



**Universitat de Lleida**

Document downloaded from:

<http://hdl.handle.net/10459.1/65073>

The final publication is available at:

<https://doi.org/10.1007/s00122-007-0665-3>

Copyright

(c) Springer Verlag, 2008

1 **Patterns of genetic and eco-geographical diversity in Spanish barleys**

2  
3 S. Yahiaoui, E. Igartua, M. Moralejo, L. Ramsay, J.L. Molina-Cano, F.J. Ciudad, J.M. Lasa,  
4 M.P. Gracia, A.M. Casas

5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18 S.Yahiaoui, E. Igartua, J.M. Lasa, M.P. Gracia, A.M. Casas  
19 Department of Genetics and Plant Production  
20 Aula Dei Experimental Station, CSIC  
21 Av. Montañana, 1005  
22 50059 Zaragoza, Spain  
23 Surface mail: PO Box 202, 50080-Zaragoza  
24 Corresponding author, Ernesto Igartua, [igartua@eead.csic.es](mailto:igartua@eead.csic.es)

25  
26  
27 M. Moralejo, J.L. Molina-Cano  
28 Centre UDL-IRTA, Av Rovira Roure 177,  
29 25006 Lleida, Spain

30  
31  
32 L. Ramsay  
33 Scottish Crop Research Institute, Invergowrie,  
34 Dundee, DD2 5DA, Great Britain

35  
36  
37 F.J. Ciudad  
38 ITA, Junta de Castilla y León  
39 P.O. Box 172.  
40 47071 Valladolid, Spain  
41

1 **Abstract**

2 The pool of Western Mediterranean landraces has been under-utilised for barley breeding so far. The  
3 objectives of this study were to assess genetic diversity in a core collection of inbred lines  
4 derived from Spanish barley landraces; to establish its relationship to barleys from other origins;  
5 and to correlate the distribution of diversity with geographical and climatic factors. To this end,  
6 sixty four SSR were used to evaluate the polymorphism among 225 barley (*Hordeum vulgare*  
7 *ssp. vulgare*) genotypes, comprising two-row and six-row types. These included 159 landraces  
8 from the Spanish Barley Core Collection (SBCC) plus 66 cultivars, mainly from European  
9 countries, as a Reference set. Out of the 669 alleles generated, a large proportion of them were  
10 unique to the six-row Spanish barleys. An analysis of molecular variance revealed a clear  
11 genetic divergence between the six-row Spanish barleys and the Reference cultivars, whereas  
12 this was not evident for the two-row barleys. The genetic information was also used to estimate  
13 population structure. A model-based clustering analysis identified four main populations for the  
14 whole genotype set, and suggested further possible subdivision within two of these populations.  
15 Most of the six-row Spanish landraces clustered into two groups that corresponded to  
16 geographic regions with contrasting environmental conditions. The existence of wide genetic  
17 diversity in Spanish germplasm, possibly related to adaptation to a broad range of  
18 environmental conditions, and its divergence from current European cultivars confirm its  
19 potential as a new resource for barley breeders, and make the SBCC a valuable tool for the  
20 study of adaptation in barley.

21

22

23 **Keywords:** Barley, Landraces, Adaptation, Microsatellites, AMOVA, Population Structure

## 1 **Introduction**

2 For many crops, high-yielding cultivars developed by modern plant breeding programmes have  
3 replaced traditional landraces. This phenomenon, which in turn reduces the genetic base of  
4 current cultivars, is especially true in developed countries (Briggs 1978). Several recent studies  
5 have shown that this is the case for European barley (*Hordeum vulgare* ssp. *vulgare*) (Graner et  
6 al. 1994; Melchinger et al. 1994; Ellis et al. 1997; and Russell et al. 2000) though the loss in  
7 diversity due to modern breeding may have been partially offset by the subsequent introgression  
8 of disease resistances (Koebner et al. 2003). Historical records indicate that genetic erosion may  
9 have occurred as a consequence of the use of a very limited number of landraces and primitive  
10 cultivars in crosses during the earlier stages of modern breeding in Europe (Fischbeck 1992). It  
11 is likely, therefore, that genetic diversity present across original European landraces has not  
12 been fully exploited.

13 In Spain, the National Centre for Plant Genetic Resources holds a collection of over 2000  
14 accessions of cultivated barley, most of which are native landraces collected in the first half of  
15 the 20th century (Lasa et al. 2001). Given their history of selection under Mediterranean  
16 conditions, over a long period of time (i.e., barley cultivation in Spain dates back to prehistoric  
17 times), they may harbour adaptive genes and alleles that have escaped mainstream breeding. An  
18 evaluation of the diversity represented by this genetic resource is necessary in order to facilitate  
19 its use in future cultivar development. To this end, a core collection of the accessions held at the  
20 national repository (Spanish Barley Core Collection or SBCC) was systematically assembled  
21 (Igartua et al. 1998).

22 We chose microsatellite markers, as this system has been used effectively in diversity studies in  
23 barley (Struss and Plieske 1998, Russell et al. 2000; Matus and Hayes 2002; Russell et al. 2003,  
24 Malysheva-Otto et al. 2006; Pandey et al. 2006; Orabi et al. 2007). The main objective of this  
25 study was the characterization of the genetic diversity present in Spanish barleys, by means of

1 molecular markers, and their relationship to the standard breeding gene pool, represented by a  
2 set of reference mostly European cultivars.

3

4

## 5 **Materials and Methods**

### 6 *Plant material*

7 A total of 225 six-row and two-row barley genotypes were included in this study, 159 of which  
8 are inbred lines from the SBCC (Igartua et al. 1998). These inbred lines from the SBCC were  
9 derived from single spikes chosen from each original landrace population, followed by single-  
10 spike selfing for at least four times to ensure high levels of homozygosity (Lasa et al. 2001).  
11 Entries were divided into four groups, according to their geographic origin and row type:  
12 Spanish six-row, Reference six-row, Spanish two-row and Reference two-row (Table 1,  
13 Supplementary Material Table S1). The Spanish six-row group was comprised of 148 SBCC  
14 inbred lines and four cultivars, three of which were derived directly from landraces. The  
15 Reference six-row set comprised 33 genotypes: 27 European, 3 American (Mammuth, Morex,  
16 Steptoe), and 3 from CIMMYT-ICARDA origins (Orria, S-36, and S-45). The Spanish two-row  
17 group consisted of 11 SBCC inbred lines, whereas the Reference two-row set comprised 27  
18 cultivars from other European countries (including some bred in Spain from European parents),  
19 1 American (Logan) and 1 from ICARDA (S-7). The majority of accessions in the six-row  
20 Reference group were chosen because they represented material that formed the ancestors of  
21 most modern European cultivars (Baumer and Cais, 2000). Other cultivars in this group were  
22 either among the most used cultivars in Spain (Steptoe, Barberousse, Dobra, Hatif), or were  
23 parents in the Spanish National Barley Breeding Programme (Plaisant, Logan, Orria, ICARDA  
24 materials). The accessions in the two-row Reference group were either widely used cultivars in

1 Spain or parents in the Breeding Programme. Most of them were European, and provide a  
2 representative sample of European two-row barley diversity.

### 3 *Molecular marker analyses*

4 Samples of leaf tissue were taken from 8-10 plants per genotype, 14 days after sowing in paper-  
5 pots in the greenhouse. DNA extraction was carried out following a CTAB procedure, as  
6 described in Casas et al. (1998). The entire set of 225 genotypes was genotyped for 64  
7 microsatellites, also known as simple sequence repeats (SSRs), (Supplementary Material Table  
8 S2) and for one sequence tagged site (STS) MWG699. This STS is closely linked to *vrs1*, the  
9 gene controlling ear type and that has been proposed as a diagnostic marker for barley origin  
10 (Tanno et al 1999; 2002). SSR primer pair sequences and amplification conditions were  
11 obtained from Pillen et al. (2000), Ramsay et al. (2000) and Macaulay et al. (2001). PCR  
12 amplifications were carried out in a final volume of 15 µl, containing 50 ng of genomic DNA,  
13 1x PCR Buffer (Biotools, Madrid, Spain), 2 mM MgCl<sub>2</sub>, 15 pmol of forward and reverse  
14 primers, dNTPs at 0.2 mM each, and 0.4 U of Tth DNA Polymerase (Biotools, Madrid, Spain).

