Theor Appl Genet (2009) 119:43–52
DOI 10.1007/s00122-009-1015-4

ORIGINAL PAPER

Population structure and linkage disequilibrium in barley assessed by DArT markers

Li Yi Zhang · Suzanne Marchand · Nicholas A. Tinker · François Belzile

Received: 19 November 2008 / Accepted: 11 March 2009 / Published online: 3 April 2009

Abstract Diversity Array Technology (DArT) markers were used to investigate the genetic diversity, population structure, and extent of linkage disequilibrium (LD) on a genome-wide level in Canadian barley (*Hordeum vulgare* L.). Approximately 1,000 DArT markers were polymorphic and scored with high confidence among a collection of 170 barley lines composed mostly of Canadian cultivars and breeding lines. The reproducibility of DArT markers proved very high, as 99.9% of allele calls were identical among seven replicated samples. The polymorphism information content (PIC) of DArT markers ranged between 0.04 and 0.50 with an average of 0.38. Using principal coordinate analysis (PCoA), most lines fell into one of two major groups reflecting inflorescence type (two-row versus six-row). Within these two large groups, evidence of geographic clustering of genotypes was also observed. A cluster analysis Unweighted Pair Group Method with Algorithmic Mean suggested the existence of three subgroups within the two-row group and four subgroups within the six-row group. An analysis of molecular variance (AMOVA) revealed highly significant ($P < 0.001$) genetic variance within subgroups, among subgroups, and among groups. Values of LD, expressed as r^2, declined with increasing genetic distance, and mean values of r^2 fell below 0.2 for markers located 2.6 cM apart. Approximately 8% of marker pairs located on the same chromosome and 3.4% of pairs located on different chromosomes were in LD ($r^2 > 0.2$). Within both the subsets of two-row and six-row lines, LD extended slightly further (3.5 cM) than for the entire set, while 7.5% of intra-chromosomal locus pairs and <2% of inter-chromosomal pairs were in LD. We discuss the implications of these findings with regard to the prospects of association mapping of complex traits in barley.

Communicated by J.-L. Jannink.

L. Y. Zhang · S. Marchand · F. Belzile (✉)
Département de Phytologie, Université Laval,
1243 Marchand Bldg, Québec, QC G1K 7P4, Canada
e-mail: francois.belzile@fsaa.ulaval.ca

N. A. Tinker
Eastern Cereal and Oilseed Research Centre,
Agriculture and Agri-Food Canada, Central Experimental Farm,
960 Carling Ave, Ottawa, ON K1A 0C6, Canada

Introduction

Barley (*Hordeum vulgare* L.) was domesticated approximately 10,000 years ago and stands among the four most important cereal crops today (FAO; http://www.fao.org/). Breeding programs around the world are developing improved barley varieties with better quality, disease resistance, and agronomic traits. Detection of loci controlling these complex traits, i.e. quantitative trait locus (QTL) mapping, opens the way to marker-assisted selection. In barley, numerous QTL mapping studies have been performed. However, conventional QTL mapping efforts using segregating progeny of a bi-parental cross allow the detection of only a subset of loci/alleles within the crop and offer a limited resolution due to the small number of informative recombination events between linked genetic loci. Recently, linkage disequilibrium (LD) mapping has been used to study the genetics of complex traits in agricultural crops such as rice, maize, and barley (Iwata et al. 2007; Remington et al. 2001; Kraakman et al. 2004). These association studies make much broader use of available germplasm, thus ensuring a more comprehensive and

precise mapping of QTLs. Recent studies in barley suggest that LD mapping could be a valuable approach for identifying QTLs for complex traits such as yield and adaptation (Kraakman et al. 2004; Caldwell et al. 2005; Rostoks et al. 2006).

However, the resolution and power of association studies in a collection of cultivars depend on the extent of LD across the whole genome. LD is affected by many factors, such as allele frequency and recombination rate, population structure and size, mutation, mating system and admixture (see review by Flint-Garcia et al. 2003). Therefore, a thorough understanding of population structure and patterns of LD across the whole genome will be a key to designing and carrying out association mapping.

Of the various factors affecting the extent of LD, reproductive behavior would seem to play an important role in plants. Highly inbreeding species tend to exhibit a higher degree of LD. In barley, studies with a limited number (<200) of AFLP or SSR markers suggested that LD can extend up to 10 cM in this species (Kraakman et al. 2004; Maliysheva-Otto et al. 2006). With a total genetic size estimated at 1155 cM,[1] and a physical size estimated at 5000 Mbp,[2] a 10 cM genetic distance in barley can represent a vast amount of physical distance. In *Arabidopsis thaliana*, another highly inbreeding species, Nordborg et al. (2002) documented even longer stretches of LD (>50 cM). With an estimated 250 kbp per cM (Jander et al. 2002), distances of this magnitude in *A. thaliana* actually represent much smaller physical distances than in barley. In contrast, Remington et al. (2001) found that LD in maize (a highly outbreeding species) was limited to a few thousand base pairs. With a total genetic map estimated at 1750 cM (see footnote 1), and a physical size of >2400 Mbp (see footnote 2), this physical distance corresponds to fractions of a cM.

