0   100   200 Kilometers

France

Lang aridity index

- 0 - 20
- 21-40
- 41-60
- 61-100
- 101-160
- >160

Portugal

- ■ *Ae. biuncialis*
- ♦ *Ae. ventricosa*
- ★ *Ae. neglecta*
- ▲ *Ae. geniculata*
- ● *Ae. triuncialis*
- ✶ *Ae. biuncialis +*
  Ae. geniculata
- ✚ *Ae. neglecta +*
  Ae. triuncialis

Balearic Islands

Canary Islands

Morocco

(b)

France

Portugal

Topsoil salinity
(dS/m)

- 0,00 - 0,10
- 0,11 - 0,20
- 0,21 - 0,30
- 0,31 - 0,40
- 0,41 - 0,70
- 0,71 - 1,00
- 1,01 - 3,90
- 3,91 - 12,70

- ■ *A. biuncialis*
- ♦ *A. ventricosa*
- ★ *A. neglecta*
- ▲ *A. geniculata*
- ● *A. triuncialis*

Balearic Islands

Canary Islands

Morocco

Autecology and mating system of the targeted *Aegilops* species. Information obtained from Van Slageren (1994) unless otherwise stated.

	Ae.biuncialis	*Ae. geniculata*	*Ae. neglecta*	*Ae. triuncialis*	*Ae. ventricosa*
General habitat	Dry and disturbed habitats (fallow, roadsides, edges of cultivation).				
Specific habitat	Dry, rocky mountain slopes.	Dry, rocky mountain slopes, wastelands.	Stony fields and hill slopes. Marginal habitats where parent rock surfaces and only pockets of the top soil remain.	Wastelands, sandy wadis (dry riverbeds), and dry rocky slopes of hills and mountains.	Sandy wadis, including saline locations, and even marshy riversides.
Vegetation	Various forest types (frequently with *Pinus halepensis* and *Quercus* sp.). Also in grasslands, maquis vegetation and steppe. More rarely in river valleys.	Vegetation types include garrigue, maquis, grassland, shrub, woodlands, forests and scrubs (e.g., of *Quercus, Pinus, Juniperus* and *Pistacia*). Steppe, and, more rarely, also humid pastures, dunes and even swamps.	Vegetation types such as, grasslands, stony fields and hillslopes, maquis, garrigue, in forests or scrubs of e.g., *Acacia, Quercus, Ceratonia* and *Pinus*. Occasionally found on river banks and generally more humid habitats.	Vegetation types include garrigue, maquis, grassland, shrub- and woodlands, (open) forests and scrubs, e.g., of *Acacia, Quercus, Pinus, Juniperus*, and of cultivated *Pistacia, Ceratonia siliqua* and *Liquidambar* trees. Also found in the steppe up to the margin of the desert, but, more rarely, also in humid pastures, river terraces.	Grasslands. Also found in scrubs of *Pistacia* and *Juniperus,* oak forests, and in *Poterium-dominated* vegetation.
Soil	Variety of bedrock types: mainly limestone but also on schists, shales, basalt, granite, and pillow lavas. Soil textures are mainly clay- or sandy loam, or clay, or (rarely) pure sandy soils.	Bedrock is predominantly limestone but shales, pillow lava, silicate, Mediterranean terra rosa, karst, basalt and sandstone are also reported. Soil texture also varies widely: often on clay- and sandy loam, clay, and gravel; more rarely on pure	The parent rock is mainly limestone, but less frequently also alluvium, basalt, pillow lava, schists, silicates, and sandstone Recorded soil textures include loam, clayloam, sandy loam, and, more rarely, sands and clay.	Bedrock is predominantly limestone and basalt, but shales, pillow lava, silicate, Mediterranean terra rosa, karst, schist, and sandstone are also reported. Soil texture varies widely; often on clay- and sandy loam, (sandy) clay, and gravel; more	Predominantly on soils with a limestone bedrock, far less on basalt or sandstone. Recorded soil textures include clay- and sandy loams, less often more pure clay or loam. Growth on very poor, stony soils.

		sand and highly organic soil such as turf.		rarely on loss, pure sands, and marly soils.	
Climate	Annual rainfall data of 225-800 mm indicate some drought tolerance, but it also occurs in areas with as much as 1250 mm.	Wide annual rainfall amplitude, varying from less than 100 mm up to 1100 mm.	Rainfall data vary from 450 to 750 mm, and in some sites it can be as high as 1400 mm.	Wide annual rainfall amplitude, varying from 125 mm up to 1400 mm.	Rainfall data vary widely: from less than 100 mm up to 600 mm, but most are from the range 200-350 mm.
Mating system	Considered largely autogamous (Hammer, 1980)	Considered largely autogamous (Hammer, 1980) but mixed mating has been observed (Arrigo et al., 2011).	Considered largely autogamous (Hammer, 1980) but mixed mating has been observed (Arrigo et al., 2011).	Considered largely autogamous (Hammer, 1980) but mixed mating has been observed (Arrigo et al., 2011).	Considered largely autogamous (Hammer, 1980) but it can occasionally hybridize with wheat (Van Slageren, 1994).