15 Equal volumes of electrophoresis loading buffer containing 95% formamide were added to the  
16 samples, which were then denatured at 95°C, quickly cooled and electrophoresed in 5%  
17 polyacrylamide gels. A 30-330 bp AFLP Ladder (Invitrogen) was also loaded and products  
18 visualized by silver staining (Bassam et al. 1991). Gels were scanned with a Molecular Imager  
19 FX (Bio-Rad) and product sizes estimated using the Diversity Database software (Bio-Rad).  
20 Two cultivars were included as checks in each gel with these checks being polymorphic for the  
21 markers assayed. A set of 40 cultivars was assayed first for all markers, providing information  
22 on allele diversity and size at each marker. After all samples had been tested, the  
23 polymorphisms found were confirmed by running a set of verification gels for each marker with  
24 genotypes representing all apparent allele sizes.

1 Twelve SSRs, Bmag136, Bmag013, Bmag384, Bmag353, EBmac701, EBmac970, Bmac113,  
2 Bmag223, EBmac806, Bmag206, Bmag120, and Bmag135 were analysed using an ABI 310  
3 automated sequencer (Applied Biosystems). PCR products were run together with the internal  
4 lane size standard GeneScan 500 [TAMRA] according to supplier's instructions. The results  
5 were processed with GeneScan software.

6 Nearly all SSR used produced a single band per genotype. Double bands were observed in low  
7 frequency, and were even rarer among the SBCC landraces, which confirms their  
8 homozygosity. A single SBCC genotype showed double bands, for six SSRs and three cultivars  
9 showed double bands, at 13 SSRs in total. All potential cases of double bands were confirmed  
10 with additional PCR runs. When double bands were detected, the most intense was taken as  
11 representative and used for subsequent analyses.

## 12 *Genetic diversity*

13 The level of polymorphism of each locus was calculated according to Nei (1973), by using the  
14 gene diversity index, also known as polymorphism information content,

$$15 \text{ PIC} = 1 - \sum P_i^2$$

16 in which  $P_i$  is the frequency of the  $i^{\text{th}}$  SSR allele.

17 Manhattan (city block) distances between individuals were calculated with the computer  
18 program NTYSYS pc v. 2.1. (Rohlf 2000). This distance is based on the sum of the absolute  
19 number of repeat differences between genotypes and represents the analogous non-squared  
20 version of the  $(\delta\mu)^2$  distance measure (Goldstein et al. 1995),

$$21 M_{ij} = \frac{1}{n} \sum_k |x_{ki} - x_{kj}|$$

1 where  $x_{ki}$  and  $x_{kj}$  are the repeat sizes of the alleles in the  $i$ th and  $j$ th individuals at the  $k$ th locus,  
2 and  $n$  is the number of loci analysed. The distance matrix was used for a Principal Coordinate  
3 Analysis (PcoA module of NTSYS).

#### 4 *Analysis of molecular variance (AMOVA)*

5 The analysis of the distribution of genetic variation across the germplasm groups established *a*  
6 *priori* was done using the AMOVA option of Arlequin software package (Arlequin software,  
7 Schneider et al. 2000). Fixation statistics ( $F_{ST}$  and  $R_{ST}$ , corresponding to the *infinite allele* and  
8 the *stepwise mutation* models of SSR evolution, IAM and SMM, respectively) were produced  
9 for individual SSRs and groups of germplasm. The significance of the estimates was obtained  
10 through permutation tests, using one thousand permutations. The significance level chosen was  
11 0.0008, which corresponds to a genome-wise significance level of 0.05 for 64 independent tests  
12 (one for each marker), applying a Bonferroni correction.

#### 13 *Analysis of genetic structure*

14 Genotypes were classified into genetic clusters according to molecular markers, using a model-  
15 based approach with the software package STRUCTURE (Falush et al. 2003). Given a value for  
16 the assumed number of subpopulations (clusters), this method assigns genotypes from the entire  
17 sample to clusters in a way that Hardy-Weinberg equilibrium is maximized within clusters, and  
18 linkage disequilibrium is accounted for by differences in allele frequencies among clusters. As  
19 the lines used for this study are homozygous, we used the method to detect exclusively  
20 association between marker loci rather than including within-marker locus variation (Kraakman  
21 et al. 2004). The analyses were done according to the linkage model in STRUCTURE (Falush et  
22 al. 2003).



1 Several runs of STRUCTURE were done by setting the number of subpopulations (K) from 1 to  
2 10. For each run, batches of runs were carried out using burn-in time and replication numbers  
3 set to 10,000 (as several runs with these numbers set at 50,000 gave very similar results).

#### 4 *Distribution of genetic diversity according to geographic and climatic factors*

5 The relationship of the populations defined by cluster analysis with climate and geography was  
6 examined. Factors considered were: altitude, latitude, rainfall, temperature, total  
7 evapotranspiration (ETP), Turc index, Papadakis climatic index, and agroecological region (a  
8 division of geographic zones based on historic barley production, defined in Igartua et al. 1998).  
9 Rainfall, temperature and ETP were calculated as annual means, and as seasonal means: *Autumn*  
10 (mean of October, November and December), *Winter* (January, February, March), and *Spring*  
11 (April, May, June). Turc index is an estimation of the agronomic productivity of a region, based  
12 on correlations between climatic factors and production over a long period of time, and is  
13 expressed as tons of dry matter per hectare of an adapted plant under standard cultivation  
14 (Ruimy et al. 1996). Papadakis climatic index is an international standard for this purpose. It  
15 classifies climates according to the factors that affect crop development most, namely  
16 temperature and humidity (Papadakis 1975). Data were extracted from the SIGA service  
17 (Spanish acronym for Geographic Information System for Agriculture) of the Spanish Ministry  
18 of Agriculture, Fisheries, and Food (<http://www.mapa.es/es/sig/pags/siga/intro.htm>). This site  
19 provides monthly averages of climatic data for 4186 locations over the country (averaged from  
20 1960 until 1996) collected by the National Institute of Meteorology. We had coordinates for the  
21 collection sites of all landrace-derived accessions from the SBCC. For 77 accessions,  
22 collection localities also had weather stations. For the rest, the nearest most similar location  
23 with climatic records was chosen. In 34 cases, climatic stations were less than 10 km away from  
24 collection sites; for an additional 35 cases, weather stations were within a 20 km radius; leaving  
25 only 13 cases where weather stations were over 20 km from the collection sites. When several

1 weather stations were available at similar distances we chose the one that most closely  
2 resembled the collection site in altitude and orientation.

3 Association of markers and alleles with geographic and climatic factors was explored by means  
4 of linear regression analyses. Only alleles that were present in over 5% of the entries analysed  
5 were used for this purpose. Regressions of markers on climatic and geographic factors were  
6 carried out using PROC GLM (SAS 1988). Alleles were introduced in the model as dependent  
7 variables, whereas geographic and climatic factors were the independent variables. Significance  
8 was calculated for the model, which included only one allele, with the significance threshold set  
9 at 0.05, using a Bonferroni correction, as mentioned before. The association to climate types  
10 according to Papadakis index was done by analysis of variance (using similar significance levels  
11 as for the regression analyses).

12

## 13 **Results**

### 14 *SSR diversity*

15 Overall, 669 alleles were found for the 225 genotypes and 64 SSRs (Supplementary Material  
16 Table S2). Null alleles were found at five loci (HvLTPPB, HvDHN7, EBmac806, Bmag135,  
17 and HvGLB2). The number of alleles per locus varied between 2 and 38, with a mean of 10.5  
18 alleles per locus. A sizeable proportion of alleles (34.1%) were restricted to one of the four  
19 germplasm groups. Although differences in sample size need to be considered, the number of  
20 unique alleles was much higher within the group of six-row Spanish barleys (184 out of 228).  
21 Most of these were rare alleles, present at very low frequencies, but nine of them were found in  
22 more than 10% of the individuals of the corresponding germplasm group.

1 Overall the Spanish six-row genotypes were more diverse than the Reference six-row group  
2 (average diversity index across all loci of 0.62 vs 0.58, Supplementary Material Table S2). The  
3 level of diversity in both two-row groups was lower (0.53 and 0.54). A few alleles were fixed at  
4 some groups. For instance, all the genotypes of the Spanish six-row group had the same allele at  
5 the HvHVA1 (136 bp) and HvCMA (133 bp) loci.

