Theor Appl Genet
DOI 10.1007/s00122-012-1824-8

ORIGINAL PAPER

Genetic structure and linkage disequilibrium in landrace populations of barley in Sardinia

Monica Rodriguez · Domenico Rau · Donal O'Sullivan · Anthony H. D. Brown · Roberto Papa · Giovanna Attene

Received: 16 June 2011 / Accepted: 11 February 2012

Abstract Multilocus digenic linkage disequilibria (LD) and their population structure were investigated in eleven landrace populations of barley (*Hordeum vulgare* ssp. *vulgare* L.) in Sardinia, using 134 dominant simple-sequence amplified polymorphism markers. The analysis of molecular variance for these markers indicated that the populations were partially differentiated ($F_{ST} = 0.18$), and clustered into three geographic areas. Consistent with this population pattern, STRUCTURE analysis allocated individuals from a bulk of all populations into four genetic groups, and these groups also showed geographic patterns. In agreement with other molecular studies in barley, the general level of LD was low (13 % of locus pairs, with $P < 0.01$) in the bulk of 337 lines, and decayed steeply with map distance between markers. The partitioning of multilocus associations into various components indicated that genetic drift and founder effects played a major role in determining the overall genetic makeup of the diversity in these landrace populations, but that epistatic homogenising or diversifying selection was also present. Notably, the variance of the disequilibrium component was relatively high, which implies caution in the pooling of barley lines for association studies. Finally, we compared the analyses of multilocus structure in barley landrace populations with parallel analyses in both composite crosses of barley on the one hand and in natural populations of wild barley on the other. Neither of these serves as suitable mimics of landraces in barley, which require their own study. Overall, the results suggest that these populations can be exploited for LD mapping if population structure is controlled.

Communicated by A. Graner.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1824-8) contains supplementary material, which is available to authorized users.

M. Rodriguez (✉) · G. Attene
Centro per la Conservazione e Valorizzazione della Biodiversità Vegetale, Università degli Studi di Sassari, Via E. de Nicola, 07100 Sassari, Italy
e-mail: mrodrig@uniss.it

D. Rau · G. Attene
Sezione di Agronomia, Coltivazioni Erbacee e Genetica (SACEG), Dipartimento di Agraria, Università degli Studi di Sassari, Via E. de Nicola, 07100 Sassari, Italy

D. O'Sullivan
John Bingham Laboratory, National Institute of Agricultural Botany, Huntington Road, Cambridge CB3 0LE, UK

A. H. D. Brown
Centre for Australian National Biodiversity Research CSIRO Plant Industry, GPO Box 1600, Canberra ACT 260, Australia

R. Papa
Dipartimento di Scienze Agrarie, Alimentari ed Ambientali, Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy

R. Papa
Cereal Research Centre, CRA-CER Agricultural Research Council (CRA), S.S. 16, km 675, 71122 Foggia, Italy

Introduction

Landraces are populations that have evolved in response to human selection in subsistence agriculture. In such systems, traditional farmers have maintained a wealth of genetic variation mediated through human migration, seed exchange, and natural selection (Harlan 1975; Brush 2000). Since landraces pre-date modern plant breeding and its intense purifying selection that produces modern varieties,

Published online: 13 March 2012

they constitute an important source of genetic diversity for plant breeding (Frankel et al. 1995). Landraces are therefore the result of evolutionary processes that lead to heterogeneous populations rather than to few superior genotypes (Brown 2000; Ceccarelli and Grando 2000). In selfing species such as barley, this diversity, held together as blocks of genes in chromosomal regions with low frequency of recombination, may confer a specific adaptation to stress environments (Ceccarelli and Grando 2000). Landraces thus provide an interesting model for association mapping to identify genes that control adaptive variation in crop species (Vigouroux et al. 2002; Mazzucato et al. 2008; Bitocchi et al. 2009; Comadran et al. 2009).

The non-random association of alleles at different loci (linkage disequilibrium, LD), namely between a marker locus and a phenotypic trait locus, is the starting point for association mapping studies. Levels of LD depend on the amount and distribution of the genetic diversity, the mating system, selection regimes and recombination events in the ancestry of the genotypes. High LD levels arise from inbreeding, small population size, population structure, admixture, low recombination rate, or natural and artificial selection. The decay or decrease of LD with increasing map distance between markers is usually faster in allogamous than in autogamous species, and in wild relatives than in modern varieties (Flint-Garcia et al. 2003; Gaut and Long 2003; Morrell et al. 2005; Caldwell et al. 2006; Rostoks et al. 2006; Mather et al. 2007; Song et al. 2009).

The level and distribution of LD play a central role in association studies because they determine the number and the density of markers needed for the analysis (Flint-Garcia et al. 2003; Slatkin 2008). In particular, when using populations where LD is low and decays within few thousands of base pairs, a candidate gene approach is usually preferred; whereas for populations with moderate or high LD, a whole genome scan can be more appropriate. However, population structure may lead to confounding effects in LD mapping studies contributing to spurious statistical correlations between unlinked markers and phenotypic variation (Pritchard et al. 2000; Mackay and Powell 2007). Nonetheless, association mapping is achievable when account is taken of population structure (Cockram et al. 2008; Comadran et al. 2011).

Landraces are often included in LD studies mainly as isolated genotypes from different, and often unrelated, populations in the attempt to capture high levels of genetic variation (Maccaferri et al. 2005; Caldwell et al. 2006; Comadran et al. 2009). However, very little is known of the levels and structure of LD within and among populations of the same landrace, growing in a range of agro-ecological conditions and in a context where complex seed exchange networks could exist.

In the present study, we investigate eleven populations of a Sardinian Barley Landrace (SBL), all of which local farmers call "S'orgiu sardu" (i.e. Sardinian barley). Farmers' management practices together with natural selection have resulted in populations with high levels of genetic diversity and adaptation to the local environment (Attene et al. 1996; Gorham et al. 1994). Papa et al. (1998) documented the value of the Sardinian barley germplasm as a source of genetic variation using morphological and molecular markers (random amplified polymorphic DNA markers and isozymes). Subsequent genotype by environment studies conducted in six different environments with SBL lines, their recombinant derivatives and commercial varieties point to their potential utility in breeding barley adapted to Mediterranean environments (Rodriguez et al. 2008).

The material studied comprised five populations collected in Sardinia in 1990, some of which were also analysed in Papa et al. (1998), and a further six populations collected in 1999. We performed molecular analyses using the retrotransposon-based markers named simple-sequence amplified polymorphism (S-SAP). These markers are reliable and appropriate for a range of genetic analyses (Leigh et al. 2003; Tam et al. 2005; Rodriguez et al. 2006) including the study of LD in eukaryotic genomes (Charlesworth et al. 1994; Fu et al. 2002). However, compared to other kinds of markers relatively few population genetic studies have used the S-SAP technique (e.g. Queen et al. 2004; Soleimani et al. 2007; Tam et al. 2007).

Several studies have investigated the level of linkage disequilibrium among genetic markers in various barley populations (such as different accessions extracted from cultivars, breeding lines, landraces collected from different regions or countries) (Caldwell et al. 2006; Malysheva-Otto et al. 2006; Rostoks et al. 2006; Comadran et al. 2009; Zhang et al. 2009). Such sets often comprise lines with contrasting diversity (such as winter vs. spring barley; malting vs. feed barley). The level of disequilibrium so constructed is of direct interest to mapping and breeding, but may not reflect the genetic structure of the source populations. Our samples are random plants from populations in situ chosen to examine the standing disequilibrium at local and regional levels using S-SAP markers.

