Theor Appl Genet (2006) 113:1409–1420
DOI 10.1007/s00122-006-0365-4

ORIGINAL PAPER

Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome

Mona Akbari · Peter Wenzl · Vanessa Caig · Jason Carling · Ling Xia · Shiying Yang · Grzegorz Uszynski · Volker Mohler · Anke Lehmensiek · Haydn Kuchel · Mathew J. Hayden · Neil Howes · Peter Sharp · Peter Vaughan · Bill Rathmell · Eric Huttner · Andrzej Kilian

Received: 8 November 2005 / Accepted: 6 July 2006 / Published online: 11 October 2006

Abstract Despite a substantial investment in the development of panels of single nucleotide polymorphism (SNP) markers, the simple sequence repeat (SSR) technology with a limited multiplexing capability remains a standard, even for applications requiring whole-genome information. Diversity arrays technology (DArT) types hundreds to thousands of genomic loci in parallel, as previously demonstrated in a number diploid plant species. Here we show that DArT performs similarly well for the hexaploid genome of bread wheat (*Triticum aestivum* L.). The methodology previously used to generate DArT fingerprints of barley also generated a large number of high-quality markers in wheat (99.8% allele-calling concordance and approximately 95% call rate). The genetic relationships among bread wheat cultivars revealed by DArT coincided with knowledge generated with other methods, and even closely related cultivars could be distinguished. To verify the Mendelian behaviour of DArT markers, we typed a set of 90 Cranbrook × Halberd doubled haploid lines for which a framework (FW) map comprising a total of 339 SSR, restriction fragment length polymorphism (RFLP) and amplified fragment length polymorphism (AFLP) markers was available. We added an equal number of DArT markers to this data set and also incorporated 71 sequence tagged microsatellite (STM) markers. A comparison of logarithm of the odds (LOD) scores, call rates and the degree of genome coverage indicated that the quality and information content of the DArT

Communicated by F. Ordon.

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-006-0365-4 and is accessible for authorized users.

M. Akbari · P. Wenzl · V. Caig · J. Carling · L. Xia · S. Yang · G. Uszynski · N. Howes · P. Sharp · P. Vaughan · B. Rathmell · E. Huttner · A. Kilian (✉)
Triticarte P/L, 1 Wilf Crane Crescent, Yarralumla, Canberra, ACT 2600, Australia
e-mail: a.kilian@DiversityArrays.com

P. Wenzl · V. Caig · J. Carling · L. Xia · S. Yang · G. Uszynski · E. Huttner · A. Kilian
Diversity Arrays P/L, 1 Wilf Crane Crescent, Yarralumla, Canberra, ACT 2600, Australia

V. Mohler
Department of Plant Breeding, Technical University Munich, Am Hochanger 2, 85350 Freising, Germany

A. Lehmensiek
Faculty of Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia

H. Kuchel
Australian Grain Technologies P/L, University of Adelaide, Roseworthy, SA 5371, Australia

V. Mohler · M. J. Hayden · N. Howes · P. Sharp · P. Vaughan · B. Rathmell
Value Added Wheat Cooperative Research Centre, Plant Breeding Institute, University of Sydney, PMB11, Camden, NSW 2570, Australia

Present Address:
M. J. Hayden
Molecular Plant Breeding Cooperative Research Centre, University of Adelaide, PMB1, Glen Osmond, SA 5064, Australia

data set was comparable to that of the combined SSR/RFLP/AFLP data set of the FW map.

Introduction

Crop improvement relies on the effective utilisation of genetic diversity. Molecular marker technologies promise to increase the efficiency of managing genetic diversity in breeding programmes. Numerous marker technologies have been developed over the last 25 years. The most widely used systems, adopted at different stages in the evolution of marker technologies, are restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP®), microsatellites or simple sequence repeats (SSR) and single nucleotide polymorphism (SNP) (Botstein et al. 1980; Weber and May 1989; Williams et al. 1990; Vos et al. 1995; Chee et al. 1996). These technologies can genotype agricultural crops with varying degrees of efficiency. They have various degrees of limitations associated with their capability to quickly develop and/or rapidly assay large numbers of markers. Although some of these limitations can be alleviated by equipment (e.g. highly parallel capillary electrophoresis), most of them are inherently linked to the sequential nature, low reproducibility, or high assay costs of the marker technologies, or the reliance on DNA sequence information.