6

### 7 *Genetic variance among the four groups of germplasm defined a priori*

8 The analysis of molecular variance for the germplasm groups defined *a priori* revealed  
9 significant  $F_{ST}$  and  $R_{ST}$  parameters, for both six-row and two-row comparisons (Supplementary  
10 Material Table S3). Genetic divergence between the Spanish and Reference six-row groups was  
11 10.4 % for  $F_{ST}$ , whereas it was 21.4% for  $R_{ST}$ . For the comparison between the two-row groups,  
12 the  $F_{ST}$  was similar (11.9%), but  $R_{ST}$  was much lower than the value for between the six-row  
13 groups (7.5%).

14 A principal coordinate analysis was carried out on the Manhattan distance matrix for the 64 loci  
15 and 225 genotypes (Fig. 1). The first axis alone explained 12.7% of the variance, and divided  
16 mainly the Spanish six-row group from the other three germplasm groups. The second axis  
17 explained 5.6% of the marker variance, and separated mostly both two-row groups from the  
18 Reference six-row group. Twenty nine of the 66 cultivars analysed, are included in the  
19 European Barley Core Collection (<http://www.barley.ipk-gatersleben.de/ebdb.php3>), among  
20 them the Spanish cultivars Albacete and Candela. These 29 cultivars are identified in Fig. 1, to  
21 provide anchor points for interpretation of the results.

22 The grouping of genotypes observed in Fig. 1 did not follow entirely the groups made *a priori*.  
23 After these results, we considered appropriate to investigate the genetic structure of the

1 germplasm studied, especially to address the question of whether highly diverse Spanish six-  
2 row barleys constitute a single group.

### 3 *Genetic structure*

4 We detected an underlying structure, with at least four populations, based on the criterion of  
5 maximization of the natural log probability of the data, which is proportional to the posterior  
6 probability of K (Falush et al. 2003). The results did not point to a clear cut-off point for the  
7 number of populations in our sample but, for practical purposes, the K=4 solution seemed  
8 sensible, as the increase of the log tapered off after this value (Fig. 2). At this level, every run  
9 produced almost exactly the same populations, whereas this consistency decreased notably for  
10 values of K above 4. Also, 72% of genotypes were assigned to one of the four populations with  
11 membership probabilities  $\geq 0.75$ . Therefore, this level seemed the smallest value of K that  
12 captured the major structure of the data.

13 The first split divided a majority of six-row Spanish lines from the rest (Fig. 3). The second split  
14 (K=3) separated two populations among the group of Spanish six-row landraces differentiated  
15 in the previous step (colours red and blue Fig. 3). The third step (K=4) separated Reference six-  
16 row genotypes, together with a few Spanish, from all two-row genotypes (grey and green in Fig.  
17 3).

18 The genotypes were spread among the four populations as follows: Population I had 55  
19 genotypes, 31 of them from the Reference six-row group, 6 from the Reference two-row (all of  
20 them winter cultivars), and 18 lines from the Spanish six-row group. Population II comprised 34  
21 genotypes, including all spring and four winter cultivars from the Reference two-row group, 9  
22 lines from the Spanish two-row group, and 2 from the Reference six-row group (Dobla, Olli).  
23 Population III included 50 lines from the Spanish six-row group. Population IV comprised 86  
24 lines, also all of them of the Spanish six-row group.

1 Genetic divergence among populations was highly significant (Table 2).  $F_{ST}$  and especially,  $R_{ST}$   
2 values were higher when the comparison involved populations dominated by genotypes from  
3 the Reference sets (I or II) with populations dominated by SBCC genotypes (III and IV). The  
4 lowest values corresponded to the comparisons between the populations dominated by Spanish  
5 genotypes (III and IV). The second analysis, run separately for populations III and IV, also  
6 indicated a subpopulation structure, more evident for population III (Table 2).

### 7 *Association of groups with geographic and climatic factors*

8 The distribution of the populations revealed in the previous section followed quite closely the  
9 distribution of climates in the Iberian peninsula, after the index of Papadakis (Fig. 4a,b). This  
10 was particularly true for genotypes belonging in the populations III and IV. Population III  
11 genotypes came from areas with Temperate Mediterranean climates, whereas population IV  
12 came from regions with Maritime, Subtropical Mediterranean and Continental Mediterranean  
13 climates. This apparently non-random distribution was confirmed by the analysis of population  
14 averages for a set of ecogeographic variables (Table 3). Population II, with 9 out the 11 SBCC  
15 two-row genotypes, came mostly from inland Northern Spain, with the lowest yearly  
16 temperatures on average, largest overall rainfall, and highest Turc productivity index. There  
17 were also marked differences between the two main populations of SBCC genotypes (III, and  
18 IV). Population IV had the lowest average for altitude, latitude, and spring rainfall, whereas it  
19 had the highest values for temperature, evapotranspiration, and autumn rainfall. The resulting  
20 Turc index was rather high. Overall, this is the profile of a relatively warm area, with sufficient  
21 water in the beginning of the growth cycle, followed by terminal water stress. Population I  
22 averages were intermediate between values of populations III and IV.

23 The analyses of association for single alleles were carried out for a total of 277 alleles with a  
24 frequency greater than 5% among the SBCC genotypes. The climatic variables for *Autumn* are  
25 not shown, as they presented very high correlation coefficients with the corresponding *Winter*

1 variables. About 50% of the markers (31 out of 64 SSRs) showed association of at least one  
2 allele with some geographic or climatic factor (Table 4). The largest number of associations  
3 occurred with *Papadakis climate index* and *Agroecological region*, followed by associations  
4 with *Temperature* and *Evapotranspiration* variables. *Latitude*, *Altitude* and *Precipitation*  
5 variables showed less associations and Turc productivity index had the least number (five)  
6 associated markers (Table 4).

7 The distribution of MWG699 haplotypes over populations of genotypes defined in this study  
8 was clearly different between populations (Table 5). Haplotype D was the most frequent, with  
9 highest frequencies in Populations I and III (Fig. 4b). Conversely, haplotype A was more  
10 frequent in Population IV, and haplotype K was clearly predominant amongst the genotypes of  
11 Population II. All entries with the K haplotype were two-row. The distribution of haplotypes in  
12 the subpopulations of populations III and IV was even more biased.

13

## 14 **Discussion**

### 15 *Genetic diversity in groups defined a priori*

16 The average number of alleles per locus (10.5, Supplementary Material Table S2) was higher  
17 than comparable figures reported in other studies for cultivated barley (Russell et al. 1997;  
18 Macaulay et al. 2001; Karakousis et al. 2003; Sjakste et al. 2003). Although many markers are  
19 in common among these studies and ours, they surveyed smaller samples than that of the current  
20 study. The level of diversity found in our sample is closer to values reported in studies with  
21 populations of *Hordeum spontaneum* (Ivandic et al. 2002; Baek et al. 2003), or large diverse  
22 sets of cultivated barley (Russell et al. 2000; Matus and Hayes 2002). Thus, we can conclude  
23 that we are studying a sample of *Hordeum vulgare* types with considerable polymorphism.  
24 Notably, we had 14 SSRs in common with the worldwide survey of cultivated barley diversity  
25 carried out by Malysheva-Otto et al. (2006). In the present study, the average number of alleles

1 and PIC for these 14 loci were 12.1 and 0.72, respectively, compared with 18.6 and 0.79 for the  
2 worldwide study.

3 On average, SSRs derived from genic sequences were less polymorphic than SSRs from random  
4 genomic clones (6.5 vs 12.1 alleles per locus). Interestingly, most of the gene-derived markers  
5 in our study showed higher diversity values than previously reported for a set of elite German  
6 cultivars (Pillen et al. 2000). The high values of diversity detected were mostly due to the large  
7 number of alleles (591) present in the six-row Spanish group, 34.1% of which were private  
8 alleles. As pointed out by Matus and Hayes (2002), the presence of so many unique alleles  
9 could be an indication of the relatively high rate of mutation at SSR loci, or could also point to  
10 the existence of exotic germplasm that could be a reservoir for novel alleles for crop  
11 improvement. Thus, the high diversity of the Spanish six-row accessions, coupled with the high  
12 number of unique alleles, could be explained by a rather long history of isolation from other  
13 European countries and concurrent genetic drift or selection for adaptation to local constraints.  
14 The generally low frequency of most private alleles is consistent with a genetic drift explanation  
15 but the presence of some private alleles with high frequencies in Spanish barleys, however,  
16 suggests the effect of selection pressure. The distribution of five of the nine private alleles  
17 present in high frequencies in the Spanish six-row group was related to geographical factors.  
18 The association of the distribution of genetic diversity with geographic patterns also points at  
19 the presence of selection pressure favouring alleles associated with better local adaptation  
20 (Tables 3, 4).