Number of *Aegilops* accessions and external sources with coordinates included in the study before and after clearing spatial duplicates and filtering by the geo-referencing quality threshold.

	Ae. biuncialis		*Ae. geniculata*		*Ae. neglecta*		*Ae. triuncialis*		*Ae. ventricosa*	
	Germplasm accesions	Occurrence records from external sources	Germplasm accesions	Occurrence records from external sources	Germplasm accesions	Occurrence records from external sources	Germplasm accesions	Occurrence records from external sources	Germplasm accesions	Occurrence records from external sources
Initial number of georeferenced records	6	30	144	4850	33	870	191	1674	26	363
Number of records removed for having low accuracy*	0 (0%)	13 (43%)	0 (0%)	1292 (27%)	0 (0%)	372 (43%)	0 (0%)	580 (35%)	0 (0%)	99 (27%)
Number of records removed for been considered spatial duplicates	0 (0%)	1 (3%)	20 (14%)	1879 (39%)	1 (3%)	71 (8%)	12 (6%)	221 (13%)	2 (8%)	63 (17%)
Number of non-duplicated records with TOTALQUAL≤80	2 (33%)	1 (3%)	5 (3%)	317 (7%)	3 (9%)	73 (8%)	5 (3%)	148 (9%)	5 (19%)	43 (12%)
Number of non-duplicated records with TOTALQUAL>80	4 (67%)	15 (50%)	119 (83%)	1362 (28%)	29 (88%)	354 (41%)	174 (91%)	725 (43%)	19 (73%)	158 (44%)

4   *Number of records with geographic coordinates expressed in decimal degrees with less than two decimals in both latitude and
5   longitude or without textual information on the occurrence site, plus records with geographic coordinates expressed in UTM with
6   lower resolution than 1x1 km.

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).

Species	Bioclimatic component	Geophysic component	Edaphic component
Ae. biuncialis	1. January minimum temperature 2. Precipitation of the wettest month 3. Isothermality	1. Northness 2. Eastness	1. Topsoil salinity 2. Topsoil reference bulk density 3. Topsoil base saturation
Ae. geniculata	1. Minimum temperature of the coldest month 2. Annual temperature range 3. Isothermality	1. Slope 2. Altitude	1. Topsoil gravel content 2. Topsoil organic carbon 3. Topsoil sand fraction
Ae. neglecta	1. March precipitation 2. May maximum temperature 3. Isothermality	1. Altitude 2. Northness	1. Topsoil total exchangeable bases 2. Reference depth of the soil unit 3. Topsoil salinity
Ae. triuncialis	1. April minimum temperature 2. April precipitation 3. Isothermality	1. Eastness 2. Altitude	1. Topsoil clay fraction 2. Topsoil organic carbon 3. Topsoil base saturation
Ae. ventricosa	1. Mean temperature of the coldest quarter 2. April precipitation 3. Temperature seasonality	1. Altitude 2. Slope	1. Topsoil sodicity 2. Topsoil reference bulk density 3. Topsoil base saturation

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.

	Ae. biuncialis	*Ae. geniculata*	*Ae. neglecta*	*Ae. triuncialis*	*Ae. ventricosa*	TOTAL
Number of population occurrences from external sources	15	1362	354	725	158	2614
Number of spatial gaps	15	1359	317	722	158	2571
Number of priority ecogeographical gaps	12	140	133	73	35	393
Percentage of population occurrences from external sources identified as priority ecogeographical gaps	80	10	38	10	22	-

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

	Ae. biuncialis	*Ae. geniculata*	*Ae. neglecta*	*Ae. triuncialis*	*Ae. ventricosa*
Number of categories in the ELC map	26	27	27	27	27
Number of ELC categories currently represented in the Spanish Network	2	13	7	11	6
Percentage of ELC categories currently represented in the Spanish Network	8	48	26	41	22
Number of ELC categories in the spatial gaps	9	26	26	22	18
Number of ELC categories in the priority ecogeographical gaps	7	12	19	11	12
Percentage of improvement in ecogeographical representativeness	27	44	70	41	44

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

	Ae. biuncialis	*Ae. geniculata*	*Ae. neglecta*	*Ae. triuncialis*	*Ae. ventricosa*	TOTAL
Priority ecogeographic gaps with a Lang aridity index <40	10	103	76	26	8	223
Priority ecogeographic gaps with a Lang aridity index <40 and with the highest topsoil salinity values	2	21	15	5	2	45