Diversity arrays technology (DArT) was developed as a hybridisation-based alternative, which captures the value of the parallel nature of the microarray platform (Jaccoud et al. 2001). DArT simultaneously types several thousand loci in a single assay. DArT generates whole-genome fingerprints by scoring the presence versus absence of DNA fragments in genomic representations generated from samples of genomic DNA. The technology was originally developed for rice, a diploid crop with a small genome of 430 Mbp (Jaccoud et al. 2001). DArT was subsequently applied to a range of other crops. The list is expanding and currently includes 19 plant species and three fungal plant pathogens (Jaccoud et al. 2001; Lezar et al. 2004; Wenzl et al. 2004; Kilian et al. 2005; Wittenberg et al. 2005; Xia et al. 2005; Yang et al. 2006; Joseph Tohme, personal communication; DArT P/L and collaborators, unpublished data). Importantly, the large diploid genome of barley (5,200 Mbp) did not pose an obstacle to applying DArT to genetic mapping and diversity analyses (Wenzl et al. 2004).

However, to validate DArT as a generic tool for genotyping of plants, it is important to prove that the technology performs well in polyploid genomes. Such a test is important because the performance of a number of genotyping technologies seems to be adversely affected by the ploidy level of plant genomes. The presence of multiple copies of genes in polyploids represent a challenge even for SNP markers, although new approaches are being developed to deal with this issue in wheat (Mochida et al. 2004). Here we show that DArT can be effectively applied to the large 16,000-Mbp hexaploid genome of bread wheat (*Triticum aestivum*). We initially evaluate the performance of a wheat DArT array in terms of allele-calling efficiency and accuracy, and subsequently use the array for a diversity analysis of wheat cultivars, and to build a genetic map.

Materials and methods

Plant material

This study is based on a collection of 62 wheat cultivars and the doubled haploid (DH) population derived from a cross between cultivars Cranbrook and Halberd (Kammholz et al. 2001).

DNA extraction

The DNA was extracted from leaves of 2-week-old wheat seedlings using a modified cetyltrimethylammonium bromide method (Doyle and Doyle 1987). Some samples were extracted from root tissue by Jan Gooden at the South Australian Research and Development Institute (SARDI), GPO 397, Adelaide, SA 5001, Australia, using a proprietary DNA extraction procedure.

DArT procedure

Preparation of DArT arrays

Several DArT arrays were built in the course of this study. For each of these arrays, a genomic representation was generated from a mixture of wheat cultivars using the *Pst*I-based complexity reduction method described by Wenzl et al. (2004) (see Supplementary Table 1 for a list of cultivars used in this study). The procedure involved digestion of a mixture of DNA samples with *Pst*I and one of the following frequently cutting restriction enzyme: *Mse*I, *Rsa*I, *Taq*I or *Bst*NI (NEB, Beverly, MA, USA), ligation of a *Pst*I adapter with T4 DNA ligase (NEB), and amplification of small, adapter-ligated fragments (Wenzl et al. 2004).

A library was prepared from each of the amplification products, also called genomic representations, as described by Jaccoud et al. (2001) with the modifications of Wenzl et al. (2004). The cloned representation fragments were amplified (Jaccoud et al. 2001), dried at 37°C and dissolved in a spotting buffer which was initially either 50% (v/v) DMSO or a 1:1 mixture of 50% (v/v) DMSO and the printing buffer of the Vanderbilt Microarray Shared Resource (Nashville, TN, USA). We also used a new spotting buffer developed specifically for Erie Scientific (Portsmouth, NH, USA) poly-L-lysine microarray slides (Peter Wenzl et al. in preparation). The amplification products were printed on poly-L-lysine-coated slides (Polysine, Menzel Gläser, Braunschweig, Germany; CEL Associates, Pearland, TX, USA; or Erie Scientific) using either a GMS417 (Affymetrix, Santa Clara, CA, USA) or a MicroGridII arrayer (Biorobotics, Cambridge, UK). Depending on the number of clones, two or three replicate spots per clone were printed on the arrays. After printing, the slides were heated to 80°C for 2 h (unless spotted with the new buffer), denatured by incubation in hot water (95°C) for 2 min, dipped in 95% (v/v) ethanol (unless spotted with the new buffer) and dried by centrifugation.

Genotyping of DNA samples

Genomic representations of individual wheat cultivars were generated with the same complexity reduction method used to prepare the library spotted on the array. The resulting representations of individual lines were precipitated, denatured and labelled according to Wenzl et al. (2004) or using 0.3 µl of cy3-dUTP (Amersham Biosciences, Castle Hill, NSW, Australia) and the exo$^-$ Klenow fragment of *Escherichia coli* DNA polymerase I (NEB). Labelled representations, also called targets, were mixed with the cy5 or 6-FAM-labelled polylinker fragment of the plasmid used for library construction, and hybridised to slides (Wenzl et al. 2004). After overnight hybridisation at 65°C, the slides were washed according to Jaccoud et al. (2001) and scanned either on an Affymetrix 428 or a Tecan LS300 (Grödig, Salzburg, Austria) confocal laser scanner at the appropriate wavelengths (cy3: 543 nm; cy5: 633 nm; 6-FAM: 488 nm).