21 The molecular analysis of variance revealed a remarkable genetic divergence between the  
22 Spanish and Reference sets.  $F_{ST}$  values among them were similar to values found by Maestri et  
23 al. (2002) and Koebner et al. (2003) for comparisons between winter and spring barleys, two  
24 quite distinct germplasm groups. The divergence between Spanish and Reference groups was  
25 similar for both the two-row and six-row barleys, using an IAM (measured by the  $F_{ST}$ ,

1 Supplementary Material Table S3). However using a SMM ( $R_{ST}$  statistic), the Spanish and  
2 Reference six-row barleys would be more distinct than the two-row groups. This was caused by  
3 the fact that differences in allele frequencies between groups were similar in number for both  
4 row types but, for the six-row groups, the alleles with different frequencies among groups were  
5 also more distant in size. Accordingly, there were more SSRs clearly discriminating between the  
6 Spanish and Reference sets for the six-row than for the two-row groups (Supplementary  
7 Material Table S3). The level of genetic differentiation between Spanish and Reference sets was  
8 not as high as found among barley landraces from Syria and Jordan ( $R_{ST} = 32.04$ ) (Russell et al.  
9 2003) but were close to the values found between populations of wild barley from different  
10 countries ( $F_{ST} = 7.75\sim 10.54$ , and  $R_{ST} = 8.58 \sim 10.59$ ) (Ivandic et al. 2002). For some markers,  
11 both  $F_{ST}$  and  $R_{ST}$  indices were significant, whereas for others only one of the two statistics,  
12 usually  $F_{ST}$ , was significant, as found also by Ivandic et al. (2002), who recommended the use of  
13 both indices to provide maximum information on allele differentiation among groups of  
14 genotypes.

15 The evidence suggests that the Spanish six-row barleys are more distinct from their Reference  
16 counterparts than the two-row types. The principal coordinate analysis supported this, as about  
17 half of the Spanish two-row barleys clustered mostly with cultivars from the Reference two-row  
18 set (Fig. 1), especially with spring cultivars (Beka, Triumph, and Alexis). Other Spanish two-  
19 row entries clustered closer to the six-row barleys, intermediate between the Spanish and  
20 Reference sets. Nevertheless, the observations on the Spanish two-row set cannot be conclusive,  
21 because of the small sample size. This size was set to be proportional to their prevalence in the  
22 original collection of Spanish barleys (Igartua et al., 1998).

23 The Spanish six-row genotypes generally formed a distinct group, with only a few genotypes  
24 clustering with genotypes of the Reference sets (about 16 in total). The distinct cluster of  
25 Spanish six-row genotypes, however, showed a remarkable internal diversity that was



1 comparable to the two Reference sets combined, given the scatter of their respective points on  
2 the principal coordinate analysis (Figure 1).

3 Interestingly, the Mediterranean cultivar Athenais (from Greece) appeared to show a relatively  
4 higher degree of genetic relatedness to Spanish accessions compared with other non-  
5 Mediterranean cultivars.

### 6 *Genetic Structure*

7 The use of the STRUCTURE clustering algorithm allowed the identification of subpopulations  
8 of genotypes, based on their genetic similarity. This procedure has been used for clustering of  
9 collections of inbred lines in maize (Remington et al. 2001, Jung et al. 2004), Arabidopsis  
10 (Olsen et al. 2004), rice (Garris et al. 2003), wheat (Chao et al. 2007) and barley (Pandey et al.  
11 2006; Morrell and Clegg 2007), among other crops. In all these studies, authors found evidence  
12 of population substructure though Kraakman et al. (2004) did not find any substructure among a  
13 collection of recent spring barley cultivars. In our study, we have clearly and consistently  
14 identified four populations, two dominated by Spanish six-row entries, and two clearly formed  
15 around the Reference sets.

16 The two main Spanish populations (III and IV) were identified by the STRUCTURE analysis at  
17  $K=3$ , prior to the separation of populations I and II, which comprised all the entries from the  
18 Reference sets ( $K=4$ ). This does not necessarily imply that the two main Spanish populations  
19 presented larger genetic divergence than populations I and II (basically two-row vs. six-row  
20 Reference sets) as the population sizes were not equal and this may well have affected the order  
21 of population identification. Indeed the  $F_{ST}$  and  $R_{ST}$  fixation indices (Table 2) showed that the  
22 genetic divergence between populations III and IV was slightly less than that between  
23 populations I and II. It is still remarkable, however, that the split between the two large Spanish  
24 populations and the two populations dominated by Reference sets entries are roughly

1 comparable as the Reference sets were very diverse, including cultivars of all growth and row  
2 types (spring and winter, two-row and six-row) and represented the range of germplasm that  
3 largely underpins current European barley cultivars. The AMOVA of the four populations  
4 derived from the STRUCTURE analysis confirmed the genetic divergence among them. The  
5 range of values observed for  $F_{ST}$  and  $R_{ST}$  among populations was similar to the range found in a  
6 comparison of landraces from Jordan and Syria by Russell et al. (2003).

7

#### 8 *Association of populations with geographic and climatic factors*

9 The two main Spanish populations were distributed according to geographic patterns, and  
10 roughly following a North-South direction (Fig. 4), though some groupings independent from  
11 geography were also observed.

12 A few studies have shown that correlations of climatic factors with genetic diversity assessed  
13 with SSRs are prevalent in *H. spontaneum*. (Turpeinen et al. 2001; Baek et al. 2003; Ivandic et  
14 al. 2002, 2003). Nevo et al. (2005) even suggested a possible adaptive role for the SSRs  
15 themselves, in a situation dominated by abiotic stresses. Some of the associations found by  
16 these authors were confirmed in the present study (Table 4). This was true for 6 markers  
17 (HvM62, HvLTPPB, Bmag369, Bmag378, HvBTAI3, HvM67). In another two cases, we  
18 detected similar associations to those found by Ivandic et al. (2002), but using different markers  
19 in the same regions. In this study HvM62 and Bmag369 alleles correlated with temperature and  
20 rainfall, whereas Ivandic et al.(2002) found Bmac29 (close to HvM62) and Bmac273 and  
21 Bmag120 (flanking Bmag369) were associated with humidity.

22 The main climatic factors affecting the distribution of barley cultivars are temperature and water  
23 availability (Morris et al. 1991). Accordingly, we divided the markers associated with climatic  
24 and ecogeographic factors in five groups: markers with association to temperature and rainfall

1 distribution; markers related to only one of these two factors; markers related to other climatic  
2 factors; markers related only to complex geographic indexes (Table 4). As climatic factors were  
3 themselves correlated it is difficult to discriminate the effect of single factors however by  
4 looking at the number of associations, temperature seemed to be the single climatic factor which  
5 most affected the distribution of marker alleles in the SBCC landraces.

6 Thus, adaptation to ecogeographic factors could be one of the causes of the observed  
7 subpopulation structure of the SBCC. It cannot be concluded, however, that there are loci for  
8 adaptation linked to the SSRs whose distribution is associated with geography and climate. The  
9 genetic diversity of this sample of germplasm was clearly stratified in populations (Fig. 3), and  
10 these populations were seemingly distributed along eco-geographic gradients (Table 3). Thus  
11 the distribution of entries from populations III, IV-I, and IV-II followed a gradient of  
12 agroecological conditions occurring in the barley growing area of the Iberian peninsula,  
13 following the main climatic clines. Therefore, any marker that had uneven distribution across  
14 genotypic subpopulations would necessarily appear associated to geographic or climatic factors  
15 for which subpopulations also differ. Actually, there were similarities between locus by locus  
16  $F_{ST}$  (not shown) calculated for the four populations presented in Table 2, and the strength of loci  
17 association with eco-geographic traits. This was especially true when the analysis was based  
18 only on populations III and IV which included the majority of SBCC entries, and showed  
19 clearly distinct eco-geographic distributions (Table 3). Markers (31) that showed association  
20 with eco-geographic factors showed an average  $F_{ST}$  of 0.18 between these populations, which  
21 was significantly ( $P < 0.0001$ ) larger than the average  $F_{ST}$  value of 0.05 for the rest of the marker  
22 loci. For these two populations, the correlation coefficient of  $F_{ST}$  with number of associations  
23 was 0.64. The reason for uneven distribution of marker alleles over subpopulations could be the  
24 selection for adaptation to local environments (Russell et al. 2003) but could also be due to  
25 incomplete admixture of populations from different origins, with different phylogenetic  
26 histories. Further efforts to elucidate the causes of this biased marker distribution will need to

1 focus on the existence of unlinked gene complexes, distribution of functional polymorphisms,  
2 and search of molecular signatures of selection.