To achieve this objective, we estimated: (a) the levels and the structure of the S-SAP genetic diversity present in our collection of populations; (b) the levels and patterns of LD between pairs of S-SAP loci; (c) the population structure of multilocus LD in multiple SBL populations. Specifically, our main objective was testing if such populations could be suitable for future LD mapping studies. Moreover, we aimed to determine whether landrace populations depart in any key aspect from composite cross populations of cultivated barley, and from populations of wild barley (*Hordeum vulgare* ssp. *spontaneum*). This would be useful for a deeper understanding and better management of these different kinds of genetic stocks.

Materials and methods

Plant materials

Eleven populations of a barley (*Hordeum vulgare* ssp. *vulgare* L.) landrace were collected in the island of Sardinia (Italy) (see Table S1 of supplementary materials for more details). For each population, we analysed approximately 30 individuals, giving a total of 337 individuals (Table 1). Lines were randomly sampled from each field (one spike per plant) and each field was assumed to be a single population. Five of the eleven populations were collected in 1990 from different agro-ecological areas of Sardinia, and the other six were collected in 1999 from the same geographic areas sampled in 1990.

Molecular data

DNA was extracted individually from each of the 337 Sardinian lines, starting from fresh leaf tissues and using the CTAB method (Doyle and Doyle 1987). Each line was analysed using 6 S-SAP primer combinations (Rodriguez et al. 2006). The S-SAP method (Waugh et al. 1997b; Leigh et al. 2003) exploits the combination of a primer designed on the long terminal repeat (LTR) sequence of a barley retrotransposon (e.g. Sukkula, Nikita, BAGY-2, BARE-1) and an Mse primer, which usually generates high levels of polymorphism.

Each primer combination was previously tested in a set of three barley varieties (Leigh et al. 2003) and subsequently used to enrich a Steptoe × Morex (S × M) genetic map (Rodriguez et al. 2006). To infer a map position for the S-SAP markers which were polymorphic in SBL, the 337 individuals were amplified alongside Steptoe, Morex and two of the DH lines obtained from the S × M cross. The S-SAP markers showing the same molecular weight both in the S × M DH lines and in the SBL individuals were presumed to be the products of the same locus (Waugh et al. 1997a; Kraakman et al. 2004). In this way, map positions were deduced for 53 of the 134 polymorphic markers scored in the SBL populations.

Genetic diversity

Barley is a strictly autogamous species with an outcrossing rate of <1 % (Briggs 1978; Abdel-Ghani et al. 2004). Preliminary analysis conducted on SBL using 11 simple sequence repeats (SSR) markers (not shown) showed only a couple of heterozygous loci in a few individuals consistent with autogamy. Therefore, all of the individuals in the present study were assumed to be homozygous. Descriptive statistics such as Nei's gene diversity (H_E, Nei 1978) and the number of polymorphic markers were computed using PopGene 1.32 software (Yeh et al. 1999). The number of haplotypes and alleles was computed by Arlequin 3.5.1.2 (Excoffier and Lischer 2010).

Table 1 Diversity statistics within populations of SBL and overall

Population	Altitude (m.a.s.l)	No. individuals	Haplotypes	Polymorphic loci (no.)	Average gene diversity		Rare alleles[a]	
					H_E^b	SD	No.	%
CUM	50	33	33	106	0.230	0.181	41	39
N2	85	30	30	104	0.228	0.178	31	30
ORO	240	30	30	108	0.241	0.181	32	30
SIS3	20	28	28	114	0.280	0.177	23	20
PIR	35	33	33	110	0.237	0.174	27	24
SOR	110	32	32	118	0.270	0.170	39	33
STU	374	30	30	92	0.183	0.174	32	35
NXM	190	31	31	101	0.228	0.174	25	25
SEN	200	30	30	112	0.268	0.180	25	22
VI	105	30	29	115	0.260	0.181	35	30
COR	95	30	30	111	0.269	0.180	23	21
Mean		30.6	30.5	108.3	0.245	0.177	30.2	28
ALL		337	336	134	0.298	0.156	33	25

The populations are ordered according latitude (from the north to the south of Sardinia)

SD standard deviation, *m.a.s.l* meters above sea level

[a] "Rare alleles" are defined here as the number of loci at which the frequency of the dominant or the recessive allele <0.1 and reported as the number or the percentage of all loci polymorphic within each of the 11 populations

[b] Unbiased genetic diversity (Nei 1978)

Population structure

The hierarchical analysis of molecular variance (AMOVA) was used to test the significance of the partitioning of genetic variance into three levels: individuals, populations and groups of populations (grouped according the year of collection, i.e. 1990 and 1999) using Arlequin 3.5.1.2 (Excoffier and Lischer 2010). Genetic distances between populations were calculated using Nei's unbiased genetic distance (Nei 1978). A dendrogram was drawn using the unweighted pair group method with arithmetic mean (UPGMA) clustering method implemented in TFPGA (Miller 1997). The relative strength of the nodes produced by UPGMA analysis was inferred by bootstrapping over loci (1,000 permutations). To have a better insight into the genetic structure of the SBL populations, different methods were exploited. First, we applied the Bayesian model-based clustering algorithm implemented in STRUCTURE 2.3.1 (Pritchard et al. 2000; Falush et al. 2003, 2007; Hubisz et al. 2009). This method assigns each individual to several different groups, according to a membership coefficient (q_i; $\Sigma q_i = 1.0$). The options set for the admixture model were: 'correlated allele frequencies among populations' and 'infer the degree of admixture (α) by the data'. The number of hypothetical populations (K) tested was from 1 to 14, which is three more than the original number of populations. For each K, 20 runs (burn-in length l00,000 and 200,000 iterations) were carried out and the logarithm probability of data, lnP(D), was collected for each run. The most likely number of K was determined using the ΔK ad hoc statistic (Evanno et al. 2005). Because the estimation of the most likely number of clusters, K, is a notoriously difficult problem (Guillot et al. 2009), we also used the Instruct software (Gao et al. 2007) to infer the optimal number of clusters via the deviance information criterion (DIC), which is widely used as a statistic for comparing models in a Bayesian framework (Spiegelhalter et al. 2002). Genetic relationships among populations were investigated by principal coordinate analysis (PCoA) based on the Jaccard's double-centred similarity matrix. Jaccard's coefficient is appropriate for binary multilocus data (Landry and LaPointe 1996). For this analysis NTSYSpc 2.02i (Rohlf 2000) was used. The correlations between genetic divergence (F_{ST}) and geographic distance between pairs of populations (km) were tested by the non-parametric Mantel test using GenAlEx 6.3 (Peakall and Smouse 2006).

Evaluation of linkage disequilibrium

The estimates of the LD were determined using two different indices:

(a) the squared allele-frequency correlations (r^2) (Hill and Robertson 1968) for pairs of loci, calculated using the software package Tassel 2.1 (Bradbury et al. 2007; http://www.maizegenetics.net/bioinformatics). For the purpose of examining the potential of SBL for association mapping, we preferred to calculate the r^2 statistic, as it is indicative of how markers might correlate with the quantitative trait loci (QTL) of interest (Flint-Garcia et al. 2003; Mueller 2004). The significance (P values) of LD for S-SAPs was determined by permutation (100,000 reps).