Image analysis and polymorphism scoring

Groups of two or three TIF images of individual slides were analysed using DArTsoft (Version 7), a purpose-built software package developed at DArT P/L, which is available to members of the DArT network in a collaborative spirit (see www.diversityarrays.com/dartnetwork.html). DArTsoft automatically analysed batches of up to 96 slides to identify and score polymorphic markers as described previously by Wenzl et al. (2004). The programme computed several quality parameters for each marker: (a) the between-allelic-states variance of the relative target hybridisation intensity as a percentage of the total variance (*P*-value), (b) the percentage of DNA samples with defined '0' or '1' allele calls (call rate) and (c) the fraction of concordant calls for replicate assays (C. Cayla et al. in preparation).

Complexity reduction testing

Four separate libraries, each containing 1,536 clones, were produced from genomic representations prepared from a DNA mixture of 13 Australia-grown cultivars (Amery, Angus, Condor, Cranbrook, Currawong, Frame, Grebe, Halberd, Janz, More, Sunland, Trident and Westonia; Supplementary Table 1). The libraries only differed in the frequently cutting restriction enzyme used to prepare the genomic representations (*Mse*I, *Rsa*I, *Bst*NI or *Taq*I). Targets generated from each of the 13 cultivars with either of four different complexity reduction methods were hybridised to matching arrays. In this experiment, polymorphic clones were identified with DArTsoft Version 6, using $P > 85\%$ and call rate > 80% as quality thresholds.

Development and quality evaluation of a *Pst*I/*Taq*I array

Two arrays containing a total of 9,216 randomly selected *Pst*I/*Taq*I clones were built from the group of 13 cultivars used for complexity reduction testing (Supplementary Table 1). These arrays were used to genotype 411 wheat lines to select high-quality polymorphic clones and to identify germplasm that was not sufficiently represented on the array (excess of '0' scores). Clones with a consistently high *P*-value, high call rate, and low calling discordance (1,681 in total) and a set of 384 consistently non-polymorphic clones were re-arrayed into a polymorphism-enriched library and a control-clone library, respectively (MicroGridII arrayer). An additional *Pst*I/*Taq*I library with 3,072 clones was prepared from a group of cultivars from underrepresented germplasm (Supplementary Table 1). Clones from this new library were arrayed together with clones from the polymorphism-enriched library. The resulting 'Version 2.0' array with a total of 5,137 clones was used for the reminder of the study.

To test the performance of this array, three replicated experiments were performed as described in the

section entitled DArT Procedure, with exception of using cy3-dUTP from Enzo (Farmingdale, NY, USA) for the third experiment. In each of these experiments, each of the 13 cultivars used for complexity reduction testing (Supplementary Table 1) was independently assayed three times ($13 \times 3 \times 3 = 117$ assays in total). Clones with a *P*-value greater than 70% and not more than a single discordant call across the nine replicate assays were selected as markers.

Cultivar diversity analysis

A group of 62 bread wheat cultivars was genotyped on the Version 2.0 array as described in the section entitled DArT Procedure. For quality control, ten cultivars were genotyped twice. Clones with $P > 77\%$, a call rate > 85 and 100% allele-calling consistency across the ten replicated assays were selected as markers. The marker scores were subjected to principal coordinate analysis to visualise the genetic relationships among the cultivars (Anderson 2003).

Genetic mapping

The cultivars Cranbrook and Halberd as well as 90 lines of a DH population derived from a cross between the two cultivars were genotyped with the Version 2.0 array as described in the DArT Procedure section. Clones with $P > 80\%$ and a call rate of at least 80% were initially selected for mapping; clones with *P* between 75 and 80% were later incorporated into the map allowing for a single double-crossover event per marker. In addition, sequence tagged microsatellite (STM) markers, developed according to Hayden et al. (2002), were amplified with fluorescently labelled primers and screened using a GelScan2000 DNA fragment analyser (Corbett, Sydney, NSW, Australia) as described by Hayden et al. (2004). The scores of all polymorphic markers were converted into genotype codes ('A', 'B') according to the scores of the parents.

