CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Digital.CSIC

Patterns of genetic and eco-geographical diversity in Spanish barleys

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

S.Yahiaoui, E. Igartua, J.M. Lasa, M.P. Gracia, A.M. Casas
Department of Genetics and Plant Production
Aula Dei Experimental Station, CSIC
Av. Montañana, 1005
50059 Zaragoza, Spain
Surface mail: PO Box 202, 50080-Zaragoza
Corresponding author, Ernesto Igartua, igartua@eead.csic.es

M. Moralejo, J.L. Molina-Cano
Centre UDL-IRTA, Av Rovira Roure 177,
25006 Lleida, Spain

L. Ramsay
Scottish Crop Research Institute, Invergowrie,
Dundee, DD2 5DA, Great Britain

F.J. Ciudad
ITA, Junta de Castilla y León
P.O. Box 172.
47071 Valladolid, Spain

Abstract

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

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

Introduction

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

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

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

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

Materials and Methods

Plant material

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

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

Molecular marker analyses

Samples of leaf tissue were taken from 8-10 plants per genotype, 14 days after sowing in paper-pots in the greenhouse. DNA extraction was carried out following a CTAB procedure, as described in Casas et al. (1998). The entire set of 225 genotypes was genotyped for 64 microsatellites, also known as simple sequence repeats (SSRs), (Supplementary Material Table S2) and for one sequence tagged site (STS) MWG699. This STS is closely linked to *vrs1,* the gene controlling ear type and that has been proposed as a diagnostic marker for barley origin (Tanno et al 1999; 2002). SSR primer pair sequences and amplification conditions were obtained from Pillen et al. (2000), Ramsay et al. (2000) and Macaulay et al. (2001). PCR amplifications were carried out in a final volume of 15 µl, containing 50 ng of genomic DNA, 1x PCR Buffer (Biotools, Madrid, Spain), 2 mM MgCl2, 15 pmol of forward and reverse primers, dNTPs at 0.2 mM each, and 0.4 U of Tth DNA Polymerase (Biotools, Madrid, Spain).

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

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

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

Genetic diversity

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

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

in which P_i is the frequency of the i^{th} SSR allele.

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

$$M_{ij} = \frac{1}{n}\sum\nolimits_k \left| x_{ki} - x_{kj} \right|$$

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, and *n* is the number of loci analysed. The distance matrix was used for a Principal Coordinate Analysis (PcoA module of NTSYS).

Analysis of molecular variance (AMOVA)

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

Analysis of genetic structure

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

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

Distribution of genetic diversity according to geographic and climatic factors

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

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

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

Results

SSR diversity

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

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

Genetic variance among the four groups of germplasm defined a priori

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

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

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

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

Genetic structure

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

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

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

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

Association of groups with geographic and climatic factors

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

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

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

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

Discussion

Genetic diversity in groups defined a priori

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

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

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

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

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

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

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

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

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

Genetic Structure

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

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

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

Association of populations with geographic and climatic factors

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

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

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

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

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

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

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

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

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

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

Acknowledgements

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

References

Baek HJ, Beharav A, Nevo E (2003) Ecological-genomic diversity of microsatellites in wild barley, *Hordeum spontaneum*, populations in Jordan, Theor Appl Genet 106:397-410

Bassam B J, Caetano-Anolles G, Gresshoff P M (1991) Fast and sensitive silver staining of DNA in polyacrilamide gels. Anal Biochem 196:80-83

Baumer M, Cais R (2000) Abstammung der Gerstensorten. Ed. Bayerische Landesanstalt für BodenKultur und Pflanzenbau. 109 p.

Briggs D E (1978) Barley. Chapman & Hall, London, pp 445-480

Casas AM, Igartua E, Vallés MP, Molina-Cano JL (1998) Genetic diversity of barley cultivars grown in Spain, estimated by RFLP, similarity and coancestry coefficients. Plant Breed 117:429-435

Casas AM, Yahiaoui S, Ciudad F, Igartua E (2005) Distribution of MWG699 polymorphism in Spanish European barleys. Genome 48:41-45

Chao S, Zhang W, Dubcovsky J, Sorrells M (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. wheat (Triticum aestivum L) germplasm representing different market classes. Crop Sci 47:1018-1030.