In order to achieve a high resolution for association analysis, large numbers of molecular markers are needed. For example, it has been estimated that complete coverage of the maize genome could require between 5,000 and 750,000 markers depending on the germplasm used (Flint-Garcia et al. 2003). In barley, Rostoks et al. (2006) estimated that approximately 1,000 evenly distributed markers could be sufficient to ensure good coverage of the genome. This estimate is based primarily on the objective of finding markers with detectable genetic linkage to QTL with moderate effects. Higher numbers of markers may be required to initiate physical map-based cloning, and this density requirement can vary depending on factors such as genome location, which can influence the relationship between physical and genetic distance. Fortunately, the Diversity Array Technology (DArT) has been implemented in barley (Wenzl et al. 2004) and offers an inexpensive and rapid means to genotype over 2,000 marker loci in parallel across a large set of diverse germplasm.

Briefly, DArT analysis is based on detection of hybridization signals from a complexity-reduced genomic sample to an arrayed set of DNA clones, with each clone representing a potentially unique marker. The standard DArT protocols used in barley involved a complexity reduction based on amplification of *Pst*I/*Bst*NI enzyme-restricted fragments, as described by Jaccoud et al. (2001). The arrayed barley clones were developed through random isolation from a bulked genomic sample of diverse barley varieties subjected to an identical complexity reduction (Wenzl et al. 2004). Most of the clones on the standard barley DArT array have been assigned a position on a consensus map (Wenzl et al. 2006) totaling almost 3,000 markers of various types (DArT, SSR, RFLP, and STS). The polymorphisms detected in DArT analysis include SNPs, insert-deletion events, and heritable methylation changes (Jaccoud et al. 2001).

The objectives of this study were (1) to investigate the population structure and genetic diversity in Canadian barley using DArT markers and (2) to estimate the extent of LD in the barley genome as a first step toward the goal of performing association mapping in future research.

Materials and methods

Plant material

A set of 170 barley cultivars were selected to represent germplasm in use across Canada in the last 20 years. This set was split approximately equally between two-row and six-row types (84 and 86 lines, respectively) and between geographic areas (79 from eastern Canada, 83 from western Canada, and 8 accessions from elsewhere in the world). These foreign accessions were included, as they constitute many of the most frequently used sources of resistance to Fusarium head blight (FHB). A complete listing of these genotypes is provided in Table 1.

DArT genotyping

DNA was extracted from young leaf tissue from a single plant of each genotype using the protocol recommended by

[1] Due to variation in map lengths from different sources, total map distances were estimated as the average map lengths based on all maps presented at 'Plant Genome Central', NCBI (http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html) as of 2009-Feb-02.

[2] Estimated genome sizes were obtained from 'Plant Genome Central' at NCBI (http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html) as of 2009-Feb-02.

Table 1 Barley cultivars and lines used in this study

Two-row		Six-row	
Eastern	*Western* (*cont.*)	*Eastern*	*Eastern* (*cont.*)
AB241-1	Deuce	AC Alma	UL028.3
AC Kings	Ellice	AC Klinck	UL033.9
AC Sirius	Elrose	AC Legend	UL052.10
AC Sterling	EX645LD	AC Malone	UL075.8
Albany	Fairfield	AC Nadia	Viviane
Almonte	H93124008	AC Stephen	Yielder
Birka	Harrington	AC Westech	21FHD-2
Bristol	HB348	ACCA	21FHD-7
CFO284AA6	HDE84194-622-1	Bruce	
CH9419-9	Hector	Brucefield	*Western*
Chief	Herta	Cadette	
Craig	Klages	Célinex	AC Ranger
Helena	Manley	Corcy	AC Rosser
Iona	MC0040-42	CFO300AA12	Argyle
Island	McLeod	Chambly	B1602
Lester	Newdale	Chapais	Bedford
Micmac	Norbert	Cyane	Bonanza
Morrison	Phoenix	Cybel	Brier
Rodeo	Ponoka	Encore	CDC Battleford
Sabrina	Seebe	Étienne	CDC Clyde
Sunderland	SH00844	Foster	CDC Earl
Symko	SH01431	Gamine	Conquest
T303-013	SH01690	Joly	Creme
Volla	Stein	Laurier	Diamond
Winthrop	TR02185	Léger	Duel
	TR03373	Loyola	Duke
Western	TR04282	Lucky	Empress
	TR04378	Maskot	Galt
Abee	TR04719	Massey	Gateway 63
AC Metcalfe	TR05287	Mingo	Heartland
B1215	TR251	Myriam	Johnston
Betzes	TR306	OAC Chesley	Klondike
BM9732-491-1	Xena	OAC Kawarta	Leduc
BM9733-130-2	03T505-01-038	OAC Kippen	Melvin
BM9764-493	2^{ND}16092	Océanik	Niska
Bridge		Païdia	Noble
CDC Bold		Perseis	Samson
CDC Candle		Perth	Tankard
CDC Copeland	*Alien lines*	Synasolis	Tradition
CDC Cowboy		Raquel	Virden
CDC Dolly	CIho4196	Rosalie	Vivar
CDC Fleet	CIho9831	Sabina	
CDC Freedom	Doneckij 6	Sandrine	*Alien lines*
CDC Rattan	Fredrickson	Sophie	
Centennial	Svanhals	Sumosan	Chevron
Conlon	Zhedar 1	UL016.6	F104-250-9

Triticarte Pty (http://www.triticarte.com.au). A total of 188 DNA samples (2 × 94) were sent to Triticarte Pty for DArT analysis using the Barley *Pst*I(*Bst*NI) v2.0 array which comprises 2,304 clones known to be polymorphic in a wide range of barley cultivars (Wenzl et al. 2004). For 10 lines, DArT data had to be discarded either because DNA quality was insufficient and led to too many missing data or because there was a suspected error in the seed stock obtained (plant material did not match its description). To verify the reproducibility of the genotyping, seven samples were analyzed in duplicate (six genotypes) or triplicate (one genotype) using samples that were not identified to the data provider.