3 All evidence found points to the existence of at least two large distinct populations of Spanish  
4 six-row barleys (III and IV), spread over different environments. In order to investigate whether  
5 these populations have a different origin all material in this study were genotyped with the STS  
6 MWG699, which was proposed by Tanno et al. (1999, 2002) as a marker of barley  
7 domestication. The marker shows three haplotypes, named A, D, and K, the last one being only  
8 reported in two-row barleys. Tanno et al. (2002) found the A haplotype widely spread, whereas  
9 D was confined to the Mediterranean region, though Casas et al. (2005) found a more widely  
10 spread distribution of D haplotype over Europe and a possible association of this marker with  
11 plant growth type.

12 The distribution of MWG699 haplotypes over the populations and subpopulations found in this  
13 study reflects the population structure and the distribution of the SSRs used to identify the  
14 populations (Table 5). If this marker does reflect domestication history, then the evidence found  
15 suggests the presence of two different origins for SBCC barleys, that may be related to the  
16 influx of different human populations coming into the Iberian peninsula in historic or prehistoric  
17 times. Alternatively, there could be a substratum of barleys in Spain that originate from a  
18 domestication event in the Western Mediterranean region (possibly represented by the  
19 MWG699 D haplotype), as several authors propose a polyphyletic origin for barley (Komatsuda  
20 et al. 1994; Morrell and Clegg 2007; Orabi et al. 2007), and some evidence points to a Western  
21 Mediterranean centre of origin or diversity for barley (Moralejo et al. 1994; Molina-Cano et al.  
22 1987, 2005). Interestingly both Casas et al. (2005) and Tanno et al. (2002) found that the  
23 MWG699 D haplotype was present in some *H. spontaneum* accessions from Morocco.

24 In conclusion, we propose the hypothesis that the SBCC genotypes have at least two different  
25 origins, and that the distribution of their original landrace populations over the Iberian peninsula

1 followed patterns of adaptation to local conditions. This adaptation may have been due to  
2 distinct fitness of founder populations to different climates, to new variability created through  
3 admixture and recombination of original populations, or to specific adaptations that developed  
4 locally. Further studies on the SBCC should shed light on the causes of adaptation, specifically  
5 on traits and genes that led to its distribution. Such studies will also facilitate the utilization of  
6 this important genetic resource as a source of useful novel alleles in future breeding  
7 programmes.

## 8 **Acknowledgements**

9 This research was funded by project RTA01-088-C3, granted by the INIA (Instituto Nacional de  
10 Investigación y Tecnología Agraria y Alimentación), of the Spanish Ministry of Science and  
11 Technology, and co-funded by the European Regional Development Fund. Samia Yahiaoui was  
12 supported by a scholarship from the AECI (Agencia Española de Cooperación Internacional), of  
13 the Spanish Ministry of Foreign Affairs.

## 14 **References**

- 15 Baek HJ, Beharav A, Nevo E (2003) Ecological-genomic diversity of microsatellites in wild  
16 barley, *Hordeum spontaneum*, populations in Jordan, *Theor Appl Genet* 106:397-410  
17 Bassam B J, Caetano-Anolles G, Gresshoff P M (1991) Fast and sensitive silver staining of  
18 DNA in polyacrilamide gels. *Anal Biochem* 196:80-83  
19 Baumer M, Cais R (2000) Abstammung der Gerstensorten. Ed. Bayerische Landesanstalt für  
20 BodenKultur und Pflanzenbau. 109 p.  
21 Briggs D E (1978) Barley. Chapman & Hall, London, pp 445-480  
22 Casas AM, Igartua E, Vallés MP, Molina-Cano JL (1998) Genetic diversity of barley cultivars  
23 grown in Spain, estimated by RFLP, similarity and coancestry coefficients. *Plant Breed*  
24 117:429-435  
25 Casas AM, Yahiaoui S, Ciudad F, Igartua E (2005) Distribution of MWG699 polymorphism in  
26 Spanish European barleys. *Genome* 48:41-45  
27 Chao S, Zhang W, Dubcovsky J, Sorrells M (2007) Evaluation of genetic diversity and genome-  
28 wide linkage disequilibrium among U.S. wheat (*Triticum aestivum* L) germplasm  
29 representing different market classes. *Crop Sci* 47:1018-1030.  
30 Ellis RP, McNicol JW, Baird E, Booth A, Lawrence P, Powell W (1997) The use of AFLPs to  
31 examine genetic relatedness in barley. *Mol Breed* 3:359-369  
32 Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus  
33 genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567-1587

- 1 Fischbeck G (1992) Barley cultivar development in Europe – Success in the past and possible  
2 changes in the future. Barley Genetics VI, Vol. II (ed. L. Munck), Munksgaard  
3 International Publishers Ltd., Copenhagen, Denmark, pp. 885-901
- 4 Garris AJ, McCouch SR, Kresovich S (2003). Population structure and its effect on haplotype  
5 diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.).  
6 Genetics 165:759-769
- 7 Graner A, Ludwig WF, Melchinger AE (1994) Relationships among European barley  
8 germplasm: II. Comparison of RFLP and pedigree data. Crop Sci. 34:1199-1205
- 9 Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of  
10 genetic distances for use with microsatellite loci. Genetics 139:463-471
- 11 Igartua E, Gracia MP, Lasa JM, Medina B, Molina-Cano JL, Montoya JL, Romagosa I (1998)  
12 The Spanish barley core collection. Genet Resour Crop Ev 45:475-481
- 13 Ivandic V, Hackett CA, Nevo E, Keith R, Thomas WTB, Forster B (2002) Analysis of simple  
14 sequence repeats (SSRs) in wild barley from the Fertile Crescent: associations with  
15 ecology, geography and flowering time. Plant Mol Biol 48:511-527
- 16 Ivandic V, Thomas WTB, Nevo E, Zhang Z, Forster B (2003) Associations of simple sequence  
17 repeats with quantitative trait variation including biotic and abiotic stress tolerance in  
18 *Hordeum spontaneum*. Plant Breed 122: 300-304
- 19 Jung M, Ching A, Bhattraramakki D, Dolan M, Tingey S, Morgante M, Rafalski A (2004)  
20 Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus  
21 in elite maize germplasm. Theor Appl Genet 109:681-689
- 22 Karakousis A, Barr AR, Chalmers KJ, Ablett GA, Holton TA, Henry RJ, Lim P, Langridge P  
23 (2003) Potential of SSR markers for plant breeding and variety identification in  
24 Australian barley germplasm. Aust J Agr Res, 54:1197-1210
- 25 Koebner RMD, Donini P, Reeves JC, Cooke RJ, Law JR (2003) Temporal flux in the  
26 morphological and molecular diversity of UK barley. Theor Appl Genet 106:550-558
- 27 Komatsuda T, Maxim P, Sentil N, Mano Y (2004) High-density AFLP map of nonbrittle rachis  
28 1 (*btr1*) and (*btr2*) genes in barley (*Hordeum vulgare* L.). Theor Appl Genet 109:986-995
- 29 Kraakman ATW, Niks RE, Van der Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage  
30 disequilibrium mapping of yield and yield stability in modern spring barley cultivars.  
31 Genetics 168:435-446
- 32 Lasa JM, Igartua E, Ciudad FJ, Codesal P, García EV, Gracia MP, Medina B, Romagosa I,  
33 Molina-Cano JL, Montoya JL (2001) Morphological and agronomical diversity patterns in  
34 the Spanish barley core collection. Hereditas 135:217-225
- 35 Macaulay M, Ramsay L, Powell W, Waugh R (2001) A representative, highly informative  
36 ‘genotyping set’ of barley SSRs. Theor Appl Genet 102:801-809
- 37 Maestri E, Malcevski A, Massari A, Marmioli N (2002) Genomic analysis of cultivated barley  
38 (*Hordeum vulgare*) using sequence-tagged molecular markers. Estimates of divergence  
39 based on RFLP and PCR markers derived from stress-responsive genes, and simple-  
40 sequence repeats (SSRs). Mol Genet Genomics 267:186-201
- 41 Malysheva-Otto LV, Ganal MW, Röder MS (2006) Analysis of molecular diversity, population  
42 structure and linkage disequilibrium in worldwide cultivated barley germplasm (*Hordeum*  
43 *vulgare* L.). BMC Genetics 7:6-19
- 44 Matus IA, Hayes PM (2002) Genetic diversity in three groups of barley germplasm assessed by  
45 simple sequence repeats. Genome 45:1095-1106
- 46 Melchinger AE, Graner A, Singh M, Messmer MM (1994) Relationships among European  
47 barley germplasm: I. Genetic diversity among winter and spring cultivars revealed by  
48 RFLPs. Crop Sci. 34:1191-1199
- 49 Molina-Cano JL, Fra-Mon P, Salcedo G, Aragoncillo C, Roca de Togores F, García-Olmedo F  
50 (1987) Morocco as a possible center for barley: biochemical and agromorphological  
51 evidence. Theor Appl Genet 98:913-918