Ellis RP, McNicol JW, Baird E, Booth A, Lawrence P, Powell W (1997) The use of AFLPs to examine genetic relatedness in barley. Mol Breed 3:359-369

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567-1587

Fischbeck G (1992) Barley cultivar development in Europe – Success in the past and possible changes in the future. Barley Genetics VI, Vol. II (ed. L. Munck), Munksgaard International Publishers Ltd., Copenhagen, Denmark, pp. 885-901

Garris AJ, McCouch SR, Kresovish S (2003). Population stucture and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics 165:759-769

Graner A, Ludwig WF, Melchinger AE (1994) Relationships among European barley germplasm: II. Comparison of RFLP and pedigree data. Crop Sci. 34:1199-1205

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463-471

Igartua E, Gracia MP, Lasa JM, Medina B, Molina-Cano JL, Montoya JL, Romagosa I (1998) The Spanish barley core collection. Genet Resour Crop Ev 45:475-481

Ivandic V, Hackett CA, Nevo E, Keith R, Thomas WTB, Forster B (2002) Analysis of simple sequence repeats (SSRs) in wild barley from the Fertile Crescent: associations with ecology, geography and flowering time. Plant Mol Biol 48:511-527

Ivandic V, Thomas WTB, Nevo E, Zhang Z, Forster B (2003) Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in *Hordeum spontaneum*. Plant Breed 122: 300-304

Jung M, Ching A, Bhattraramakki D, Dolan M, Tingey S, Morgante M, Rafalski A (2004) Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus in elite maize germplasm. Theor Appl Genet 109:681-689

Karakousis A, Barr AR, Chalmers KJ, Ablett GA, Holton TA, Henry RJ, Lim P, Langridge P (2003) Potential of SSR markers for plant breeding and variety identification in Australian barley germplasme. Aust J Agr Res, 54:1197-1210

Koebner RMD, Donini P, Reeves JC, Cooke RJ, Law JR (2003) Temporal flux in the morphological and molecular diversity of UK barley. Theor Appl Genet 106:550-558

Komatsuda T, Maxim P, Sentil N, Mano Y (2004) High-density AFLP map of nonbrittle rachis 1 (*btr1*) and (*btr2*) genes in barley (*Hordeum vulgare* L.). Theor Appl Genet 109:986-995

Kraakman ATW, Niks RE, Van der Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-446

Lasa JM, Igartua E, Ciudad FJ, Codesal P, García EV, Gracia MP, Medina B, Romagosa I, Molina-Cano JL, Montoya JL (2001) Morphological and agronomical diversity patterns in the Spanish barley core collection. Hereditas 135:217-225

Macaulay M, Ramsay L, Powell W, Waugh R (2001) A representative, highly informative 'genotyping set' of barley SSRs. Theor Appl Genet 102:801-809

Maestri E, Malcevschi A, Massari A, Marmiroli N (2002) Genomic analysis of cultivated barley (*Hordeum vulgare*) using sequence-tagged molecular markers. Estimates of divergence based on RFLP and PCR markers derived from stress-responsive genes, and simple-sequence repeats (SSRs). Mol Genet Genomics 267:186-201

Malysheva-Otto LV, Ganal MW, Röder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in worldwide cultivated barley germplasm (Hordeum vulgare L.). BMC Genetics 7:6-19

Matus IA, Hayes PM (2002) Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. Genome 45:1095-1106

Melchinger AE, Graner A, Singh M, Messmer MM (1994) Relationships among European barley germplasm: I. Genetic diversity among winter and spring cultivars revealed by RFLPs. Crop Sci. 34:1191-1199

Molina-Cano JL, Fra-Mon P, Salcedo G, Aragoncillo C, Roca de Togores F, García-Olmedo F (1987) Morocco as a possible center for barley: biochemical and agromorphological evidence. Theor Appl Genet 98:913-918

Molina-Cano JL, Russell JR, Moralejo MA, Escacena JL, Arias G, Powell W (2005) Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. Theor Appl Genet 110:613-619

Moralejo M, Romagosa I, Salcedo G, Sànchez-Monge R, Molina-Cano JL (1994) On the origin of Spanish two-rowed barleys. Theor Appl Genet 87:829-836.

Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent. PNAS 104:3289-3294

Morris ML, Belaid A, Byerlee D (1991) Wheat and barley production in rainfed marginal environments of the developing world. Part 1 of *1990-91 CIMMYT World Wheat Facts and Trends: Wheat and Barley production in Rainfed Marginal Environments of the Developing World*. Mexico, DF: CIMMYT.

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci, USA 70:3321-3323

Nevo E, Beharav A, Meyer RC, Hackett CA, Forster BP, Russell JR, Powell W (2005) Genomic microsatellite adaptive divergence of wild barley by microclimatic stress in 'Evolution Canyon', Israel. Biol J Linn Soc 84:205–224

Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weining C, Schmitt J Purugganan MD (2004). Linkage disequilibrium mapping of *Arapidosis CRY2* flowering times alleles. Genetics 167:1361-1369

Orabi J, Backes G, Wolday A, Yahyaoui A, Jahoor A (2007) The Horn of Africa as a centre of barley diversification and a potential domestication site. Theor Appl Genet 114:1117-1127

Pandey M, Wagner C, Friedt W, Ordon F (2006) Genetic relatedness and population differentiation of Himalayan hulless barley (Hordeum vulgare L.) landraces inferred with SSRs. 167:1361-1369. Theor Appl Genet 113:715-729

Papadakis J (1975) Climates of the world and their agricultural potentialities. Edición Argentina, Buenos Aires, 200pp.

Pillen K, Binder A, Kreuzkam B, Ramsay L, Waugh R, Förster J, Léon J (2000) Mapping new EMBL-derived barley microsatellites and their use in differentiating German barley cultivars. Theor Appl Genet 101:652-660

Ramsay L, Macaulay M, degli Ivanissevich S, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997-2005

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler IV ES (2001) Structure of linkage disequilibrium and phenotypic associations in the Maize genome. Proc Natl Acad Sci, USA 98: 11479-11484.

Rohlf FJ (2000) NTSYSpc. Numerical Taxonomy and Multivariate Analysis System. Version 2.1. Exeter Software, New York

Ruimy A, Dedieu G, Saugier B (1996) TURC: A diagnostic model of continental gross primary productivity and net primary productivity. Global Biogeochemical Cycles 10 (2): 269-285.

Russell J, Fuller J, Young G, Thomas B, Taramino G, Macaulay M, Waugh R, Powell W (1997) Discriminating between barley genotypes using microsatellite markers. Genome 40:442-450

Russell JR, Ellis RP, Thomas WTB, Waugh R, Provan J, Booth A, Fuller J, Lawrence P, Young G, Powell W (2000) A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. Mol Breed 6:553-568

Russell JR, Fuller J, Baum M, Ceccarelli S, Grando S, Powell W (2003) Patterns of polymorphism detected in the chloroplast and nuclear genomes of barley landraces samples from Syria and Jordan. Theor Appl Genet 107 :413-421.

SAS Institute. 1988. SAS/ STATTM User's Guide, Release 6.03. Edition. Cary, NC: SAS 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.). Heriditas
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	Nº of accessions
Six-row	
• Spanish	
- Inbred lines derived from local landraces (SBCC)	148
- Cultivars	
Albacete, Almunia, Candela, Pané	4
• Reference set	
Ager, Asplund, Athenais, Athene, Banteng, Barberousse, Bordia, Dea, Dobla, 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	33
Two-row	
• Spanish	
- Inbred lines derived from local landraces (SBCC)	11
• Reference set	
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.	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}				
Populations	I	II	III	IV
II	13.0	-		
III	14.0	23.3	-	
IV	11.9	23.0	11.8	-
Subpopulations	III-1	III-2	III-3	
III-2	51.3	-		
III-3	25.4	21.5	-	
Subpopulations	IV-1	IV-2		
IV-2	8.3	-		