Statistical analyses

The polymorphism information content (PIC) values were calculated for each DArT marker using the formula $PIC = 1 - \sum(Pi)^2$, where Pi is the proportion of the population carrying the ith allele (Botstein et al. 1980). Consistent 0/1 data matrices were used as input for population structure analysis. The GENESTAT (version 9.2) analysis software was used to perform principal coordinate analysis (PCoA) using a genetic similarity matrix based on the Nei-Li genetic similarity (Nei and Li 1979), abbreviated here as d. The PAUP software package (Swofford 2003) was used to construct an Unweighted Pair Group Method with Algorithmic Mean (UPGMA) dendrogram based on d. The reliability and goodness of fit of dendrograms obtained from DArT data were tested through bootstrapping based on 100 samples (Felsenstein 1985). This led to 100 dendrograms summarized in a consensus tree that indicated the proportion of bootstrapped trees showing that same clade.

An analysis of molecular variance (AMOVA) was carried out using ARLEQUIN v3.11 to estimate genetic variance components and to partition the total variance within subgroups, among subgroups, and among groups. Significance of variance components was tested using 1,000 permutations.

LD between pairs of polymorphic loci was evaluated using the software package TASSEL1.9.4 (http://www.maizegenetics.net/). LD was estimated using the squared allele frequency correlations (r^2), which is a measurement of the correlation between a pair of variables (Hill and Robertson 1968). Decay of LD with genetic distance was estimated by nonlinear regression (SPSS Version 10.0) following the methods of Remington et al. (2001). The expected decay of LD was modeled as per Weir and Hill (1986)

$$E(r^2) = \left[\frac{10+c}{(2+c)(11+c)}\right]\left[1+\frac{(3+c)(12+12c+c^2)}{n(2+c)(11+c)}\right]$$

where n is the sample size and the nonlinear regression yields a least-squares estimate of c per distance (cM).

Results

Reproducibility and polymorphism of DArT markers

After initial quality screening based on standard methods optimized for barley at Triticarte, and after merging data from two separately analyzed plates, scores for 942 markers were obtained across 170 different barley lines. In order to investigate the reproducibility of DArT markers, identical samples were replicated in the same genotyping plate (three pairs) or on different plates (four pairs) with one genotype (Chapais) appearing in both sets such that it was analyzed in triplicate. For the complete set of 942 DArT markers scored as polymorphic on both genotyping plates, 99.93% of allele calls were identical for genotypes replicated on the same plate, while 99.83% were identical for genotypes replicated on two different plates. These results confirm that the DArT genotyping platform offers excellent reproducibility.

The PIC values were computed for the 942 polymorphic markers and ranged from 0.04 to 0.5 (the maximal value for a biallelic marker), with a mean value of 0.38. Mean PIC values were computed for each of the seven chromosomes, and these were found to be quite similar for six of the chromosomes (0.35–0.39) whereas chromosome 4H presented a significantly higher mean PIC value (0.46).

Population structure

The PCoA analysis was used to examine population structure in the collection. The PCoA analysis, based on decomposition of any multidimensional distance metric (in this case *d)*, is similar to the more familiar Principle Component Analysis (PCA), which is based on Euclidean coordinates. A two-dimensional scatter plot of the 170 barley genotypes is shown in Fig. 1. The first principal coordinate (PCo1) accounted for 14.9% of the variation and clearly separated two-row and six-row cultivars with a few exceptions. Joly and Niska, two-six-row genotypes fell