- 1 Molina-Cano JL, Russell JR, Moralejo MA, Escacena JL, Arias G, Powell W (2005)  
2 Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. *Theor*  
3 *Appl Genet* 110:613-619
- 4 Moralejo M, Romagosa I, Salcedo G, Sánchez-Monge R, Molina-Cano JL (1994) On the origin  
5 of Spanish two-rowed barleys. *Theor Appl Genet* 87:829-836.
- 6 Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (*Hordeum*  
7 *vulgare*) east of the Fertile Crescent. *PNAS* 104:3289-3294
- 8 Morris ML, Belaid A, Byerlee D (1991) Wheat and barley production in rainfed marginal  
9 environments of the developing world. Part 1 of *1990-91 CIMMYT World Wheat Facts*  
10 *and Trends: Wheat and Barley production in Rainfed Marginal Environments of the*  
11 *Developing World*. Mexico, DF: CIMMYT.
- 12 Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci, USA*  
13 70:3321-3323
- 14 Nevo E, Beharav A, Meyer RC, Hackett CA, Forster BP, Russell JR, Powell W (2005) Genomic  
15 microsatellite adaptive divergence of wild barley by microclimatic stress in 'Evolution  
16 Canyon', Israel. *Biol J Linn Soc* 84:205-224
- 17 Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weining C, Schmitt J Purugganan MD (2004).  
18 Linkage disequilibrium mapping of *Arapidosis CRY2* flowering times alleles. *Genetics*  
19 167:1361-1369
- 20 Orabi J, Backes G, Wolday A, Yahyaoui A, Jahoor A (2007) The Horn of Africa as a centre of  
21 barley diversification and a potential domestication site. *Theor Appl Genet* 114:1117-  
22 1127
- 23 Pandey M, Wagner C, Friedt W, Ordon F (2006) Genetic relatedness and population  
24 differentiation of Himalayan hulless barley (*Hordeum vulgare* L.) landraces inferred with  
25 SSRs. 167:1361-1369. *Theor Appl Genet* 113:715-729
- 26 Papadakis J (1975) *Climates of the world and their agricultural potentialities*. Edición  
27 Argentina, Buenos Aires, 200pp.
- 28 Pillen K, Binder A, Kreuzkam B, Ramsay L, Waugh R, Förster J, León J (2000) Mapping new  
29 EMBL-derived barley microsatellites and their use in differentiating German barley  
30 cultivars. *Theor Appl Genet* 101:652-660
- 31 Ramsay L, Macaulay M, degli Ivanissevich S, MacLean K, Cardle L, Fuller J, Edwards KJ,  
32 Tuvešson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganal M,  
33 Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley.  
34 *Genetics* 156:1997-2005
- 35 Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S,  
36 Goodman MM, Buckler IV ES (2001) Structure of linkage disequilibrium and phenotypic  
37 associations in the Maize genome. *Proc Natl Acad Sci, USA* 98: 11479-11484.
- 38 Rohlf FJ (2000) *NTSYSpc. Numerical Taxonomy and Multivariate Analysis System*. Version  
39 2.1. Exeter Software, New York
- 40 Ruimy A, Dedieu G, Saugier B (1996) TURC: A diagnostic model of continental gross primary  
41 productivity and net primary productivity. *Global Biogeochemical Cycles* 10 (2): 269-  
42 285.
- 43 Russell J, Fuller J, Young G, Thomas B, Taramino G, Macaulay M, Waugh R, Powell W (1997)  
44 Discriminating between barley genotypes using microsatellite markers. *Genome* 40:442-  
45 450
- 46 Russell JR, Ellis RP, Thomas WTB, Waugh R, Provan J, Booth A, Fuller J, Lawrence P, Young  
47 G, Powell W (2000) A retrospective analysis of spring barley germplasm development  
48 from 'foundation genotypes' to currently successful cultivars. *Mol Breed* 6:553-568
- 49 Russell JR, Fuller J, Baum M, Ceccarelli S, Grando S, Powell W (2003) Patterns of  
50 polymorphism detected in the chloroplast and nuclear genomes of barley landraces  
51 samples from Syria and Jordan. *Theor Appl Genet* 107 :413-421.
- 52 SAS Institute. 1988. *SAS/ STATTM User's Guide*, Release 6.03. Edition. Cary, NC: SAS  
53 Institute Inc., 1028p

1 Schneider S, Roessli D, Excoffier L (2000) Arlequin v. 2.000. A software for population genetic  
2 data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland  
3 Sjakste TG, Rashal I, Röder MS (2003) Inheritance of microsatellite alleles in pedigrees of  
4 Latvian barley varieties and related European ancestors. Theor Appl Genet 106:539-549  
5 Struss D, Plieske J (1998) The use of microsatellite for detection genetic diversity in barley  
6 populations. Theor Appl Genet 97:308-315  
7 Tanno K, Takaiwa F, Oka S, Komatsuda T (1999) A nucleotide sequence linked to the vrs1  
8 locus for studies of differentiation in cultivated barley (*Hordeum vulgare* L.). Hereditas  
9 130:77-82.  
10 Tanno K, Taketa S, Takeda K, Komatsuda T (2002) A DNA marker closely linked to the vrs 1  
11 locus (row type gene) indicates multiple origins of six rowed cultivated barley. (*Hordeum*  
12 *vulgare* L.). Theor Appl Genet 104:54-60  
13 Turpeinen T, Tenhola T, Manninen O, Nevo E, Nissilä E (2001) Microsatellite diversity  
14 associated with ecological factors in *Hordeum spontaneum* populations in Israel. Mol  
15 Ecol 10:1577-1591



Table 1. Barley genotypes analysed (detailed information on cultivars in Supplementary material Table S1).

Plant material	N° of accessions
<u>Six-row</u>	
• Spanish	
- Inbred lines derived from local landraces (SBCC)	148
- Cultivars	
<i>Albacete, Almunia, Candela, Pané</i>	4
• Reference set	
<i>Ager, Asplund, Athenais, Athene, Banteng, Barberousse, Bordia, Dea, Dobra, Dura, Frisia, Gerbel, Hatif de Grignon, Hauter, Herfodia, Juli, Mammuth, Maskin, Mirco, Monlon, Morex, Olli, Orria, Plaisant, Ragusa, Senta, Steptoe, S-36, S-45, Tapir, Vega Svalöf, Vindicat, Vogelsanger Gold</i>	33
<u>Two-row</u>	
• Spanish	
- Inbred lines derived from local landraces (SBCC)	11
• Reference set	
<i>Albaicín, Alexis, Alpha, Angora, Beka, Camelot, Cameo, Clarine, Gaelic, Graphic, Hassan, Hispanic, Igri, Kym, Labea, Logan, Mogador, Nevada, Pallas, PC-4, Seira, S-7, Tipper, Tremois, Triumph, Union, Volga, Wisa, Zaida.</i>	29

Table 2.  $F_{ST}$  and  $R_{ST}$  statistics for pairwise comparisons between the populations (named in Roman numerals I through IV) and subpopulations of barley genotypes defined by the analysis carried out with the STRUCTURE package for  $K=4$ .

$F_{ST}$					$R_{ST}$				
Populations	I	II	III	IV	Populations	I	II	III	IV
II	13.0	-			II	12.6	-		
III	14.0	23.3	-		III	20.9	32.0	-	
IV	11.9	23.0	11.8	-	IV	19.9	32.8	10.5	-
Subpopulations	III-1	III-2	III-3		Subpopulations	III-1	III-2	III-3	
III-2	51.3	-			III-2	70.0	-		
III-3	25.4	21.5	-		III-3	25.7	29.8	-	
Subpopulations	IV-1	IV-2			Subpopulations	IV-1	IV-2		
IV-2	8.3	-			IV-2	7.3	-		

Table 3. Means and standard deviations of geographic and climatic factors for the landrace-derived inbred lines of the SBCC of the four populations, and population IV subpopulations, deduced from the population structure analyses.