Diversity arrays technology and STM segregation data were merged with the segregation data for 339 SSR, RFLP and AFLP markers of a recent framework (FW) map (Chalmers et al. 2001; Kammholz et al. 2001; Lehmensiek et al. 2005). JoinMap 3.0 was used to assign markers to linkage groups by employing logarithm of the odds (LOD) threshold values ranging from 3.0 to 5.0 (Stam 1993; van Ooijen and Voorrips 2001). JoinMap was also used to order markers within linkage groups, initially based on the preset default settings of the program, and subsequently with a LOD linkage threshold to 3.0. An alternative map was constructed with RECORD and a purpose-built perl script that computed map distances for the marker order reported by RECORD (van Os et al. 2005; Peter Wenzl, unpublished data). The perl script was based on the algorithm of Lalouel (1977), which is also used by JoinMap (Stam 1993). A comparison of the sum of adjacent recombination fractions, a sensitive indicator of map expansion due to sub-optimal marker order (Liu and Knapp 1990), indicated that the RECORD/perl map was superior to the map built with JoinMap, which showed 11.6% expansion compared to the RECORD/perl map (Supplementary Figure 1). The RECORD/perl map, therefore, was selected for this study. The graphical representation of the map was drawn using MapChart software (Voorrips 2002).

Results

Evaluation of complexity reduction methods

An important first step in the development of DArT for a new species is to determine which complexity reduction method generates a genomic representation that reveals a large amount of genetic polymorphisms. We previously identified restriction enzyme digestion with *Pst*I in combination with a second frequently cutting enzyme, followed by adapter ligation to *Pst*I ends and amplification of adapter-ligated fragments as one of the methods of choice for plant genomes (Kilian et al. 2005). We therefore tested the four frequent cutters that worked best for barley (*Mse*I, *Rsa*I, *Bst*NI and *Taq*I) for their efficiency of revealing genetic polymorphism in wheat (Wenzl et al. 2004). The arrays built from the four genomic representations clearly differed in the number of polymorphic clones identified (Table 1). The *Pst*I/*Taq*I library showed the highest polymorphism level (9.4%), followed by *Pst*I/*Bst*NI at 5.3%. The other two methods (*Pst*I/*Rsa*I and *Pst*I/*Mse*I) yielded less than half the number of polymorphic clones obtained with the best method. The *Pst*I/*Taq*I and *Pst*I/*Bst*NI complexity reduction methods were previously found to generate the most polymorphic representations in barley (Wenzl et al. 2004). While the two methods were similarly efficient in barley, *Taq*I as a co-digesting enzyme was clearly superior to *Bst*NI in wheat (Table 1).

Test of *Pst*I/*Taq*I array performance

Based on the comparison of complexity reduction methods, we expanded the *Pst*I/*Taq*I library in two steps designed to capture the genetic diversity of a broad spectrum of wheat cultivars. The final "Version

Table 1 Polymorphism levels obtained with four different complexity reduction methods

Complexity reduction method	Number of clones on array[a]	Number of polymorphic clones	Percentage of polymorphic clones
*Pst*I/*Bst*NI	1,504	80	5.3
*Pst*I/*Mse*I	1,280	54	4.2
*Pst*I/*Rsa*I	1,536	71	4.6
*Pst*I/*Taq*I	1,536	144	9.4

Each method was tested with a separate array constructed from a pool of DNA samples of 13 Australia-grown wheat cultivars representing the predicted diversity of the Australian germplasm, including the most common alien segments (Supplementary Table 1)

[a] All arrays contained 1,536 clones, each printed in triplicate; but the *Pst*I/*Bst*NI and *Pst*I/*Mse*I arrays had 32 and 246 control clones, respectively

2.0" genotyping array contained a total of 5,137 partly polymorphism-enriched clones (see Materials and methods and Supplementary Table 1).

Before using the Version 2.0 array for routine analyses of unknown samples, we tested allele-calling efficiency and consistency by typing each of the 13 cultivars used in the initial complexity reduction tests nine times (three separate experiments × three replicate assays per experiment). There were 648 markers with 100% calling concordance for all cultivars across the nine assays. An additional 140 markers had a one discordant call for a single cultivar in one of nine replicate assays (calling concordance = 99.1%: one discordant call amongst 13 × 9 = 117 calls). If marker selection criteria were relaxed further to allow up to two discordant intra-experiment calls, there were more than 1,000 markers that were fully concordant across the three experiments.

Polymorphic DArT markers are selected by simultaneously applying thresholds for three quality parameters: *P*-value, call rate and discordance (see Materials and methods). We examined the relationship among the values of these parameters averaged across the three experiments, by using the set of 788 polymorphic markers (648 markers with 100% calling concordance + 140 markers with 99.1% calling concordance). The *P*-value, which is the principal measure of marker quality, was positively correlated with marker call rate ($r = 0.45$). Calling discordance was negatively correlated both with both *P* ($r = -0.31$) and call rate ($r = -0.23$).