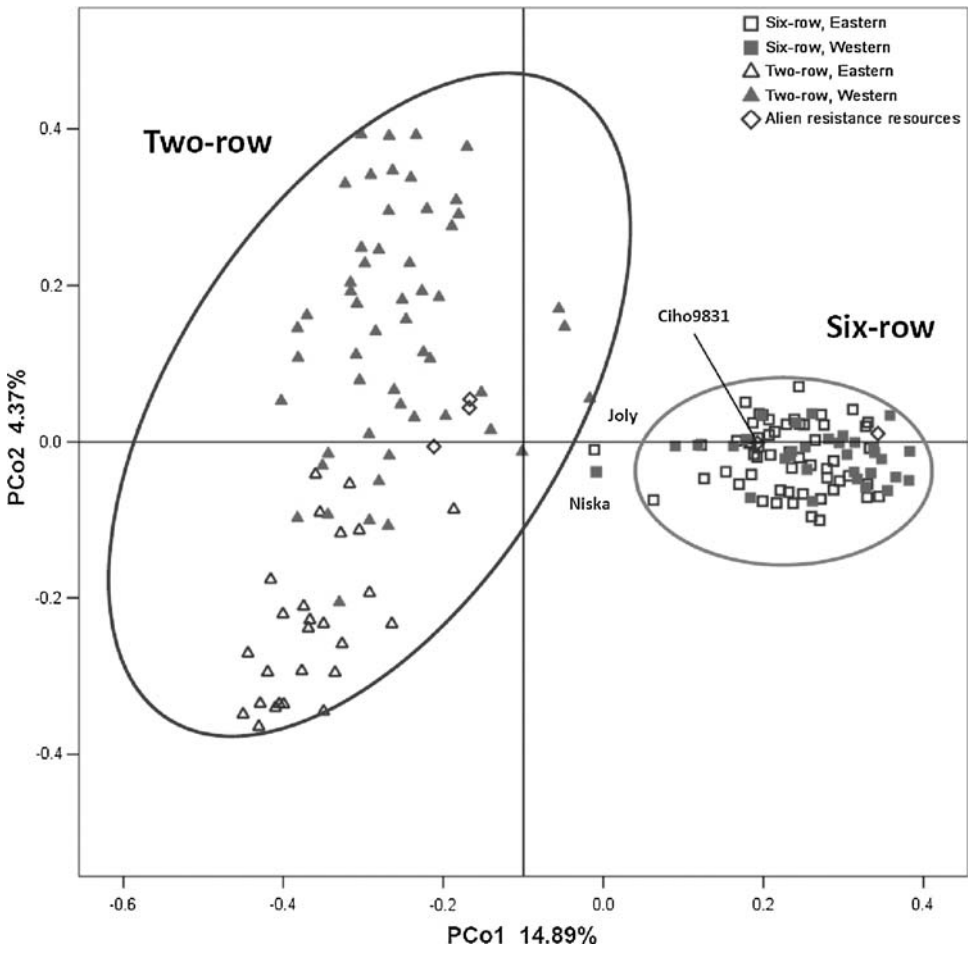

Fig. 1 PCA using pair-wise similarity for 170 barley varieties

somewhat in between the two clusters. Also, one-two-row genotype (CIho9831) fell well within the six-row cluster. The second principal coordinate (PCo2) accounted for 4.4% of the variation. The six-row group proved to be highly clustered around both axes with no obvious separation of genotypes originating from western and eastern Canada. In contrast, the two-row group was evenly spread along the PCo2 axis, with most genotypes from western Canada located in the upper left quadrant and genotypes from eastern Canada falling into the lower left quadrant. Varieties with known resistance to FHB fell within their respective groups (two-row/six-row) with the sole exception of CIho9831 mentioned above.

Genetic relationships among barley cultivars

The Nei-Li genetic distance coefficients (*d)* ranged from 0.00034 to 0.125, with an overall mean of 0.0724. The greatest distance was observed between two western genotypes, Betzes (two-row) and Creme (six-row). A consensus dendrogram obtained from UPGMA of *d* was constructed and is shown in Fig. 2. Again, a clear separation of two-row and six-row types was observed, and bootstrap analysis also supported this clustering. Within the group formed by two-row lines, three sizable (>5 genotypes) sub-groups can be distinguished (SG1–SG3). The first, SG1, is composed exclusively of lines from western Canada while SG2 comprises all the two-row lines from eastern Canada as well as a similar number of lines from western Canada. The lines within SG3 are primarily those with reported FHB resistance (HDE84194, CIho4196, Zhedar1 and Frederickson), FHB-resistant derived lines (F104-250-9 and MC0040-42), and one other genotype (H931240008). Joly and Niska, two-six-row genotypes that clustered with the two-row types in the PCoA analysis, form a distinct branch in the cluster of two-row genotypes. Within the six-row branch of the tree, genotypes fell into four subgroups (SG4-SG7). SG4 consists of 27 cultivars that are predominantly from western Canada while SG5 contains 30 cultivars that were almost exclusively from eastern Canada. The final two subgroups (SG6 and SG7) consist primarily of eastern Canadian cultivars.

Analysis of molecular variance

As described above, seven important subgroups were identified in the cluster analysis and these were used to perform an AMOVA on the collection of lines. The collection was considered to be composed of two groups (based on row type) and seven subgroups. Highly significant ($P < 0.001$) genetic variance was observed within subgroups, among subgroups, and among groups (Table 2). The variance within subgroups accounted for the largest portion (58.3%) of total variance whereas 24.8% of the variance was observed among groups and differences among subgroups contributed 16.8% of the total variance. The estimated F_{st} value of 0.25 suggests that this germplasm is highly differentiated. A total of 428 loci (45.4% of those scored) differed significantly in allelic frequency among the various subgroups. Of these, one locus on chromosome 1H, two loci on chromosome 5H, and one locus on chromosome 7H were responsible for most of the differentiation, accounting for 62–73% of the variance. Among groups, 29.6% (279/942) of the loci differed significantly and, of these, two loci located on 1H (at 66 and 106 cM), one locus located on chromosome 2H (at 70 cM), as well as one locus on chromosome 3H (at 145 cM) accounted for 91–98% of the total variance.

Linkage disequilibrium

Of the 942 polymorphic loci, 59 were not used for LD analysis because their map location was unknown. To prevent the detection of spurious LD, the set of DArT markers was expunged of those presenting a minor allele frequency (MAF) below 0.1. Of the 883 polymorphic markers of known map position, only 37 were removed on account of a MAF < 0.1 within the entire set of barley lines. Pair-wise LD was estimated using the squared-allele frequency correlations (r^2) and was found to decay rapidly with genetic distance (Fig. 3). The r^2 values for pairs of loci located on the same chromosome (intrachromosomal LD) ranged between 0 and 1 with an average of 0.07. The mean distance at which r^2 values fell below 0.2 was 2.6 cM (Table 3). Averaged over all chromosomes, 8.0% of intrachromosomal locus pairs exhibited r^2 values exceeding 0.2 and these showed an average r^2 value of 0.47. Interchromosomal LD was lower as only 3.4% of loci pairs had r^2 values exceeding 0.2, and the average r^2 value for these pairs was 0.29.