(b) A summary measure of multilocus LD (r_d), calculated using the software MultiLocus 1.3 (Agapow and Burt 2001). The index r_d derives from the multilocus association index (I_A) (Brown et al. 1980) which was modified to avoid dependence on number of loci (Agapow and Burt 2001). The significance of the r_d estimates was tested by shuffling alleles across individuals (1,000 reps).

To display the change in LD as a function of genetic distance, LD for pairs of markers was measured for five classes of inter-marker genetic distance (≤3, 4–10, 11–30, 31–50, >50 cM) (e.g. Maccaferri et al. 2005). LD among markers located on different chromosomes was also calculated. The level of long range LD was calculated overall, within each of the eleven SBL populations, and within the genetic groups obtained by STRUCTURE analysis. An overall correlation between the genetic distance between markers on the same chromosome and LD was estimated. To derive a critical value of r^2 indicative of LD above sampling error, the r^2 estimates for pairs of unlinked markers were square root transformed to approximate a normally distributed random variable and the parametric 95th percentile of that distribution determined (Breseghello and Sorrells 2006).

Structure of multilocus associations

The structure of multilocus LD in multiple populations was analysed using the framework of Brown and Feldman (1981). This method summarises the structure of the multilocus associations in a series of subpopulations as a set of components. Specifically, "Each source is measured in terms of its contribution to the variance in heterozygous loci in two gametes randomly chosen from within the same population or from a pool of all populations." (Brown and Feldman 1981). They define three single-locus and five two-locus components. Namely, the single-locus components are the average diversity (MH), the variance among populations in gene diversity (VH) and the variance among populations in allele frequency, namely the Wahlund effect (WH). A high percentage of variance accounted for by the single-locus components indicates that alleles are randomly associated in an unstructured total population. The two-locus components include the mean disequilibria (MD), the variance of disequilibria (VD), the covariance of allele frequencies over populations, namely the Wahlund effect at two loci (WC), the interaction of disequilibria (AI) between

Table 2 Percentage (%) of significant pairwise associations, average squared correlation coefficient (r^2) and multilocus index of association (r_d) calculated within each geographical population, the four groups identified by STRUCTURE, and the overall bulk

Population	No. of individuals	Pairwise LD (%)[a]	Average r^2	r_d and significance
CUM	33	0.32	0.036	0.003**
N2	30	1.12	0.057	0.027***
ORO	30	0.74	0.051	0.019***
PIR	33	2.12	0.049	0.014***
SIS3	28	1.32	0.055	0.017***
SOR	32	0.81	0.043	0.011***
STU	30	0.41	0.052	0.022***
NXM	31	0.71	0.041	0.009***
SEN	30	0.76	0.045	0.014***
VI	30	1.48	0.056	0.019***
COR	30	1.00	0.044	0.020***
Mean	30.68	0.98	0.048	0.016
Grp1	65	1.68	0.024	0.008***
Grp2	53	3.06	0.035	0.013***
Grp3	52	2.50	0.033	0.021***
Grp4	47	1.26	0.030	0.010***
Mean	54.25	2.13	0.0305	0.013
ALL	337	13.34	0.012	0.011***

Grp1 is constituted by individuals from the North (22 from N2, 20 from ORO, 23 from CUM); Grp2 is mainly constituted by individuals from the centre (14 SIS3, 19 PIR, 19 SOR, 1 SEN); Grp3 is mainly constituted by individuals from the south (27 VI, 6 SEN, 11 COR, 2 PIR, 1 ORO, 2 SOR); Grp4 is mainly constituted by individuals from the south (27 STU, 14 NXM, 5 SEN, 1 COR)

[a] Significant at $P < 0.01$

, * Significant levels at $P < 0.01$ and $P < 0.001$, respectively

MD and WC, and CI which arises from covariation in the interaction of disequilibria and the Wahlund covariance among populations (Brown and Feldman 1981). The partitioning assists in hypothesising which of the several evolutionary forces (such as epistatic natural selection, genetic drift, population subdivision, founder effect, genetic hitchhiking) are relatively important when comparing patterns of disequilibria in different sets of populations. For example, when epistatic selection is consistent among populations MD is high, VD low and AI positive. High values of CI are associated with founder effects. High WC and low AI indicate diversifying selection and population subdivision. The analysis was performed with POPGENE 1.32 (Yeh et al. 1999).

Finally, results of multilocus analysis on SBL were compared with those obtained by Brown and Feldman (1981) on barley composite cross populations and wild barley populations (*Hordeum vulgare* ssp. *spontaneum*) (see for details Brown and Feldman 1981, last column of Table 1 and third column of Table 2, respectively).

Results

Genetic diversity and population structure

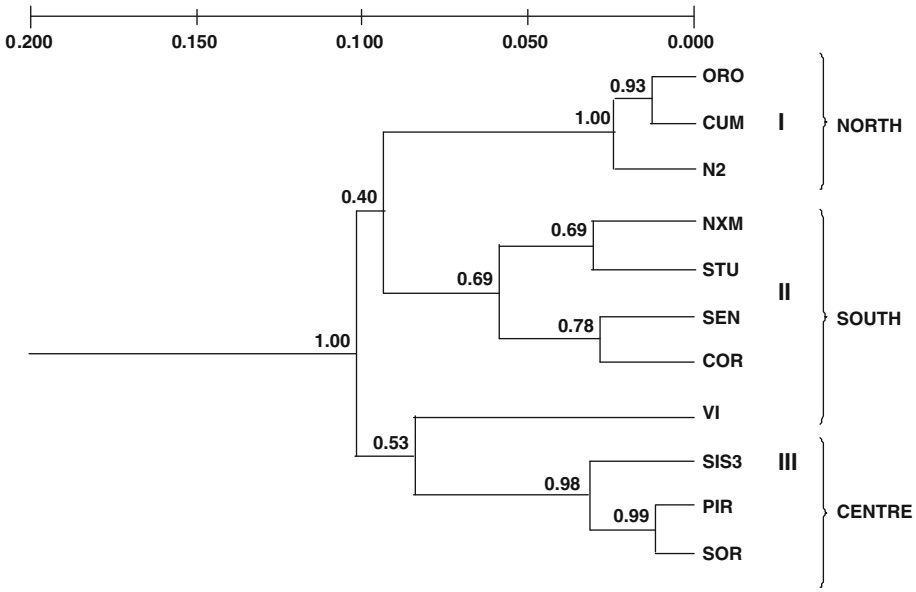
Genetic diversity

Estimates of genetic diversity, based on the 134 S-SAP markers confirm that these populations are highly polymorphic (Table 1). Indeed, the number of haplotypes nearly matches the number of individuals assayed; each individual was genetically unique (except for one pair of individuals from VI). Within the populations, the number of polymorphic markers varied from 92 (STU) to 118 (SOR). The overall gene diversity (H_E) was 0.30 and the mean within populations was 0.25. The difference in H_E between STU (0.18) and COR, SIS3, SEN, SOR and VI (from 0.26 to 0.28) was statistically significant (Tukey-Kramer HSD test, $P < 0.05$). STU, the most elevated population of the set, was the least genetically diverse. This is consistent with previous results for allozymes in wild barley from Israel (Nevo et al. 1979), and encourages further investigations to compare this result to previous studies on barley (e.g. Tanto Hadado et al. 2010). The percentage of "rare" markers (the number of loci at which the frequency of the minor allele is <0.1) varied among populations from 20 % (SIS3 and COR) to 39 % (CUM) (Table 1).