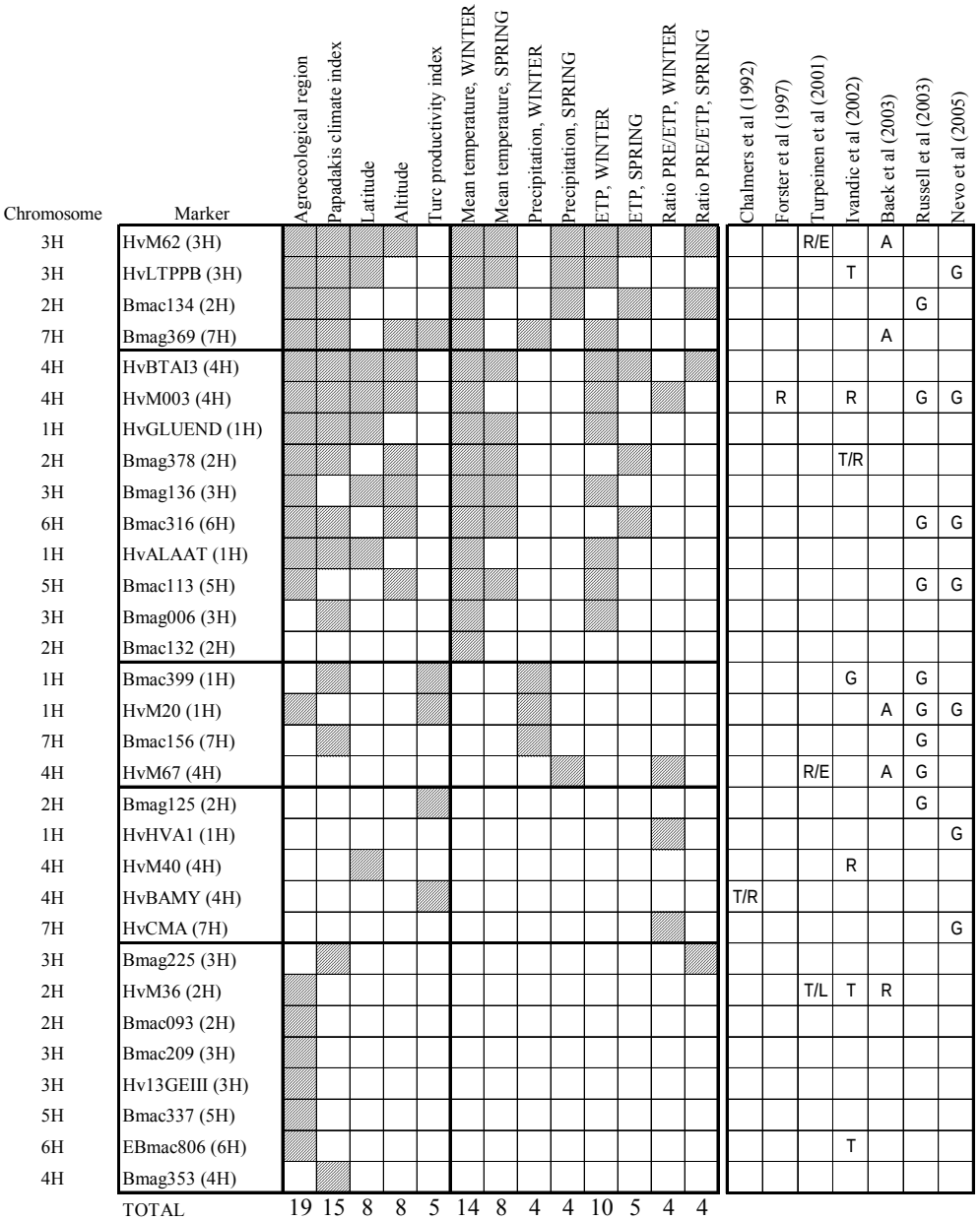
R_{ST}				
Populations	I	II	III	IV
II	12.6	-		
III	20.9	32.0	-	
IV	19.9	32.8	10.5	-
Subpopulations	III-1	III-2	III-3	
III-2	70.0	-		
III-3	25.7	29.8	-	
Subpopulations	IV-1	IV-2		
IV-2	7.3	-		

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	Altitude (m)		Latitude (degrees)		Temperature (degrees)		ETP (mm)		Autumn rainfall (mm)		Winter rainfall (mm)		Spring rainfall (mm)		Turc productivity index	
(# lines)	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.