As our population of barley lines seemed to be highly structured based on inflorescence type, a similar exercise was conducted to remove markers whose MAF was <0.1 among either the two-row or six-row groups. This resulted in the removal of 211 and 200 markers, respectively, i.e. about a quarter of the 883 DArT markers. Two different scenarios were encountered. In 37 cases (those already described above), one allele was rare throughout the entire collection. In the other cases, the two alleles were equally frequent when considered over the entire population, but were distributed very unevenly between two-row and six-row lines. In extreme cases, one allele was present only in one group while the other allele was uniquely present in the other. Of the seven chromosomes, 1H and 2H had the greatest proportion (40% and 32%, respectively) of markers with MAF <0.1 in the two-row or six-row group.

After removal of markers with MAF <0.1 within groups, we then examined LD within each of these two subsets and

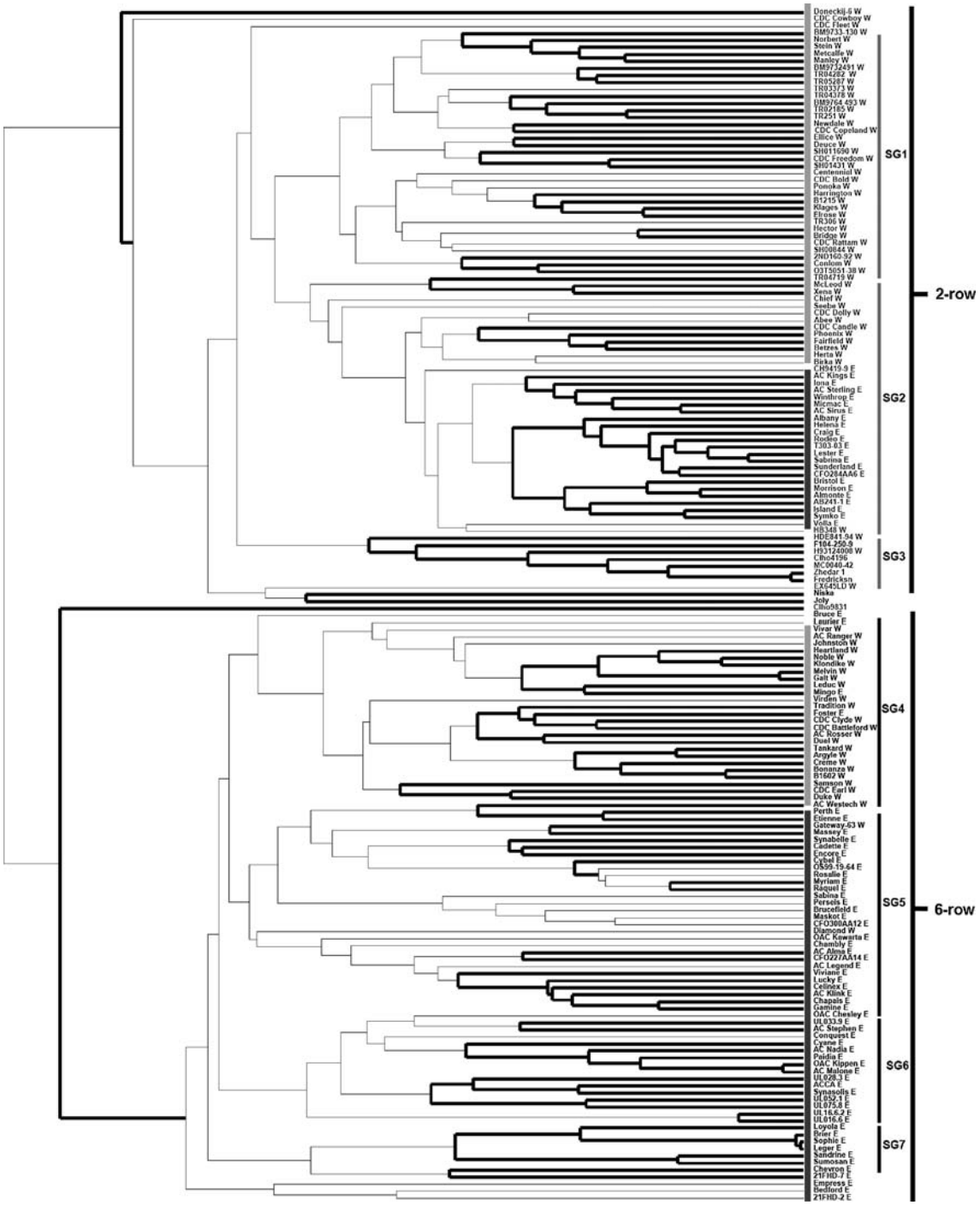

Fig. 2 Dendrogram representing the relationships among the 170 Canadian barley cultivars as revealed by UPGMA cluster analysis based on Nei/Li genetic distances. Bootstrap support values (100 replicates) are shown as *bold lines* when >50%. Two-rowed genotypes are in gray type, while 6-rowed genotypes are in black. The letter in parentheses indicates if a cultivar comes from Western (W, *gray*) or Eastern (E, *black*) Canada

found that it decayed more slowly than it did in the global analysis, extending to 3.4 cM in the two-row lines and 3.5 cM in the six-row lines. At the intrachromosomal level, LD between locus pairs was very similar to that found for the entire population; in both the two-row and six-row groups, 7.5% of the loci pairs had r^2 values >0.2 (Table 3) and mean r^2 values in such cases were, respectively, 0.51 and 0.53. As expected, the removal of markers with a highly skewed distribution between the two groups reduced the proportion of interchromosomal locus pairs with r^2