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

Species	Latitude	Longitude	Province	Municipality	Lang index	Topsoil salinity (dS/m)
Ae. neglecta	41.700833	-0.045000	Huesca	Villanueva de Sigena	29	2.1
Ae. neglecta	42.060000	-0.460000	Huesca	NA	36	2.1
Ae. neglecta	41.530000	-0.840000	Navarra	El Burgo de Ebro	26	0.8
Ae. neglecta	41.430000	-0.360000	Zaragoza	Pina de Ebro	27	0.8
Ae. neglecta	41.430000	-0.720000	Zaragoza	NA	28	0.8
Ae. neglecta	41.420000	-0.240000	Zaragoza	Bujaraloz	29	0.8
Ae. neglecta	41.510000	-0.240000	Zaragoza	La Almolda	29	0.8
Ae. geniculata	38.640000	-0.900000	Alicante	Villena	27	0.7
Ae. geniculata	36.736490	-4.118400	Málaga	Vélez-Málaga	23	0.7
Ae. neglecta	41.430000	-0.480000	Zaragoza	NA	23	0.7
Ae. neglecta	41.520000	-0.600000	Zaragoza	Osera de Ebro	23	0.7
Ae. geniculata	39.137000	-0.512550	Valencia	Alberic	24	0.7
Ae. geniculata	39.227020	-0.509370	Valencia	Alginet	25	0.7
Ae. geniculata	39.936200	-0.037130	Castellón	Vila-real	25	0.7
Ae. geniculata	39.450000	-0.440000	Valencia	Burjassot	25	0.7
Ae. geniculata	39.494530	-0.383580	Valencia	Valencia	25	0.7
Ae. geniculata	39.671850	-0.260330	Valencia	Sagunt	26	0.7
Ae. geniculata	39.260000	-0.330000	Valencia	Valencia	26	0.7
Ae. geniculata	38.603590	-0.875550	Alicante	Villena	26	0.7
Ae. geniculata	39.857140	-0.486740	Castellón	Segorbe	29	0.7
Ae. geniculata	38.856240	-0.061490	Alicante	Pego	29	0.7
Ae. geniculata	40.487170	0.463370	Castellón	Vinaròs	30	0.7
Ae. geniculata	38.810000	0.173000	Alicante	Jávea	33	0.7
Ae. neglecta	41.710000	-4.680000	Valladolid	Cabezón de Pisuerga	34	0.7
Ae. geniculata	36.706040	-4.610470	Málaga	Cártama	34	0.7
Ae. neglecta	42.270278	3.144444	Gerona	Roses	36	0.7
Ae. neglecta	41.360000	2.040000	Barcelona	Barcelona	37	0.7
Ae. neglecta	42.080000	-4.570000	Palencia	Monzón de Campos	37	0.7
Ae. biuncialis	37.021190	-4.528030	Málaga	Antequera	39	0.7
Ae. geniculata	38.691500	-0.757980	Alicante	Beneixama	30	0.7
Ae. geniculata	39.540000	-0.550000	Valencia	Bétera	26	0.6
Ae. geniculata	37.930000	-1.170000	Murcia	Murcia	17	0.6
Ae. ventricosa	38.869280	-6.637140	Badajoz	Lobón	30	0.6
Ae. neglecta	41.330000	-0.250000	Zaragoza	Bujaraloz	26	0.6
Ae. geniculata	39.046980	-0.515710	Valencia	Villanueva de Castellón	25	0.6
Ae. neglecta	41.710000	-0.830000	Zaragoza	NA	25	0.6
Ae. geniculata	39.584530	-0.380200	Valencia	Valencia	26	0.6
Ae. geniculata	36.727000	-4.405000	Málaga	Málaga	28	0.5
Ae. ventricosa	38.414300	-0.423030	Alicante	El Campello	20	0.4
Ae. triuncialis	41.790000	0.710000	Lérida	Castellón de Farfaña	35	0.4

Ae. biuncialis	36.850000	-2.330000	Almería	Almería	12	0.3
Ae. triuncialis	38.384490	-2.804990	Jaén	La Puerta de Segura	27	0.3
Ae. triuncialis	40.030000	-0.250000	Castellón	Onda	27	0.3
Ae. triuncialis	39.419940	-1.199440	Valencia	Requena	28	0.3
Ae. triuncialis	39.190000	-1.490000	Albacete	Casas Ibáñez	27	0.3

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

	Ae. biuncialis	*Ae. geniculata*	*Ae. neglecta*	*Ae. triuncialis*	*Ae. ventricosa*
Number of categories in the ELC map	26	27	27	27	27
Number of ELC categories currently represented in the Spanish Network	2	13	7	11	6
Percentage of ELC categories currently represented in the Spanish Network	8	48	26	41	22
Number of ELC categories in the spatial gaps	9	26	26	22	18
Number of populations in the PC subset	2	21	15	5	2
Number of ELC categories of the populations in the PC subset	2	3	3	2	2
Percentage of improvement in ecogeographical representativeness	8	11	11	7	7