Population structure

The AMOVA partitioning of S-SAP diversity indicated that the within-population component (82 %) significantly dominated ($P < 0.001$). No differentiation occurred between the two groups of populations collected in 1990 and in 1999 ($F_{CT} = 0.00$). The UPGMA dendrogram (Fig. 1) clustered ten landrace populations into three main groups. Group I comprised three populations from the north, group II consisted of four populations from the south and group III included three populations from the centre of Sardinia. The southern population VI weakly clustered with group III. This analysis suggested a relationship between genetic divergence and geographic distance between pairs of populations. However, the correlation was weak and barely significant based on a Mantel non-parametric test ($r = 0.27$, $P = 0.05$).

Four main genetic groups (hereafter named GRP1-4) were inferred from the STRUCTURE analysis of the pooled data and the calculation of ΔK (STRUCTURE) and DIC (InStruct) (Fig. S1 and S2 of Supplementary materials). The majority of individuals (64 %) were assigned to one of the four genetic groups (membership coefficient, $q_i, \geq 0.8$) while 36 % were classed as admixed ($q_i < 0.8$). As illustrated in Fig. 2, group 1 mainly consisted of individuals from populations of the north of the Island (CUM,



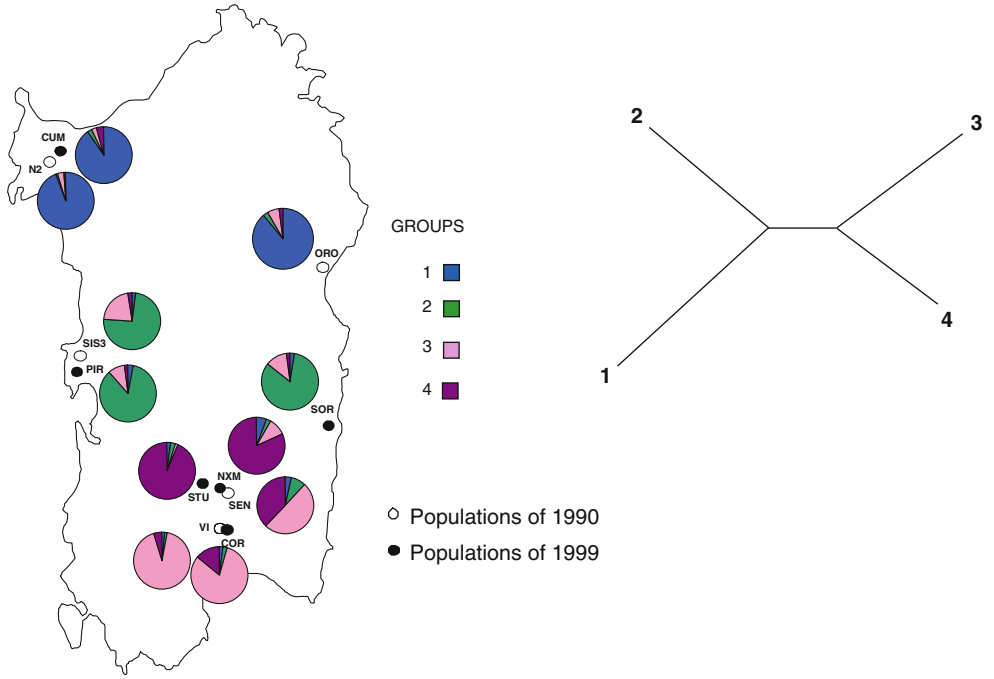


Fig. 1 UPGMA tree describing genetic relationships among the eleven geographical landrace populations. Genetic distances were calculated based on Nei 1978. Consistency of the tree topology was evaluated by bootstrapping over loci (10,000 reps)

Fig. 2 Map of the sampled sites in Sardinia illustrating the distribution of the four genetic groups. Pie charts were obtained by using the mean value of q_i (membership coefficient) per genetic group within populations including only individuals with $q_i > 0.8$. The neighbour-joining tree on the right side was obtained by the net nucleotide distances (allele-frequency divergence) among Structure groups

N2 and ORO). Individuals from the centre (SIS3, PIR, and SOR) merged into group 2. Individuals from VI, COR and SEN belonged mainly to group 3. Group 4 included individuals from the south (STU, NXM and partly from SEN). The third genetic group appeared more widespread than the other three, although only a few group 4 individuals were present in the north. The PCoA showed a pattern similar to STRUCTURE when the three principal coordinates were used, which together explained 15.7 % of the total S-SAP allelic variance (Fig. 3a, b). For all assigned individuals, the first axis separated groups 1 and 4 from 2 and 3, the second axis mainly separated group 1 from group 4, and the third axis separated group 2 from group 3. Because the results for population structure obtained from the different analyses are consistent, only the genetic groups obtained from STRUCTURE (Fig. 2) will be used further.

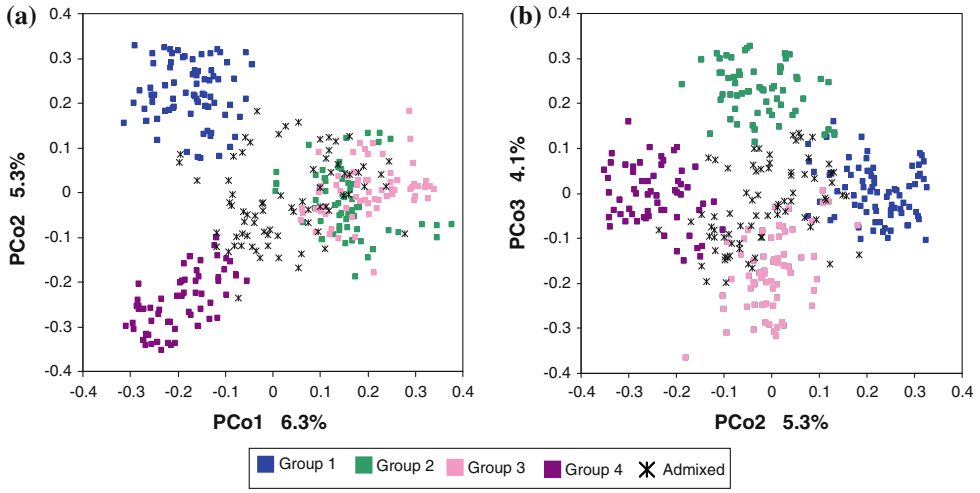


Fig. 3 Principal coordinate analysis of 337 SBL individuals based on Jaccard's distance matrix between all pair of individuals. *Scatter plots* represent the grouping of the individuals according to the first coordinate versus the second one (**a**) and the second coordinate versus the third one (**b**). The four represents the different genetic groups identified by STRUCTURE analysis ($q_i > 0.8$). Admixed individuals are indicated by "x" ($q_i < 0.8$). The three principal coordinates cumulatively explain the 15.7 % of the S-SAP genetic variance