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	Nº 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

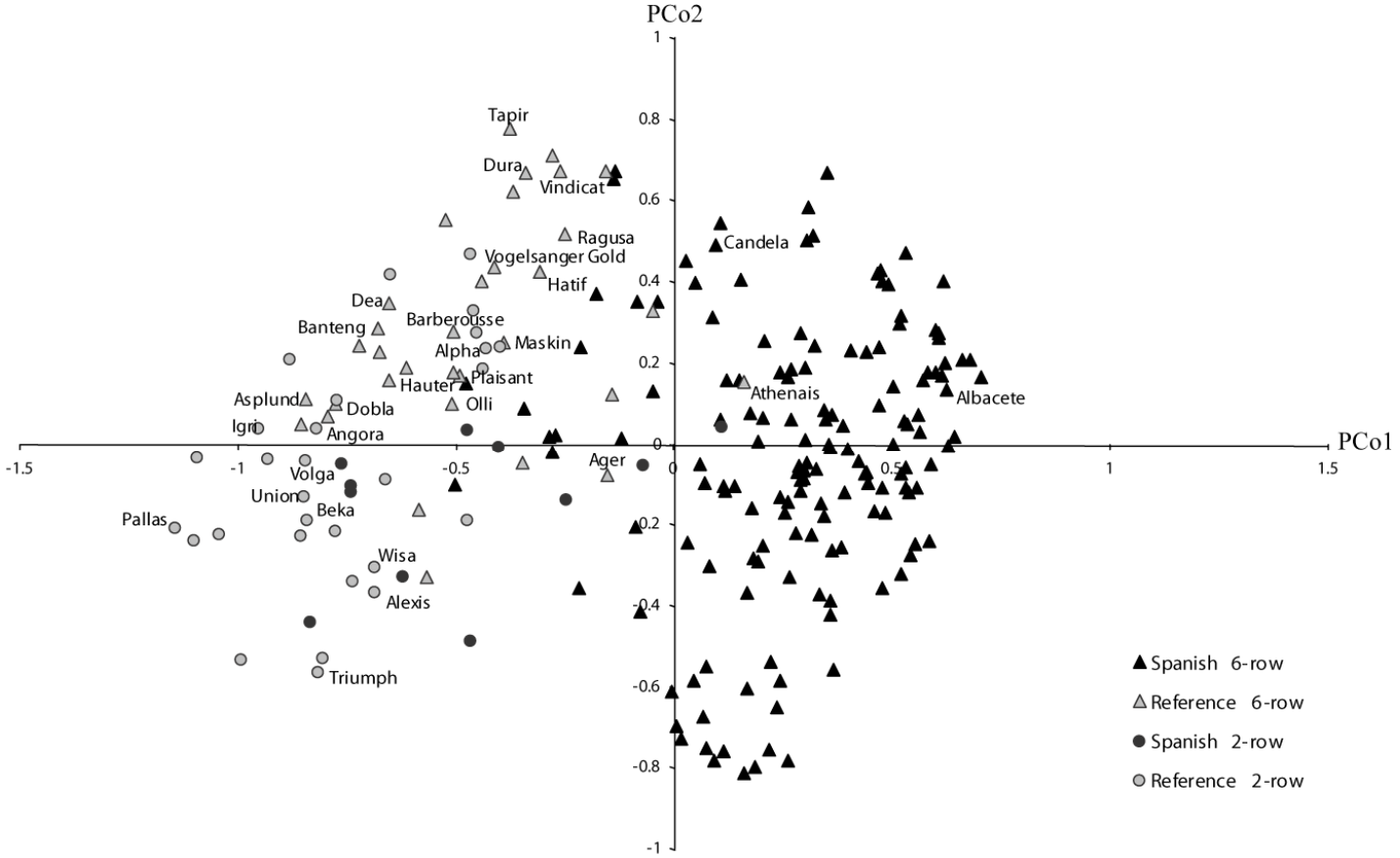


Figure 1. Associations among 225 genotypes of barley revealed by principal coordinate analysis performed on genetic distances calculated from 64 SSR data. The first principal coordinate differentiates most of the Spanish six-row landraces from other European cultivars. Genotypes included in the European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