A similar analysis based on bins of 100 markers grouped in descending order of their *P*-value revealed the same trend. Although average calling discordance increased as the average *P*-value of a marker group decreased, even low-quality markers ($P < 75\%$) had an average discordance of only 0.37% (Fig. 1, top panel). The two groups of top-quality markers ($P > 92\%$) had extremely low average discordance (0.02% or four discordant calls among almost 20,000 comparisons). As expected, the decrease in *P* for a group of markers was accompanied by a decreasing call rate, from more than 98% to just above 93% (Fig. 1, bottom panel). Overall, for the markers meeting our 'standard quality threshold' ($P > 77\%$) the average scoring concordance was 99.8% and the average call rate was 95%.

Genetic relationship between wheat cultivars revealed by DArT

Having established that the genotyping array performed well from a technical point of view, we tested its ability to resolve genetic relationships among a

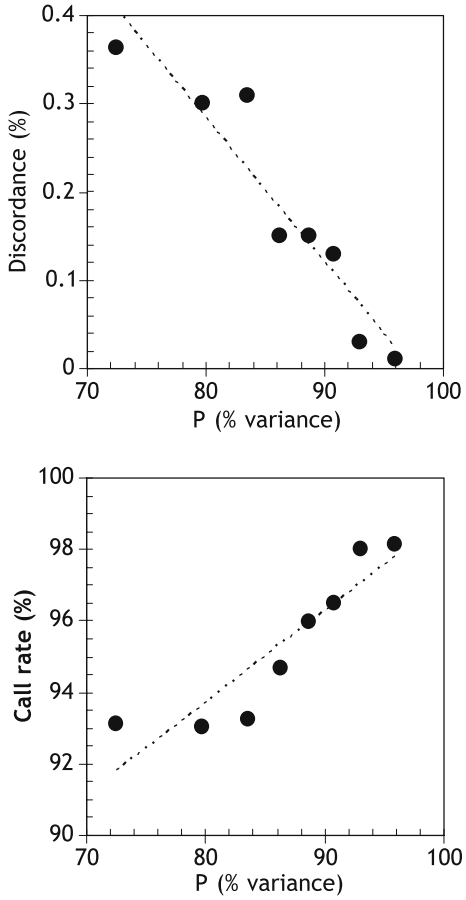

Fig. 1 Relationships among different quality parameters for DArT markers. Markers were distributed into bins of 100 markers each according to descending *P*-value. Within-group averages for *P* (*x*-axes) were plotted against average values for allele-calling discordance (*y*-axis of *top panel*) and call rate (*y*-axis of *bottom panel*)

group of 62 hexaploid wheat lines (including three potential replicates—alternative seed samples from three cultivars). The wheat lines examined were mainly winter wheats bred in Europe and spring wheats bred in Australia (Supplementary Table 2). We identified 411 polymorphic markers with $P > 77\%$, call rate > 85 and 100% calling consistency (Supplementary Table 3).

Figure 2 displays the position of each sample in the space spanned by the first two principal coordinates of a relative Hamming distance matrix derived from the scores, which jointly explained 50.6% of the total data variance. There was a clear separation between European wheat lines and the Australian materials, with the Australian genotypes being significantly more diverse than the European wheats. The genetic relationships are analysed in more detail in the Discussion section.

Genetic mapping of DArT markers

To confirm that wheat DArT markers behave in a Mendelian fashion, we constructed a linkage map for a cross between cultivars Cranbrook and Halberd (Kammholz et al. 2001). We selected 339 RFLP, SSR and AFLP markers from a curated/expanded FW map derived from the original map (Chalmers et al. 2001; Lehmensiek et al. 2005), added the data of 71 STM markers typed during this study, and merged the combined data set with the segregation data for the top 339 DArT markers (Supplementary Table 4).

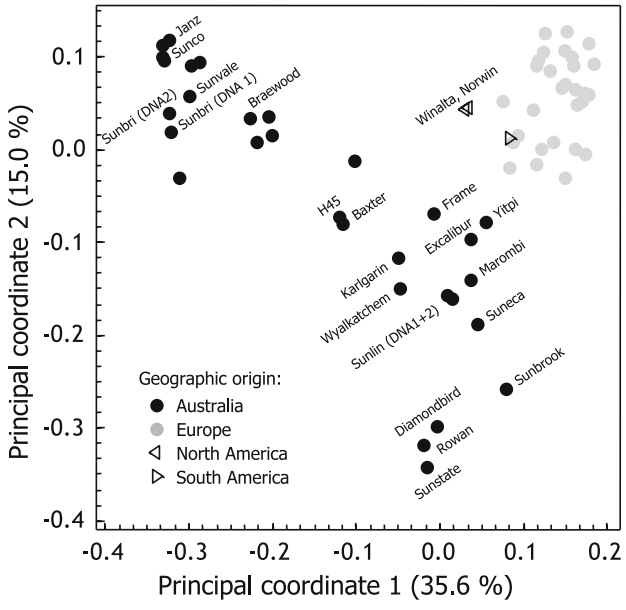

Fig. 2 Principal Coordinate analysis of 62 wheat cultivars based on 411 DArT markers. The names of cultivars mentioned in the text are inserted in the figure. The principal coordinates of all cultivars as well as information on geographic origin, growth habitat and pedigree are provided in Supplementary Table 2