Levels of linkage disequilibrium

As evident in Table 2, the levels of LD within these SBL populations and the overall bulk were low and statistically significant. In the bulk of 337 individuals, some 13.3 % of the locus pairs were in significant LD and the average r^2 was 0.012. Moreover, populations differed in their levels, with population CUM being the lowest based on all of the three measures of LD. Population N2 had the highest levels for the r^2 and r_d statistics and was above average for its percentage of significant pairwise associations. Both r^2 and r_d measures were highly significantly correlated over populations (Spearman $\rho = 0.79$, $P < 0.004$), whereas the correlations between r^2 and percentage of locus pairs in LD, or r_d and percentage of locus pairs in LD, were not significant ($\rho = 0.56$, $P < 0.08$ and $\rho = 0.28$, $P < 0.40$, respectively). Within the four genetic groups, the percentage of locus pairs in LD was higher, but the average level of LD was less intense (r^2). To check the consistency of these estimates of LD, we repeated the analyses modifying the original dataset by (a) omitting the rare alleles (frequency <0.10) (e.g. Caldwell et al. 2006; Rossi et al. 2009), and (b) taking out very similar individuals (we set ≤0.15 % of identical alleles) to obtain a "normalised" sample as suggested in Breseghello and Sorrells (2006). The analyses confirmed the low levels of LD within populations and overall. Moreover, significant correlations were observed among the same parameters of LD estimated from different subsets of the data (Table S2).

Decay of linkage disequilibrium with increasing linkage distance

The relationship between LD and linkage distance between markers was analysed in the bulk sample of 337 individuals and in the five classes of inter-marker genetic distance (Fig. 4). Despite the limited number of mapped markers, LD tended to decrease with distance. The correlation coefficient between the inverse of intra-chromosomal linkage distance and LD r^2 was 0.243 ($P < 0.01$). The level of LD decayed within 3 cM distance from an average $r^2 = 0.10$ to below the critical background value (Breseghello and Sorrells 2006). This tendency was replicated within three of the four genetic groups identified by STRUCTURE (Fig. 5), and confirmed within the populations. Among the eleven populations, all seven instances in which the class means exceeded the sampling threshold, were for the first distance class (see Fig. S3 of supplementary materials).

The distribution of pairwise r^2 estimates among unlinked loci across SBL populations varied from 0.00 to 0.30, with a median of 0.01 (results not shown). Breseghello and Sorrells (2006) used the 95th percentile of this distribution (0.05 in our case) as an empirical estimate of the mean background LD. This represents a threshold beyond which the r^2 values between unmapped markers are likely to indicate genetic linkage in mapping populations such as segregating F_2s. Across the SBL populations, LD scores above this value were observed only for the 3 cM distance class (Fig. 4). Overall, for the 337 individuals, intra-chromo-

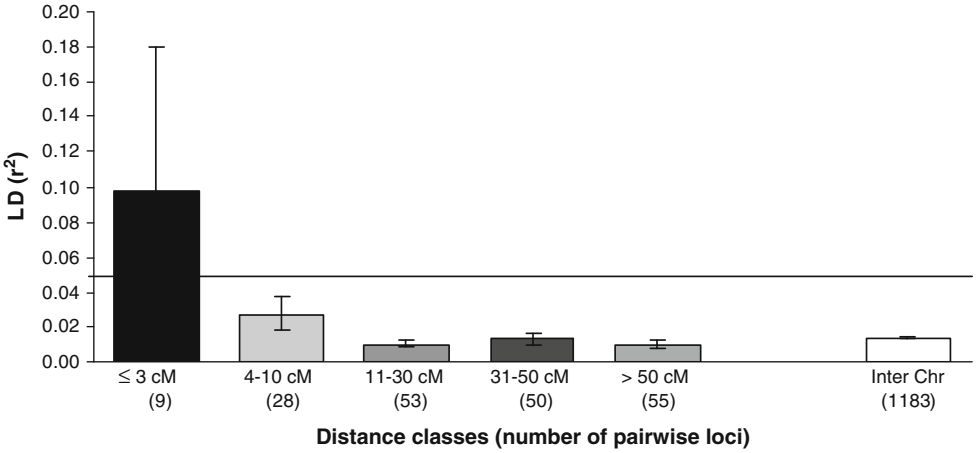


Fig. 4 LD decay as calculated across all of the 337 individuals. The *horizontal straight line* indicates the threshold above which r^2 values are likely due to genetic linkage. *Inter Chr* inter-chromosomal pairs. All *bars* are means ± SE (standard error). The number of pairs of loci is shown for each distance class

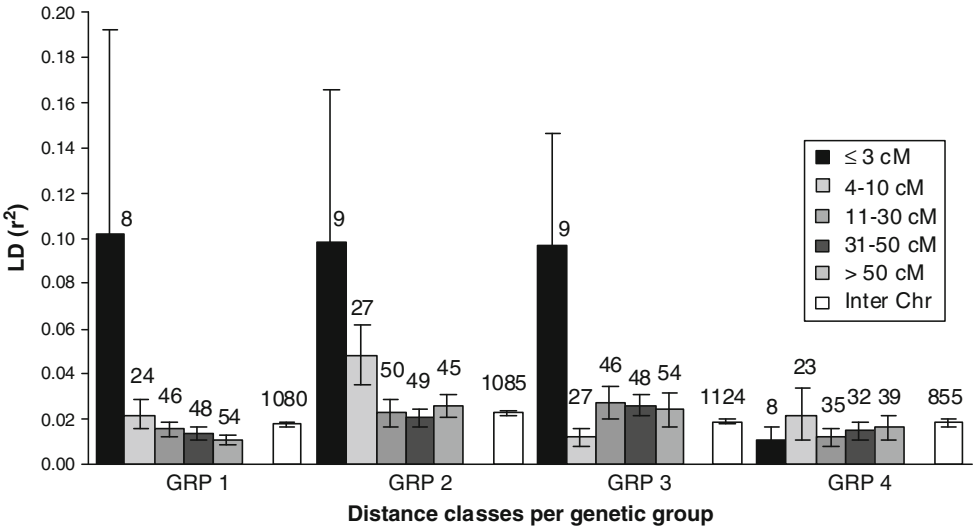


Fig. 5 LD decay within the four genetic groups identified by STRUCTURE analysis. All *bars* are means ± SE (standard error). The number of pairs of loci is shown above each bar

somal LD was 1.14 times higher than interchromosomal LD both as percentage of pairs and as r^2 (Table S3).

Structure of multilocus LD

Table 3 presents the results of the analysis of multilocus LD in structured populations following the method of Brown and Feldman (1981). The multilocus allozyme data on barley composites crosses and wild barley populations (*Hordeum vulgare* ssp. *spontaneum*) are included from that study for comparison. Overall, the SBL resembled the composite crosses more than the wild populations. The single-locus effects accounted for only about half the total variance in heterozygosity in both sets of cultivated barley populations. In sharp contrast, the single-locus proportion was much higher for the wild populations (93 %). In SBL and composite crosses, the two-locus effects show an appreciable mean disequilibrium and an average positive AI that suggest repetitive patterns of linkage disequilibria in different populations. The SBL rather differed from COM for their WC proportions (30 vs. 9 %).