Populations (# lines)	Altitude (m)		Latitude (degrees)		Temperature (degrees)			ETP (mm)		Autumn rainfall (mm)		Winter rainfall (mm)		Spring rainfall (mm)		Turc productivity index	
	Mean*	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
I (18)	689 a	337	40.8 a	1.7	14.6 b	2.5	727 b	88	54.7 ab	18.0	44.4 a	17.4	48.9 b	13.7	14.2 ab	6.0	
II (9)	817 a	411	41.6 a	0.7	13.7 b	2.3	694 b	74	63.1 ab	35.7	48.5 a	34.5	60.5 a	14.5	16.8 a	7.8	
III (50)	757 a	219	40.6 a	1.1	14.9 b	2.0	733 b	74	49.6 b	12.3	42.0 a	12.2	45.6 b	8.8	10.9 b	4.6	
IV (82)	490 b	329	38.9 b	2.8	16.8 a	2.3	804 a	87	62.2 a	25.4	50.2 a	25.0	42.5 b	16.9	15.4 a	6.5	
Subpopulations																	
IV-1 (23)	621 a	300	40.4 a	3.0	15.3 b	1.8	744 b	64	52.4 b	25.0	41.7 a	25.6	49.3 a	18.5	13.7 a	7.2	
IV-2 (59)	439 b	328	38.3 b	2.5	17.5 a	2.2	827 a	83	66.0 a	24.7	53.5 a	24.2	39.8 b	15.6	16.0 a	6.2	

\*Means followed by the same letter within columns are not significantly different for  $P < 0.05$ .

Table 4. Loci with at least some individual allele presenting significant associations with ecogeographic and climatic factors, for a locus-wise probability  $P < 0.00018$ , equivalent to  $P < 0.05$  at the genome-wise level for 277 independent comparisons.

Chromosome	Marker	Agroecological region	Papadakis climate index	Latitude	Altitude	Ture productivity index	Mean temperature, WINTER	Mean temperature, SPRING	Precipitation, WINTER	Precipitation, SPRING	ETP, WINTER	ETP, SPRING	Ratio PRE/ETP, WINTER	Ratio PRE/ETP, SPRING	Chalmers et al (1992)	Forster et al (1997)	Turpeinen et al (2001)	Ivancic et al (2002)	Baek et al (2003)	Russell et al (2003)	Nevo et al (2005)
3H	HvM62 (3H)															R/E		A			
3H	HvLTPPB (3H)																	T			G
2H	Bmac134 (2H)																			G	
7H	Bmag369 (7H)																		A		
4H	HvBTAI3 (4H)																				
4H	HvM003 (4H)															R		R		G	G
1H	HvGLUEND (1H)																				
2H	Bmag378 (2H)																	T/R			
3H	Bmag136 (3H)																				
6H	Bmac316 (6H)																			G	G
1H	HvALAAT (1H)																				
5H	Bmac113 (5H)																			G	G
3H	Bmag006 (3H)																				
2H	Bmac132 (2H)																				
1H	Bmac399 (1H)																	G		G	
1H	HvM20 (1H)																		A	G	G
7H	Bmac156 (7H)																			G	
4H	HvM67 (4H)																R/E		A	G	
2H	Bmag125 (2H)																			G	
1H	HvHVA1 (1H)																				G
4H	HvM40 (4H)																	R			
4H	HvBAMY (4H)															T/R					
7H	HvCMA (7H)																				G
3H	Bmag225 (3H)																				
2H	HvM36 (2H)																T/L	T	R		
2H	Bmac093 (2H)																				
3H	Bmac209 (3H)																				
3H	Hv13GEIII (3H)																				
5H	Bmac337 (5H)																				
6H	EBmac806 (6H)																	T			
4H	Bmag353 (4H)																				
TOTAL		19	15	8	8	5	14	8	4	4	10	5	4	4							

A Altitude  
 E Evapotranspiration  
 G Geographical differences among landrace populations  
 L Latitude  
 R Rainfall, water availability, humidity  
 T Temperature

Loci related with similar variables are grouped as follows (from top down): association to temperature and rainfall; association to either temperature or rainfall; association to other climatic factors; association to complex geographic indexes). Also shown, associations of distribution of these loci with geographic and climatic factors described in the literature

Table 5. Distribution of STS MWG699 haplotypes across populations derived from population structure analyses.

Populations and subpopulations	N° of genotypes	Haplotype		
		A	D	K
I	55	22	26	7
II	34	4	2	28
III	50	9	40	1
III-1	8	8	0	0
III-2	12	0	12	0
III-3	30	1	28	1
IV	86	51	35	0
IV-1	24	4	20	0
IV-2	62	47	15	0
Total	225	86	103	36

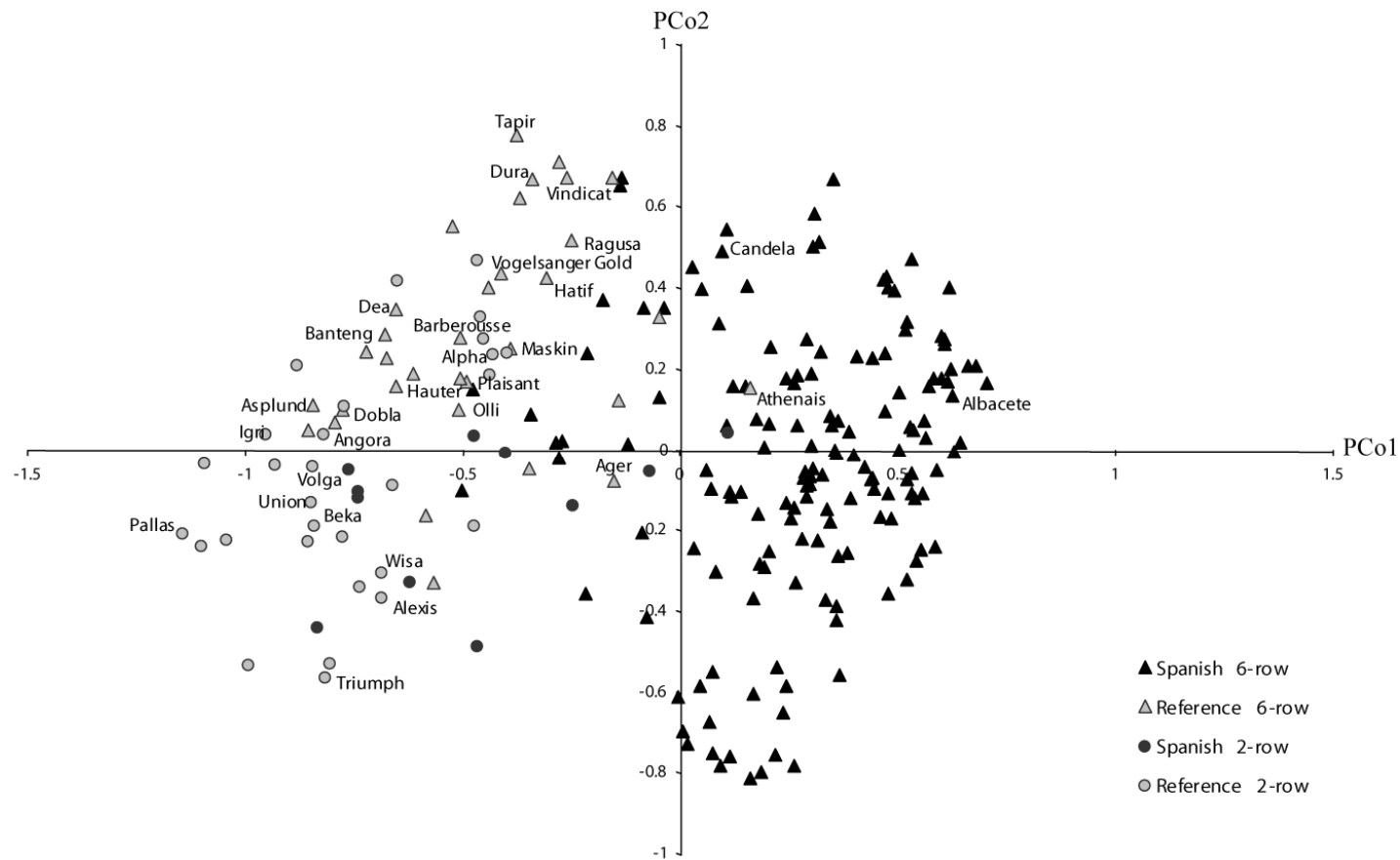


Figure 1. Associations among 225 genotypes of barley revealed by principal coordinate analysis performed on genetic distances calculated from 64 SSR data. The first principal coordinate differentiates most of the Spanish six-row landraces from other European cultivars. Genotypes included in the European Barley Core Collection are labelled.