Table 2 Analysis of molecular variance for 170 barley cultivars

Source of variation	Degrees of freedom	SS	CV	% Total	*P*
Among groups	1	4299.82	42.20	24.82	0
Among subgroups within groups	5	3430.61	28.64	16.84	
Within subgroups	153	15172.81	99.17	58.33	
Total	159	22903.24	170.00		
Fixation index (Fst) = 0.42					

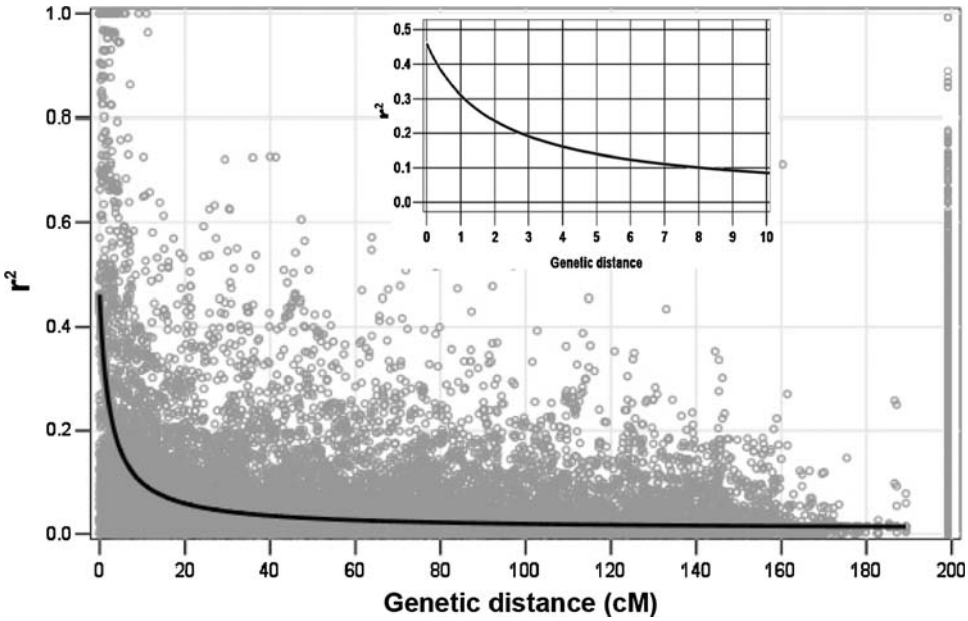

Fig. 3 Intrachromosomal LD (r^2) decay plot of marker pairs on all seven chromosomes as a function of genetic distance (in cM) for the 170 barley lines. The curve illustrates LD decay based on the nonlinear regression of r^2 on genetic distance, as calculated using the equation of Hill and Weir (1988). The observed LD values for unlinked markers are compiled in single file at the 200 cM mark. The inset shows a more detailed view of the LD decay curve for markers located <10 cM apart

Table 3 Percentage of loci pairs and mean r^2 in intrachromosomal and interchromosomal LD ($P < 0.001$, $r^2 \geq 0.2$) for the whole dataset as well as the two-row and six-row subsets

LD ($P < 0.001$)	Whole set	Two-row subset	Six-row subset
Intrachromosomal			
Percentage of loci pairs in LD ($r^2 \geq 0.2$)	7.96%	7.50%	7.48%
Mean r^2	0.47	0.51	0.53
Interchromosomal			
Percentage of loci pairs in LD ($r^2 \geq 0.2$)	3.43%	1.52%	0.55%
Mean r^2	0.29	0.26	0.25
Mean distance at which $r^2 = 0.2$ (cM)	2.6	3.4	3.5

For each set, the average distance (in cM) at which $r^2 = 0.2$ is indicated

values above 0.2. Among the two-row group, this fell to 1.5% while among the six-row group it dropped to only 0.6%. Similarly, the mean r^2 values also decreased to 0.26 and 0.25, respectively.

One case of complete LD ($r^2 = 1$ with $P = 0$) between loci on different chromosomes was observed consistently in the entire collection as well as the two-row and six-row groups. This pair involved loci bPb-9788 on chromosome 1H (27.5 cM) and bPb-7626 on chromosome 2H (21.9 cM).

Discussion

Comparison of DArT and SNP markers for genome-wide scans

DArT markers have been demonstrated to be useful in the current study, and will be useful in the context of integrating this work with future objectives and marker-based results in the barley research community. DArT markers proved cost-effective and ideally suited to high-throughput

parallel analysis that was not dependent on in-house infrastructure. These results will be cross-applicable with many other projects in barley that are now using DArT-based markers. Furthermore, when DArT clone sequences are available, these will provide access to other methods for marker design and for orthology-based comparison among species. A recent analysis of DArT clone sequences in oat has shown that over one-third of these markers contain DNA sequence with strong homology to functionally annotated genes (Tinker et al. 2009). Forthcoming analyses of DArT clone sequences from barley are expected to provide similar results.

We found that DArT markers offer a highly reproducible tool for genome-wide analyses in barley. When replicated on the same or different plate, allele calls proved identical in close to 99.9% of cases. Such high reproducibility is similar to what has been reported for SNPs (Bao et al. 2005). Among our set of cultivated barleys from Canada, approximately 1,000 markers proved polymorphic among the ~2,300 amplicons found on the DArT barley array. To obtain a similar number of informative markers, one would likely need to use at least one oligonucleotide pool assay (OPA) of 1,536 SNPs using the Illumina Golden Gate technology.