As well as for single-locus effects, the SBL differed from WILD also for the two-locus effects, in having an appreciable MD and a higher WC (Table 3).

The fraction [(AV-MH)/MH] is a standardised measure of multilocus structure within populations, and its value for the SBL is more than threefold higher than composite cross and wild populations (Table 3). The variance of disequilibrium is the most important two-locus source in both SBL and wild populations (50 and 32 % of the average variance, respectively). This indicates that non-systematic disequilibria across populations (i.e. the frequency of multilocus

Table 3 Components of variance in the number of heterozygous comparisons between haploid genotypes for the eleven populations of a Sardinian landrace, composite crosses (COM) and wild populations (WILD)

Code	Component description	% of the total variance			% of the average variance		
		SBL	COM	WILD	SBL	COM	WILD
	Single-locus effects	43.4	54.7	93.0			
MH	Mean gene diversity	35.3	47.9	57.7	23.7	58.0	51.8
VH	Variance of diversity	3.9	3.0	18.1			
WH	Wahlund's allele-frequency variance	4.2	3.8	17.2			
	Two-locus effects	56.6	45.3	7.0			
MD	Mean disequilibrium	20.4	22.9	1.3	13.7	27.7	1.2
WC	Covariance of allele frequencies (Wahlund)	30.1	8.9	4.8			
AI	MD × WC interaction	6.1	13.5	0.4	4.1	16.3	0.4
VD	Variance of disequilibrium				50.2	1.8	31.6
CI	Covariance of interaction				8.3	−3.8	15.8

Components are expressed as percentages of the total variance (TV = MH + VH + WH + MD + WC + AI) in an equiproportional mixture of all populations sampled, or as percentages of the average within-population variances (AV = MH + MD + AI + VD + CI). Original values obtained for SBLs were: MH = 20.00; VH = 2.19; WH = 2.40; MD = 11.58; WC = 17.08; AI = 3.43; VD = 42.32; CI = 56.69; TV = 56.69; AV = 84.37

haploid genotypes varies substantially among different populations) were a major contributor to multilocus structure of SBL and wild populations.

Discussion

The present study investigated the genetic diversity, the population structure and the LD levels of 11 populations of a barley landrace from the island of Sardinia (Italy) as evident from polymorphism for 134 S-SAP fragments.

Genetic diversity and population structure

The high number of unique haplotypes detected in the populations studied indicates substantial genetic variation within the SBL, comparable to a previous study of Sardinian barley (Papa et al. 1998). Overall, the level of the S-SAP diversity in these populations is appreciable ($H_S = 0.24$), and similar to that detected by isozymes ($H_S = 0.35$, Papa et al. 1998). The extent of genetic divergence among populations was of a similar order for S-SAP ($F_{ST} = 0.18$) as for allozymes ($G_{ST} = 0.16$) (Papa et al. 1998). Considerable genetic marker diversity is present within traditional varieties of barley (Jaradat and Shahid 2006) and in other selfing cereals such as rice (Thomson et al. 2007; Pusadee et al. 2009), as might be expected from the many factors affecting diversity in landraces (Teshome et al. 2001). No relevant differences were observed between the two groups of genotypes collected in 1990 and 1999, suggesting that one decade was insufficient for any significant genetic temporal differentiation among these populations to arise.

A genetic distance-based analysis clustered the populations into groups according to their geographic origin and a weak correlation was found between genetic distance and geographic distance ($r = 0.27$, $P < 0.05$). A STRUCTURE analysis of the bulk of 337 individuals allocated a majority of them to four genetic groups. The hierarchical island structure of genetic variation had four main genetic groups which tended to have distinct geographic occurrence, in agreement with the distance clusters. In our study the structural tendencies were not absolute, as different genetic groups co-occur within each population and more than 30 % of the individuals were apparently admixed i.e. derived from a hybrid between different genetic groups. Thus, this suite of barley landrace populations exhibited a complex population structure that will influence LD patterns and their exploitation in barley breeding.

Linkage disequilibrium extent and decay with linkage distance between markers

In the bulk sample of 337 individuals the level of LD was relatively low, with some 13 % of locus pairs showing statistical correlations ($P < 0.01$), or 22 % at $P < 0.05$ significance level. This value slightly exceeded that between loci in a sample of 25 accessions of the wild subspecies (*Hordeum vulgare* ssp. *spontaneum*) from across its range (c. 15 %, $P < 0.05$) (Morrell et al. 2005). The similar proportion of significant r^2 values probably reflects the net outcome of two opposing factors. On the one hand, the widespread geographic origin of the samples, autogamy and the isolation of wild populations might predict considerable LD within the species. On the other hand, our geographically restricted sample of individuals was 13-fold

larger than that for the wild species and therefore likely to detect a greater fraction of low LD values. Yet the "relatively" low overall levels of disequilibrium in both of these studies were unexpected, given the autogamous mating system of barley.

Table 4 summarises six studies of interlocus LD in barley. In a study with a comparable number of mapped markers, Malysheva-Otto et al. (2006) found an average r^2 of 0.10 (range 0.062–0.191) at intra-chromosomal level and of 0.064 (range 0.050–0.136) at inter-chromosomal level (see Table 4 of Malysheva-Otto et al. 2006). Our r^2 values tended to be less than theirs, but the proportions of locus pairs with $r^2 > 0.05$ were similar. Our data also parallel those from other studies employing a higher density of markers (e.g. Zhang et al. 2009; Comadran et al. 2009). Moreover, in the present study r^2 and P values were similar for the 53 mapped S-SAP pairs of loci and the total 134 markers, which suggests that increased genome coverage would not significantly alter the overall conclusion of limited disequilibrium (data not shown).

In Table 4, different criteria were used in these studies to gauge the extent of LD and its rate of decay with increasing map distance between markers. Despite the low number of mapped and linked markers in our study, LD decreased below the estimate background threshold (0.05) for marker pairs separated by >3 cM. This outcome is similar to those obtained by Zhang et al. (2009) and Comadran et al. (2009). The reason why the decay is more rapid in barley than might be expected in a selfing species (Zhang et al. 2009; Comadran et al. 2009; Morrell et al. 2005; Caldwell et al. 2006) is unclear. Rostoks et al. (2006) propose this is a consequence of "unique human-induced pseudo-outbreeding" coupled with "strong selection for advantageous alleles" in agriculture. They consider that a collection of lines from diverse breeding programs might approach the recombinational dynamism of an outbreeding species like maize. However, wild and landrace populations which presumably have not been subject to such intense crossing and selection also show low LD and rapid decay.

Within the groups inferred by STRUCTURE, the number of locus pairs in significant LD was only approximately 2 % (versus 13 % in the total sample). Fewer individuals than the total (~60 vs. 337) constitute each group, so the observed drop in significant LD is likely due in part to a reduced power of the LD test (Remington et al. 2001; Liu et al. 2003). Moreover, the reduction in the number of marker pairs in significant LD was more evident for the inter-chromosomal comparison, than for the intra-chromosomal comparison (see Table S3 of supplementary materials). Thus, high LD appears to be more associated with close physical linkage within STRUCTURE groups than in the overall sample. Fewer spurious associations are expected from the analysis of the four genetic groups, as was also the case in durum wheat (Maccaferri et al. 2005).