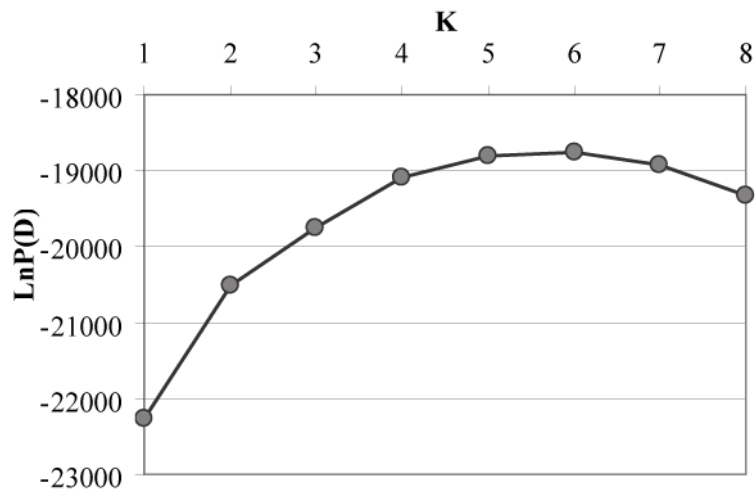


Figure 2. Evolution of the natural log probability of the data, which is proportional to the posterior probability of  $K$ , against  $K$  (number of populations). Values are the mean of 25 runs of the package STRUCTURE.

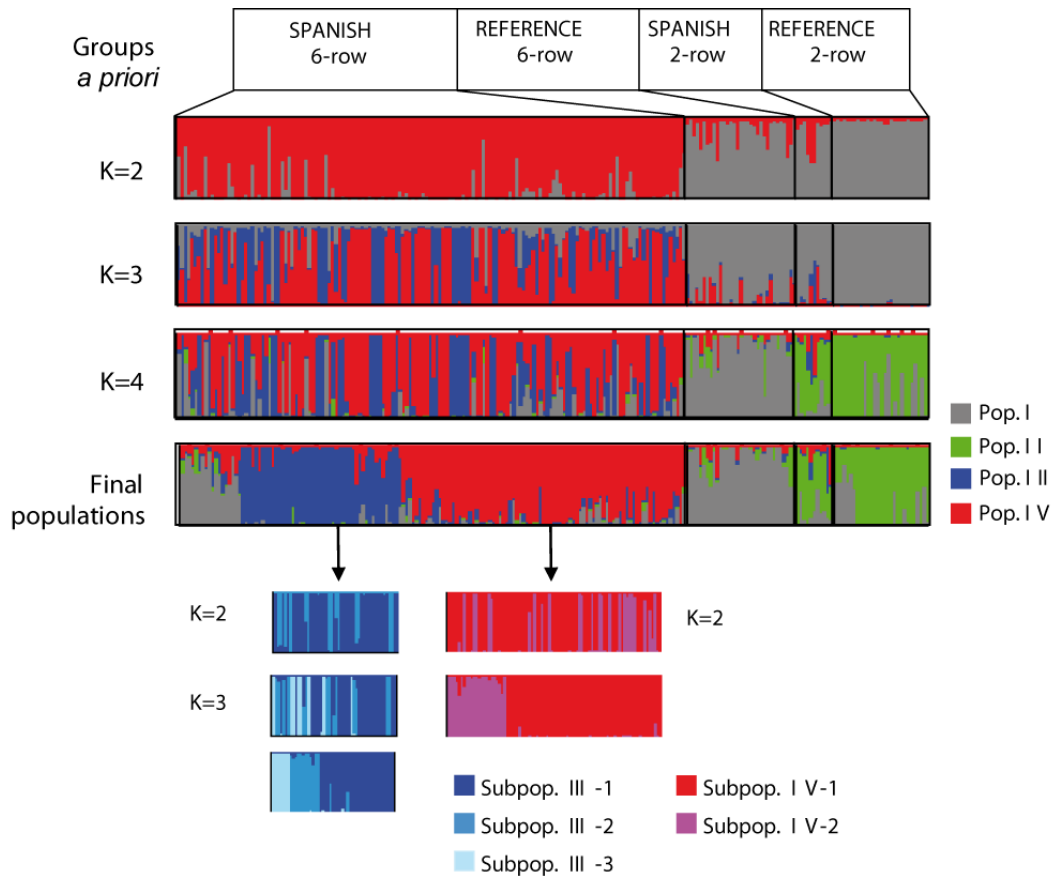


Figure 3. Clustering process using SSR information for 225 barley genotypes, and the package STRUCTURE. Genotypes are represented in columns. K is the number of populations. Genotypes are classified according to the a priori germplasm group, except in the last tier (*Final populations*), where they are classified by Q (the probability of membership of every genotype to each population, identified by colour). A second STRUCTURE analysis for populations III and IV is shown in the lower part of the Figure.



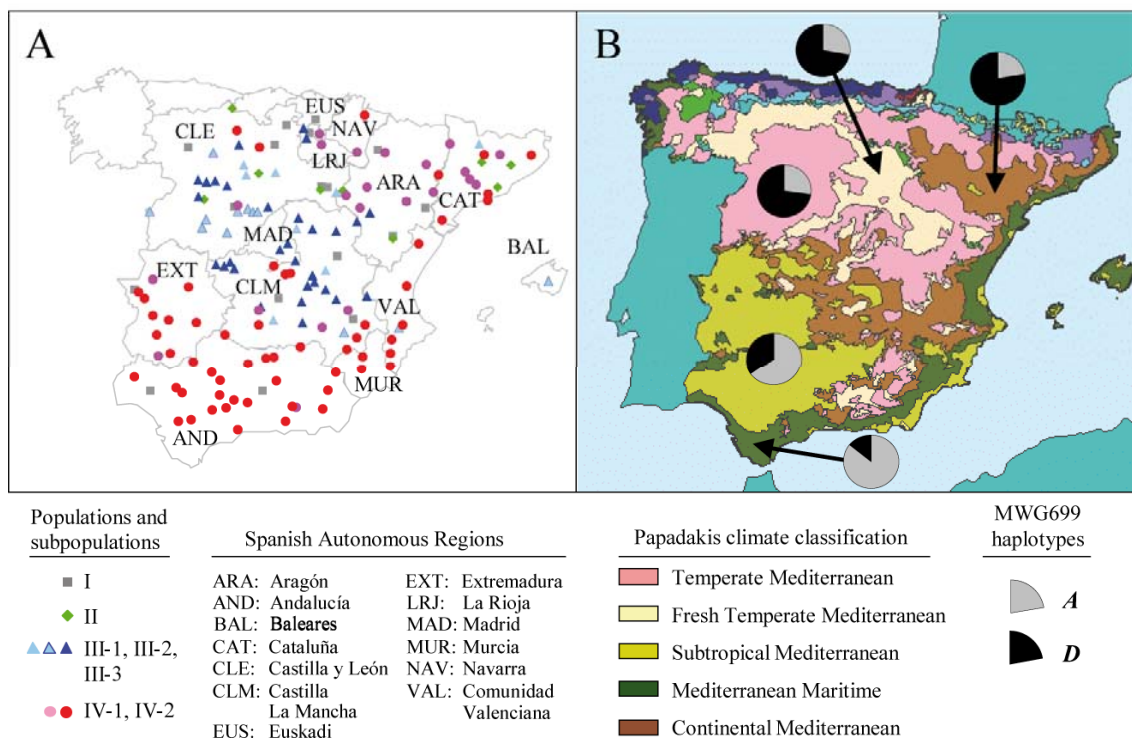


Figure 4. Distribution of 159 Spanish landraces of the Spanish Barley Core Collection, according to their classification in four populations (see text). Genotypes are placed according to latitude and longitude of collection sites. The key for the climates according to Papadakis index is given at the lower right corner (TM, Temperate Mediterranean; FTM, Fresh Temperate Mediterranean; STM, Subtropical Mediterranean; MM, Maritime Mediterranean; CM, Continental Mediterranean).