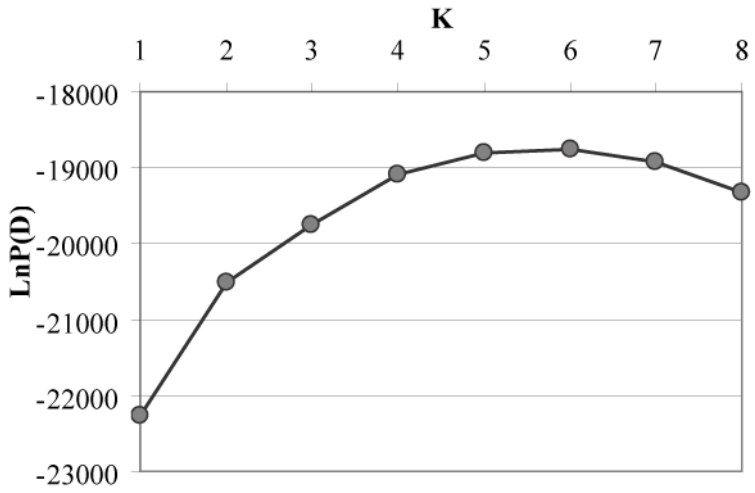


Figure 2. Evolution of the natural log probability of the data, which is proportional to the posterior probability of K, against K (number of populations). Values are the mean of 25 runs of the package STRUCTURE.