Our results agree with others on the importance of inferring genetic structure and detect genetic differences among populations for intra-specific biodiversity assessment, evolutionary studies and association mapping (Caldwell et al. 2006; Rostoks et al. 2006; Mazzucato et al. 2008; Rossi et al. 2009; Comadran et al. 2011).

Multilocus LD dissection

Morrell et al.'s (2005) study of disequilibrium in wild barley accessions particularly stressed that geographic divergence was a major source of interlocus disequilibrium. The method of Brown and Feldman (1981) is one approach to analysing the relative importance of various sources of disequilibrium and the biology of multilocus association. Furthermore, the partitioning lends itself to comparisons among different kinds of populations, for example the pattern observed in SBL populations compared with that in composite crosses of cultivated barley and in wild populations of *H. vulgare* ssp. *spontaneum* (Brown and Feldman 1981).

In our analysis, the landraces tended to resemble the composite cross populations more than the wild ones. This was mainly due to the more prominent role of two-locus effects. At the two-locus level, each set of populations displayed a distinct LD pattern, with landraces tending to a "hybrid" pattern between wild and composite cross populations. Bulked hybrid populations such as composite crosses have been proposed as a model system for studying the evolution of genetic diversity and co-evolution with biotic factors, for the conservation in situ of agro-biodiversity (Brown 2000). However, our results suggest that neither wild populations nor composite crosses are complete mimics of landraces. Specifically, the multilocus structure of composite crosses seemed to be partly associated with systematic, repeatable selection (high MD and significant positive AI) (Brown and Feldman 1981). In contrast, wild populations showed a pattern consistent with the hypothesis that founder effects (or epistatic localised diversifying selection) might dominate (high VD and positive CI). In the SBL populations, on one hand we observed a pattern of LD that indicates systematic repeatable selection, similar to composite crosses. On the other hand, similar to wild populations a prominent role of genetic drift or founder effect, and diversifying selection was found.

When allele frequencies differ among populations, population subdivision might generate a reduction of heterozygosity (WH) and/or non-random associations between alleles at multiple loci (WC). Whereas both for SBL and composite crosses, the WC was greater than WH (seven- and twofold, respectively), the contrary was true for wild

Table 4 Estimates of interlocus LD using different criteria in various kinds of samples, of domesticated (*Hordeum vulgare* ssp. *vulgare*) and its wild progenitor (*Hordeum vulgare* ssp. *spontaneum*), compared to the results for the same criteria in eleven SBL populations using 134 S-SAP markers

Criterion	Authors	Barley sample (mapped markers)	LD estimate	SBL populations		
				Whole sample	Within populations	Within genetic groups
% of locus pairs with $r^2 > 0.20$ over 10 cM[a]	Kraakman et al. (2004)	146 European two-row spring cultivars (123 AFLPs)	0.11 %	0.1 %	0.3 %	0.2 %
% of locus pairs with $P < 0.05$	Morrell et al. (2005)	c. 25 Wild barley accessions (418 mutations at 18 loci)	14.7 %	22.1 %	3.4 %	5.9 %
% of locus pairs with $r^2 > 0.05$	Malysheva-Otto et al. (2006)	Bulk of 953 worldwide cultivars (48 SSRs)	3.3 %[b]	4.2 %[c]	3.4 %[c]	6.0 %[c]
		Subset of 207 two-row	4.7 %[b]			
Mean (median) interlocus distance for which $r^2 > 0.5$	Rostoks et al. (2006)	91 elite European cultivars (612 SNPs)	3.9 (1.2) cM[d]	1.5[e]	7.5 (1.5)	1.5 (1.5)[f]
		Subset of 53 two-row	1.5 (0.8) cM[g]			
Distance at which $r^2 < 0.2$ ($P < 0.001$)	Zhang et al. (2009)	170 Canadian cultivars (846 DArTs)	2.6 cM	1.5 cM	1.5–19.2 cM	2.7–6.7 cM
		Subset of 84 two-row	3.4 cM			
		Subset of 86 six-row	3.5 cM			
Distance at which r^2 is less than the critical value for a given sample size (Breseghello and Sorrells 2006)	Comadran et al. (2009)	Bulk of 192 landrace and cultivar accessions (916 DArTs)	3.2 cM	3 cM	3 cM	3 cM

[a] Derived from visual inspection of Fig. 2 in the source paper

[b] In the case of $P < 0.001$

[c] In the case of $P < 0.05$

[d] In the case of $r^2 < 0.5$ and $P < 0.001$ up to 60 cM

[e] Only 1 pairwise association was found with $r^2 > 0.5$

[f] Only 2 pairwise associations were found with $r^2 > 0.5$

[g] In the case of $r^2 < 0.5$ and $P < 0.001$ up to 15 cM

populations. This suggests that, wild populations may have had more time to "accumulate" recombination events that led to a reduction of the association between alleles at two or more loci. Under this scenario, an appreciable amount of LD persists in SBL and composite crosses even without selection. Similar results are expected for predominantly inbreeding species, such as barley, because associations among haplotypes that are generated by mutation and random genetic drift can persist in inbred and partially isolated subpopulations (Allard 1999; Morrell et al. 2005).

Conclusions

The relatively low level of LD found in the SBL populations is a desirable property for association mapping studies. However, the high variance of disequilibrium among populations means that barley lines should not be pooled indiscriminately and indicates the need to control for the presence of population structure when conducting association studies. The multilocus analyses revealed features of SBL midway between wild and composite cross populations. This result begs the question of whether the observed pattern of LD is the signature of "evolutionarily sustainable production" that has led to the formation of landraces. Further comparative studies are needed to test this hypothesis.

Acknowledgments MR performed the experiments under the supervision of DOS and analysed the data; MR and DR interpreted the data and wrote the manuscript; AHDB contributed ideas and co-wrote the manuscript; DOS, RP and GA contributed ideas and commented on the manuscript. RP and GA conceived and designed the study. Molecular analyses were carried out by the first author during the 2003 at NIAB (Cambridge, UK) while supported from the EC by a Marie-Curie Training Site Fellowship. This work was also supported by the Sardinian Region (Master and Back Program). The reviewers of the manuscript are thanked for helpful comments.