Map length and coverage

The linkage map derived from the composite data set spanned 2,937 cM. The 339 FW markers alone spanned 2,534 cM (86% of total length). An identical number of DArT markers spanned 2,383 cM, 81% of the total length (Table 2). DArT markers mapped to all 21 chromosomes with the exception of chromosome 4D, which was only represented by a small linkage group (21.3 cM) with one STM and four FW markers (Fig. 3). It is possible that the DArT array contained 4D marker(s), which were removed in the course of map construction because of a lack of linked anchor markers. DArT markers, however, formed two small linkage groups spanning approximately 30 cM in total, which did not contain any FW markers. Preliminary mapping data from seven additional mapping populations indicate that group 1 belongs to chromosome 6B (Neil Howes, unpublished observation).

Segregation distortion

There was no statistically significant difference between DArT and non-DArT markers in the distribution of parental alleles. Only in two areas of the map was segregation distortion significant at the $P < 0.01$ level (chromosome regions 4AL and 1DS). These two regions contained both DArT and non-DArT markers.

Marker distribution among chromosomes and genomes

Diversity arrays technology markers were distributed among chromosomes in a similar way as FW markers, suggesting that the density of both groups of markers roughly followed the distribution of DNA polymorphism across the genome (Fig. 4). There was, however, a statistically significant deficit of DArT markers on the D genome, which contained approximately twice as many FW markers as DArT markers (Supplementary Table 5). This deficit of DArT markers is consistent with the well established low level of molecular marker (Bryan et al. 1997) and DNA sequence (Caldwell et al. 2004) variation in the D genome, a phenomenon, which has triggered the targeted development of SSR markers for the D genome (Pestsova et al. 2000). There was no significant difference in the distribution of the two marker types among the seven homologous chromosome groups (Supplementary Table 5).