Another aspect to consider is the extent of coverage offered by the markers. DArT markers do suffer from some amount of clustering (Wenzl et al. 2006) whereas it is, in theory, easier to select SNPs that are evenly distributed when assembling an OPA. To estimate the coverage afforded by the 883 DArT markers used to analyze LD in our collection, we assumed that markers spaced <5.2 cM apart offer a good coverage of the intervening region, given that LD ($r^2 > 0.2$) extends on average to 2.6 cM. By adding up the size of all gaps in excess of 5.2 cM, we estimate that coverage totaled 86.2% over the entire genome and varied between 54.2% on chromosome 4H and 99.1% on chromosome 1H. Unfortunately, in previous work in barley using SSRs and SNPs (Maliysheva-Otto et al. 2006; Russell et al. 2004), no similar estimation of genome coverage has been provided.

Population structure in Canadian barley

Using various approaches (PCoA, UPGMA, and AMOVA), we found that the inflorescence type is a key determinant of population structure within Canadian barley. With very few exceptions, lines were very neatly categorized as belonging to the two- or six-row type based on their marker genotype. This occurred despite the fact that the region around the *VRS1* locus (the main locus controlling inflorescence type) is not very densely populated with DArT markers. On the consensus map of Wenzl et al. (2006), *Vrs1* is located at position 92.6 on chromosome 2H. The closest polymorphic markers in this work were bPb-6055 (90.5 cM) and bPb-7991 (101.3 cM). Based on the AMOVA, we found that four markers (two on 1H, one on 2H, and one on 3H) were responsible for most of the separation between these two groups. This is very similar to the observations of Hayes and Szucs (2006) on North American barleys using SNPs. A less striking contrast between two-row and six-row barleys was seen in other work focusing either mainly on European lines (Rostoks et al. 2006) or on a worldwide collection of 953 accessions (Maliysheva-Otto et al. 2006). In the latter cases, more overlap between the two-row and six-row lines was seen and lines also clustered according to growth habit (spring versus winter). In Canada, as only spring barley is grown, the inflorescence type appears to be the predominant feature separating groups. This suggests that breeders tend to work preferentially within each of these two groups and seldom exploit two-row by six-row crosses.

Another striking feature is the difference in genetic diversity within the six-row and two-row groups. Whereas six-row lines are tightly clustered together irrespective of their origin (eastern or western Canada), two-row lines are much more diverse and there is a clear clustering of lines according to geography. This could suggest that two-row barleys have been bred using more diverse germplasm. This is consistent with the fact that in most barley-growing regions outside of North America, two-row barley is dominant and offers a rich array of genotypes to use in local breeding programs.

Among Canadian lines, two-six-row genotypes were consistently positioned between clusters of six-row and two-row types. In a separate analysis conducted using 50 SSR markers on a similar set of Canadian barley lines, Lamara et al. (unpublished) also found that these two genotypes were intermediate. Niska has a complex pedigree including both six-row and two-row parents and, despite its inflorescence type, it may have retained certain genome blocks that are more typical of two-row types. However, nothing in the pedigree of Joly indicated a contribution from a two-row parent. Another set of lines that clustered somewhat separately was comprised of FHB resistance donors. Most of these lines originate from outside of North America and are likely to be distinct from the Canadian germplasm. This situation is likely to have some effect on the mapping of QTLs for FHB resistance within such a collection of germplasm. If these foreign lines are the predominant source of FHB resistance alleles and because this germplasm often exhibits distinctive alleles compared to the Canadian lines, markers linked to resistance loci may be excluded from association analyses on the basis of the low frequency of such alleles (when MAF <0.1; see below for more on this point).

Linkage disequilibrium

In the present work, LD was measured both in the entire collection of lines and within the two-row and six-row subsets. LD was found to decrease below a critical value ($r^2 = 0.2$) when the genetic distance between neighboring markers exceeded 2.6 cM (in the entire set) or ~3.5 cM (in the two subsets). This represents a more rapid decay than what has been reported in previous studies in barley. In the work that is most comparable to this one in terms of the number of markers assayed (656 SNPs), Rostoks et al. (2006) reported that the distance between markers presenting an r^2 value above 0.5 was 3.9 cM. At a lower level of correlation (e.g., $r^2 > 0.2$), this distance would necessarily have been greater. Using 134 AFLP markers on 146 elite European barley cultivars, Kraakman et al. (2004) reported that LD ($r^2 > 0.2$) rarely extended beyond 10 cM, but do not provide an estimate of the mean distance at which LD falls below such a threshold. In the work of Maliysheva-Otto et al. (2006), the small number of markers used (48 SSRs) made it difficult to detect high r^2 values as markers are obviously relatively distantly spaced. In fact, no r^2 values above 0.2 were found among any of the pairs of loci located on the same chromosome. The fact that we observed a more rapid decline in LD among our set of lines may reflect the larger size (170 vs. 102 lines) or greater genetic diversity of our population relative to that of Rostoks et al. (2006).