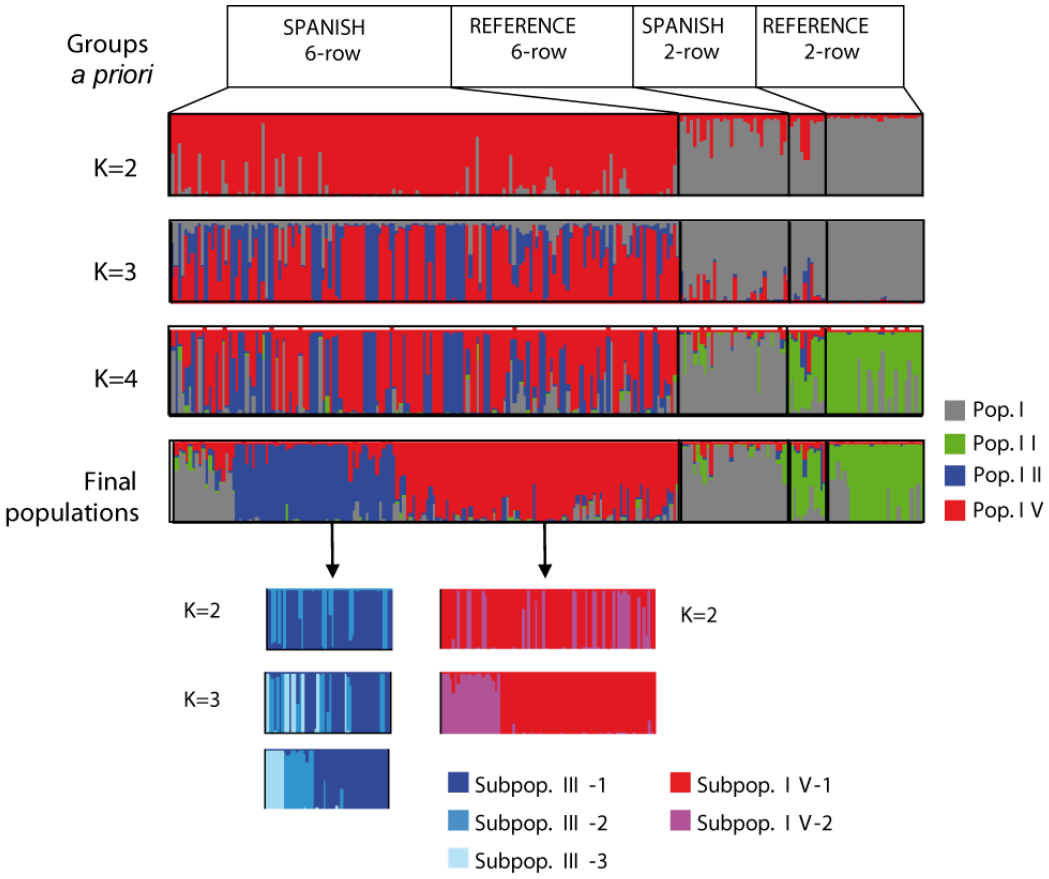


Figure 3. Clustering process using SSR information for 225 barley genotypes, and the package STRUCTURE. Genotypes are represented in columns. K is the number of populations. Genotypes are classified according to the a priori germplasm group, except in the last tier (*Final populations*), where they are 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.

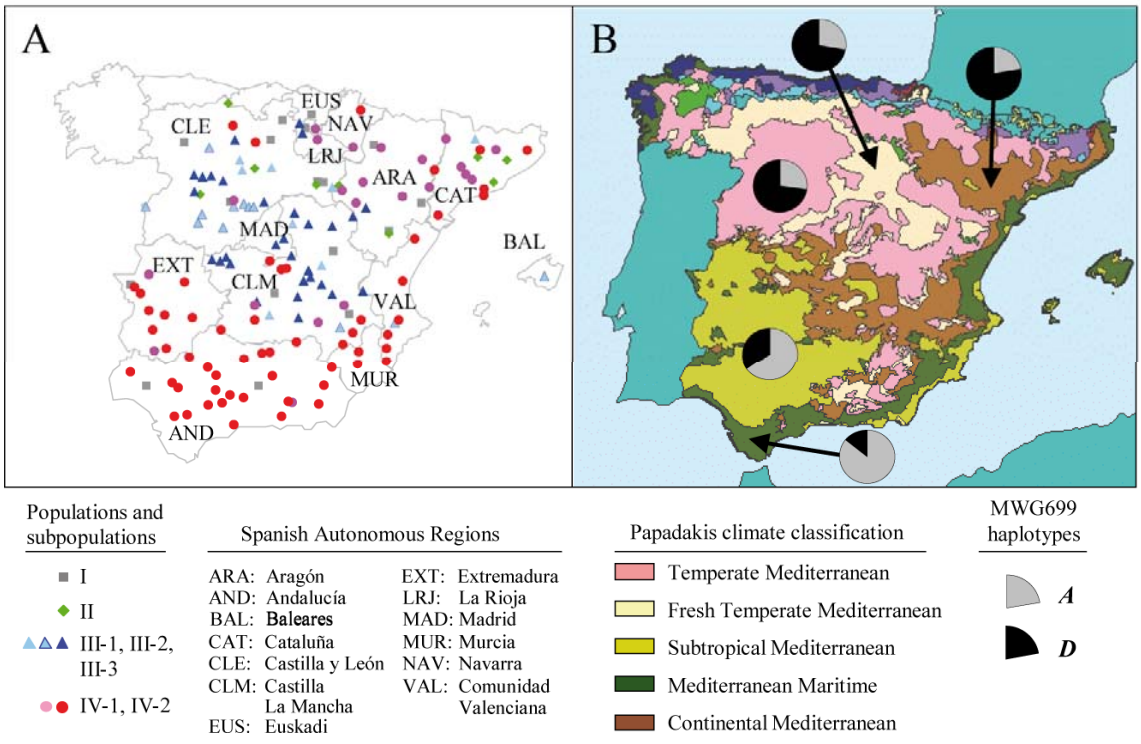


Figure 4. Distribution of 159 Spanish landraces of the Spanish Barley Core Collection, according to their classification in four populations (see text). Genotypes are placed according to latitude and longitude of collection sites. The key for the climates according to Papadakis index is given at the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