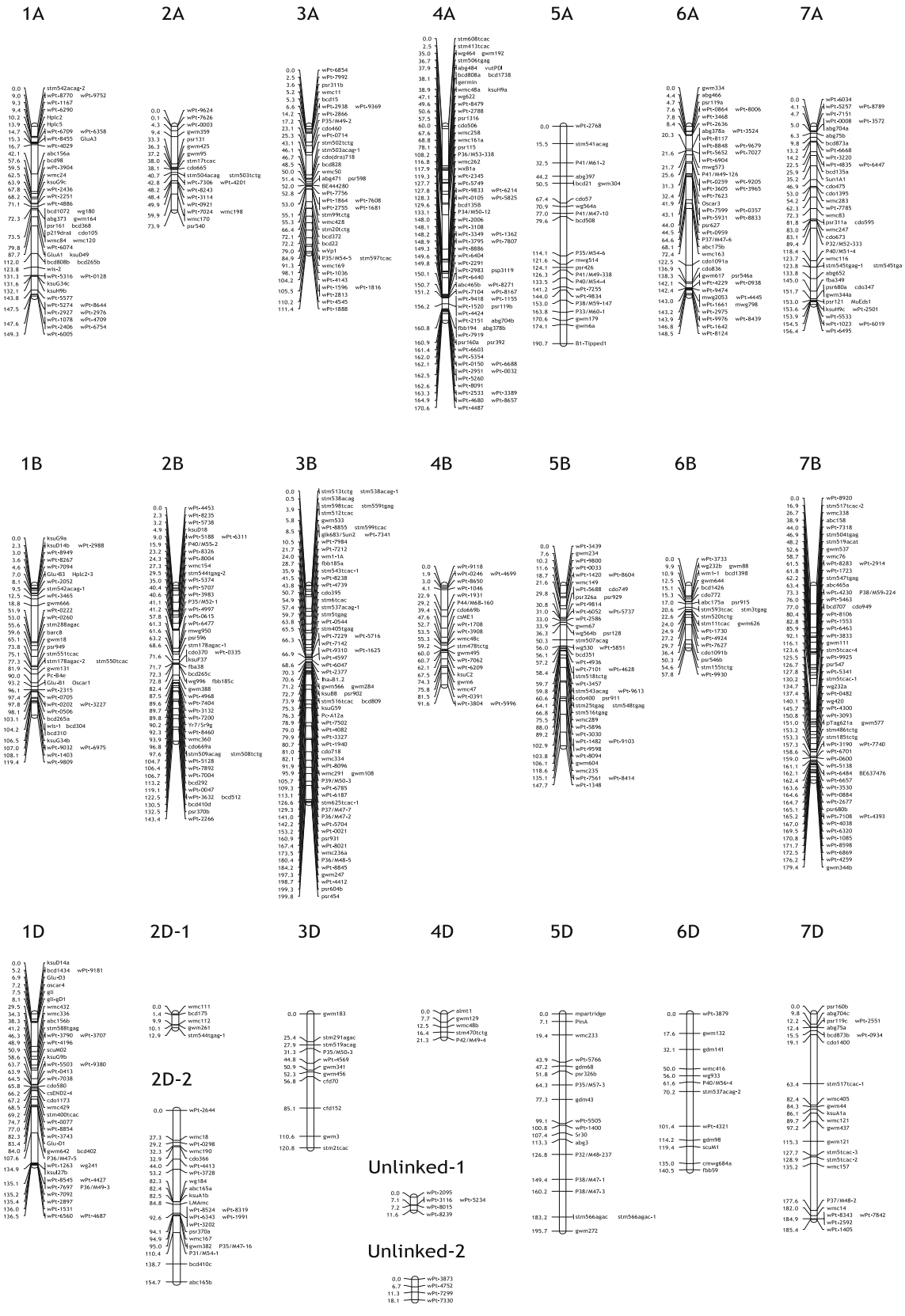

Fig. 3 Genetic map for a cross between wheat cultivars Cranbrook and Halberd. The segregation data of DArT and STM markers used to construct the map are available as Supplementary Table 4

Table 2 Mapping quality features of different sets of markers

Marker type	Number	Call rate (%)[a]	LOD (mean ± SD)[b]	Unique segregation patterns (%)[c]	Coverage (cM)[d]
FW[e]	339	93 ± 7.4	17 ± 5.3	80	2,534
DArT	339	94 ± 4.2	20 ± 4.4	69	2,383
STM	71	92 ± 5.9	16 ± 4.2	86	–[f]
All	749	93 ± 5.8	18 ± 5.0	71	2,937

[a] Call rates were computed after removing lines that were not scored at all for a particular marker type (STM and some FW markers)

[b] Dataset mean of the averages of the LOD score pairs linking each marker with its two neighbours

[c] Percentage of markers with unique segregation patterns (i.e. number of unique loci as a percentage of the number of markers)

[d] Coverage was computed by adding, for FW and DArT markers separately, the distances between the two most distal markers on each of the chromosomes. Because of insufficient numbers of markers, chromosomes 3D and 4D were not included when calculating coverage of DArT markers

[e] Quality-filtered and curated SSR, RFLP and AFLP loci

[f] The STM data set was too small to derive a meaningful value for genome coverage

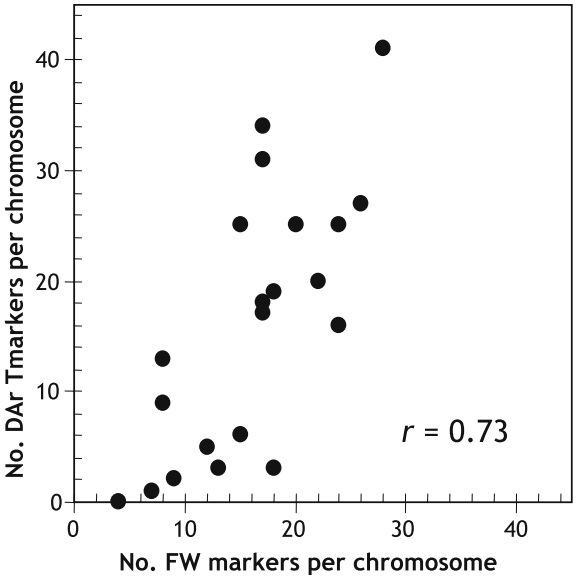

Fig. 4 Relationship between the number of DArT markers and the number of other markers across the 21 chromosomes

Centromeric clustering

We selected the ten chromosomes with the largest numbers of markers (1A, 3A, 4A, 6A, 7A, 2B, 3B, 5B, 7B, 1D) to compare the degree of centromeric clustering between DArT and non-DArT markers by calculating for each dataset and chromosome the percentage of markers located in 'centromeric regions' (the central one-third of the genetic length of a chromosome). On average there were fewer DArT markers (24 ± 17%) than non-DArT markers (39 ± 16%) in centromeric regions. This difference was significant at the $P < 0.03$ level as deduced from a one-tailed *t*-test based on the expectation that *Pst*I-based DArT markers should show a bias towards non-centromeric, hypomethylated regions (Peter Wenzl et al. submitted).