References

Abdel-Ghani AH, Parzies HK, Omary A, Geiger HH (2004) Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. Theor Appl Genet 109:588–595

Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. Mol Ecol Notes 1:101–102

Allard RW (1999) History of plant population genetics. Annu Rev Genet 33:1–27

Attene G, Ceccarelli S, Papa R (1996) The barley (*Hordeum vulgare* L.) of Sardinia, Italy. Genet Resour Crop Evol 43:385–393

Bitocchi E, Nanni L, Rossi M, Bellucci E, Giardini A, Buonamici A, Vendramin GG et al (2009) Introgression from modern hybrid varieties into landrace populations of maize (*Zea mays* ssp. *mays* L.) in central Italy. Mol Ecol 18:603–621

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Briggs DE (1978) Barley. Chapman and Hall, London

Brown AHD (2000) The genetic structure of crop landraces and the challenge to conserve them in situ on farms. In: Brush SB (ed) GENES in the FIELD. On-farm conservation of crop diversity. IPGRI/IDRC/Lewis Publishers, Boca Raton, pp 19–48

Brown AHD, Feldman MW (1981) Population structure of multilocus associations. Proc Natl Acad Sci USA 78:5913–5916

Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure in natural populations of *Hordeum spontaneum*. Genetics 96:523–536

Brush SB (2000) The issues of in situ conservation of crop genetic resources. In: Brush SB (ed) GENES in the FIELD. On-farm conservation of crop diversity. IPGRI/IDRC/Lewis Publishers, Boca Raton, pp 3–26

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172:557–567

Ceccarelli S, Grando S (2000) Barley landraces from the Fertile Crescent: a lesson for plant breeders. In: Brush SB (ed) GENES in the FIELD. On-farm conservation of crop diversity. IPGRI/IDRC/Lewis Publishers, Boca Raton, pp 3–26

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

Cockram J, White J, Leigh FJ, Lea VJ, Chiapparino E, Laurie DA, Meckay IJ, Powell W, O'Sullivan DM (2008) Association mapping of partitioning loci in barley. BMC Genet 9:16. doi:10.1186/1471-2156-9-16

Comadran J, Thomas WTB, van Eeuwijk FA, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, Romagosa I, Hackett CA, Russell JR (2009) Patterns of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. Theor Appl Genet 119:175–187

Comadran J, Ramsay L, MacKenzie K, Hayes P, Close TJ, Muehlbauer G, Stein N, Waugh R (2011) Patterns of polymorphism and linkage disequilibrium in cultivated barley. Theor Appl Genet 122:523–531. doi:10.1007/s00122-010-1466-7

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

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

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes. doi:10.1111/j.1471-8286.2007.01758.x

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Frankel OH, Brown AHD, Burdon JJ (1995) The conservation of plant biodiversity. Cambridge University Press, Cambridge

Fu HH, Zheng ZW, Dooner HK (2002) Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. Proc Natl Acad Sci USA 99:1082–1087

Gao H, Williamson S, Bustamante CD (2007) An MCMC approach for joint inference of population structure and inbreeding rates from multi-locus genotype data. Genetics 176:1635–1651

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1502–1506

Gorham J, Papa R, Aloy-Lleonart M (1994) Varietal differences in sodium uptake in barley cultivars exposed to soil salinity or salt spray. J Exp Bot 45:895–901

Guillot G, Leblois R, Coulon AL, Frantz AC (2009) Statistical methods in spatial genetics. Mol Ecol 18:4734–4756

Harlan JR (1975) Our vanishing genetic resources. Science 188:618–621

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332

Jaradat AA, Shahid M (2006) Population and multilocus isozyme structures in a barley landrace. Plant Genet Resour Charact Util 4:108–116

Kraakman ATW, Niks RE, Van den 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

Landry PA, LaPointe FJ (1996) RAPD problems in phylogenetics. Zool Scr 25:283–290

Leigh F, Kalendar R, Lea V, Lee D, Donini P, Schulman AH (2003) Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. Mol Genet Genomics 269:464–474

Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–289

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12:57–63

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

Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). Genetics 177:2223–2232

Mazzucato A, Papa R, Bitocchi E, Mosconi P, Nanni L, Negri V, Picarella ME et al (2008) Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.). Theor Appl Genet 116:657–669

Miller MP (1997) Tools for population genetic analysis (TFPGA) 1.3: a Windows program for the analysis of allozyme and molecular population genetic data (distributed by the author)

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. Proc Natl Acad Sci USA 102:2442–2447

Mueller JC (2004) Linkage disequilibrium for different scales and applications. Brief Bioinforma 5:355–364

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Nevo E, Zohary D, Brown AHD, Haber M (1979) Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum*, in Israel. Evolution 33:815–833

Papa R, Attene G, Barcaccia G, Ohgata A, Konishi T (1998) Genetic diversity in landrace populations of *Hordeum vulgare* L. from Sardinia, Italy, as revealed by RAPDs, isozymes and morphophenological traits. Plant Breed 117:523–530

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Pusadee T, Jamjod S, Chiang Y, Rerkasem B, Schaal BA (2009) Genetic structure and isolation by distance in a landrace of Thai rice. Proc Natl Acad Sci USA 106:13880–13885

Queen RA, Gribbon BM, James C, Jack P, Flavell AJ (2004) Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. Mol Genet Genomics 271(1):91–97. doi:10.1007/s00438-003-0960-x

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Rodriguez M, O'Sullivan D, Donini P, Papa R, Chiapparino E, Leigh F, Attene G (2006) Integration of retrotransposon-based markers in a linkage map of barley. Mol Breed 17:173–184

Rodriguez M, Rau D, Papa R, Attene G (2008) Genotype by environment interactions in barley (*Hordeum vulgare* L.): different responses of landraces, recombinant inbred lines and varieties to Mediterranean environment. Euphytica 163:231–247

Rohlf FJ (2000) NTSYS-pc. Numerical taxonomy and multivariate analysis system, version 2.1. Exeter Software, Setauket

Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, Papa R (2009) Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. Evol Appl 2:504–522

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR et al (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 103:18656–18661

Slatkin M (2008) Linkage disequilibrium: understanding the evolutionary past and mapping the medical future. Nat Rev Genet 9:477–485

Soleimani VD, Baum BR, Johnson DA (2007) Analysis of genetic diversity in barley cultivars reveals incongruence between S-SAP, SNP and pedigree data. Genet Resour Crop Evol 54:83–97

Song B-H, Windsor AJ, Schmid KJ, Ramos-Onsins S, Schranz ME et al (2009) Multilocus patterns of nucleotide diversity, population structure and linkage disequilibrium in *Boechera stricta*, a wild relative of Arabidopsis. Genetics 181:1021–1033

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. J R Stat Soc B 64:583–640

Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, Grandbastien MA (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based S-SAP, AFLP and SSR. Theor Appl Genet 110:819–831

Tam SM, Causse M, Garcher C, Burck H, Mhiri C, Grandbastien M-A (2007) The distribution of copia-type retrotransposons and the evolutionary history of tomato and related wild species. J Evol Biol 20:1056–1072

Tanto Hadado T, Rau D, Bitocchi E, Papa R (2010) Adaptation and diversity along an altitudinal gradient in Ethiopian barley (*Hordeum vulgare* L.) landraces revealed by molecular analysis. BMC Plant Biol 10:121

Teshome A, Brown AHD, Hodgkin T (2001) Diversity in landraces of cereal and legume crops. Plant Breed Rev 21:221–261

Thomson MJ, Septiningsih EM, Suwardjo F, Santoso TJ, Silitonga TS, McCouch SR (2007) Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. Theor Appl Genet 114(3):559–568

Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y et al (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc Natl Acad Sci USA 99:9650–9655

Waugh R, Bonar N, Baird E, Thomas B, Graner A et al (1997a) Homology of AFLP products in three mapping populations of barley. Mol Gen Genet 255:311–321

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W (1997b) Genetic distribution of BARE-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253:687–694

Yeh FC, Yang R, Boyle T (1999) Popgene, version 1.32. Microsoft window-based freeware for population genetic analysis. University of Alberta, Edmonton. http://www.ualberta.ca/~fyeh/index.htm

Zhang LY, Marchand S, Tinker NA, Belzile F (2009) Population structure and linkage disequilibrium in barley assessed by DArT markers. Theor Appl Genet 119:43–52