The structure of the population had an important impact on the extent of LD, especially of the interchromosomal type. As described in the results section, the proportion of intrachromosomal locus pairs in LD ($r^2 > 0.2$) remained relatively constant (7.5–8.0%) in the entire population and in the two subsets. The proportion of interchromosomal locus pairs in LD, however, dropped from 3.4% in the entire set to 1.5% and 0.5% in the two-row and six-row subsets, respectively. A similar reduction in interchromosomal LD has been reported by Rostoks et al. (2006). In their case, interchromosomal pairs represented over 20% of all highly significant ($r^2 > 0.5$) associations in the entire set of 102 lines, but only 2% within the set of 53 two-row lines. This likely results from the elimination of much spurious interchromosomal LD that is due to uneven distribution of the alleles within the two-row and six-row subsets. All loci, irrespective of their map location, will appear to be in LD with one another if they are nearly monomorphic for alternate alleles within each subset of lines.

We also documented one case of extreme interchromosomal LD where bPb-9788 on chromosome 1H and bPb-7626 on chromosome 2H showed absolute LD ($r^2 = 1$). When the DArT data were examined carefully, it was found that markers immediately flanking these two markers were not in LD. One possible explanation for this is that certain DArT clones can occasionally reveal polymorphism at two or more loci; thus a clone may identify different loci in different populations. In one biparental mapping population, the polymorphism being scored resides at one locus and, in another population, this same locus is fixed for the 'absent' allele and the polymorphism scored is due to a completely different locus. In cases where this has been documented, Triticarte indicates the two known map locations for a clone on the array. In our case, the map position for the two loci mentioned above is based on the polymorphism segregating in the biparental mapping populations described in Wenzl et al. (2006). It may be that the polymorphism being scored among our Canadian barley genotypes is different at one of these two loci. Thus, it is possible that the two loci are in fact tightly linked, which would explain the high LD observed. In any case, the fact that some DArT clones may provide information on different loci, depending on the materials under study, must be kept in mind.

The high incidence of nonlinkage-based disequilibrium that was observed in the full population has important implications for genome-wide association mapping in barley. Generally, two types of association mapping strategies are envisioned, and these may be most effectively deployed in a two-tiered strategy, as suggested by Caldwell et al. (2006). In a first tier, a low-resolution whole genome scan would be performed within a collection of elite germplasm from a single adaptation-type (e.g., six-row spring). This population would contain relatively far-reaching LD, but with lower incidence of spurious effects caused by population structure. If the objective were to identify QTLs for marker-assisted selection, the resolution provided by this approach might be sufficient. Otherwise, a second step would be conducted using a very diverse set of lines (e.g., including landraces or wild barley accessions) in which LD would be expected to extend over much shorter distances, and where the analysis could be restricted to a specific chromosome region.

In any association analysis, it is also important to consider the potential for pleiotropic effects (e.g., of row type or winter habit) on traits that are analyzed. It is speculated, for example, that FHB susceptibility may be partially related to the compactness of florets in the 6-row barley type. Thus, restricting a first-stage analysis to a single row type would provide better resolution of traits affected by this character. Later analyses that employ more diverse sets of germplasm could then be targeted to avoid discovery of unlinked loci, but the success of this strategy would also depend on whether alleles at the targeted locus are present in the expanded population, and whether their effects were substantial relative to other pleiotropic effects.

In conclusion, DArT markers have proven to be an effective tool to accurately evaluate genetic relationships,

population structure, and the extent of LD in barley. Marked population structure has been revealed effectively by these biallelic markers. The presence of this population structure will affect how future strategies for association mapping are designed and deployed, and it may reduce the number of polymorphic markers that can usefully contribute to association analyses.

References

Bao YP, Huber M, Wei TF, Marla SS, Storhoff JJ, Müller UR (2005) SNP identification in unamplified human genomic DNA with gold nanoparticle probes. Nucleic Acids Res 33(2):e15

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

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

Felsenstein J (1985) Confidence-limits on phylogenies—an approach using the bootstrap. Evolution 39:783–791

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

Hayes P, Szucs P (2006) Disequilibrium and association in barley: thinking outside the glass. Proc Natl Acad USA 103(49):18385–18386

Hill WG, Robertson A (1968) The effects of inbreeding at loci with heterozygote advantage. Genetics 60(3):615–628

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33(1):54–78

Iwata H, Uga Y, Yoshioka Y, Ebana K, Hayashi T (2007) Bayesian association mapping of multiple quantitative trait loci and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms. Theor Appl Genet 114:1437–1449

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) Arabidopsis map-based cloning in the post-genome era. Plant Physiol 129:440–450

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 168:345–446

Maliysheva-Otto LV, Ganal MW, Roder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). BMC Genet 7:1–14

Nei M, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonuleases. Proc Natl Acad USA 76:5269–5273

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190–193

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad USA 98(20):11479–11484

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad USA 103(49):18656–18661

Russell J, Booth A, Fuller J, Harrower B, Hedley P, Machray G, Powell W (2004) A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 47(2):389–398

Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods), vol 4. Sinauer Associates, Sunderland, MA

Tinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P, Rines HW, Bjornstad A, Howarth CJ, Jannink JL, Anderson JM, Rossnagel BG, Stuthman DD, Sorrells ME, Jackson EW, Tuvesson S, Kolb FL, Olsson O, Federizzi LC, Carson ML, Ohm HH, Molnar SJ, Scoles GJ, Eckstein PE, Bonman JM, Ceplitis A, Langdon T (2009) New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genom 10:39

Weir BS, Hill WG (1986) Nonuniform recombination within the human beta-globin gene cluster. Am J Hum Genet 38(5):776–781

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci USA 101(26):9915–9920

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesna J, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genom 7:206