Uniqueness of segregation patterns

Not surprisingly, more FW markers (80%) than DArT markers (69%) had unique segregation patterns (Table 2). A large number of markers with identical segregation patterns (mostly AFLP) were removed in the process of map curation that produced the FW map used in this study (Lehmensiek et al. 2005). The markers on the DArT array used for this study, by contrast, were not filtered for redundancy. This difference in the treatment of co-segregating loci probably has also contributed to the slightly higher average LOD score for DArT markers (20 ± 4.4) compared to FW markers (17 ± 5.3) (Table 2). A second contributing factor may have been the higher call rates of DArT markers compared to FW markers. The group of 71 STM markers, only about a fifth in size compared to the DArT and FW groups of markers, had more unique segregation patterns (Table 2). Interestingly, the average percentage of unique patterns in twelve groups of 71 randomly selected DArT markers was higher than for the STM markers (92.6 ± 3.9%), which indicates a low redundancy level for DArT markers.

Discussion

DArT performs well in the hexaploid wheat genome

Polyploidy may affect the precision of molecular marker technologies in different ways. PCR-based marker technologies such as SSR or SNP may be

affected by alternative primer binding sites on homeologous chromosomes both 'diluting' the correct annealing targets and competing for primers. Hybridisation-based marker technologies such as RFLP may suffer from the presence of multiple targets simultaneously hybridising to the same probe. At the onset of this study we considered that the hexaploid nature of the wheat genome could interfere with DArT genotyping. In particular, we were concerned that polymorphism frequency and technical reproducibility would be low as a result of homologous DNA fragments cross-hybridising to each other. We therefore evaluated the polymorphism frequency obtained with a variety of complexity reduction methods (Table 1) and evaluated the technical performance of an array built with the best method identified (*Pst*I/*Taq*I; see section entitled "Test of *Pst*I/*Taq*I array performance" in Results).

Polymorphism frequency

The four methods of complexity reduction tested in this study have previously been applied to barley (Wenzl et al. 2004), a diploid species with a similar genome size and structure as the component genomes of wheat (Feuillet and Keller 2002). Barley is believed to have a higher level of molecular marker polymorphism than wheat (Langridge et al. 2001). We found the polymorphism frequencies of the four representations in wheat to be very similar to those of barley. The frequencies ranged from 4.2 (wheat) to 4.0% (barley) for the *Pst*I/*Mse*I representation to 9.4 (wheat) and 10.4% (barley) for the *Pst*I/*Taq*I representation, with identical ranking of representations in the two species. We can therefore conclude that the three-fold increase in ploidy level does not negatively affect the ability of DArT to detect DNA polymorphism in cereals with large genomes.

This positive result can be attributed to two key features of DArT. First, the specificity of polymorphism detection does not rely on the annealing of primers to genomic targets in the presence of homologous annealing targets, but is mediated by high-fidelity restriction enzymes detecting SNPs in restriction enzyme sites (Wittenberg et al. 2005). Second, only a tiny fraction of adapter-ligated digestion products is amplified and captured in the genomic representation (approximately 10,000–20,000 fragments representing 0.1–0.2% of the genome). This selection step reduces the chance that homologous fragments are amplified together and cross-hybridise during the assay. Polymorphic fragments (DArT markers) are probably sampled from just a single homologous chromosome. Ongoing mapping efforts currently suggest that only approximately 2% of DArT markers map to more than one location in the genome (unpublished results). This number is consistent with the explanation above and only marginally higher than the frequency of multi-locus markers in diploid barley (Peter Wenzl et al. submitted).

Technical performance

The technical reproducibility of molecular marker assays is an issue that has only rarely been addressed in the scientific literature, although it is known that genotyping error rates can be relatively high and variable, and that errors may impact on the biological inferences from the data (Bonin et al. 2004). We could not find any study on the relationship between the ploidy level and genotyping error frequency. In wheat, the reproducibility of SSR markers ranged from 89.5% for genomic SSRs to 98.8% for EST-derived SSRs (Dreisigacker et al. 2004). We did not observe any differences in the reproducibility of DArT assays between barley (Wenzl et al. 2004) and wheat (this study): the average frequency of discordant scores was approximately 0.2% for both genomes. Similarly high levels of reproducibility were also observed for the small diploid genome of Arabidopsis and the medium-size diploid genomes of cassava and pigeonpea (Wittenberg et al. 2005; Xia et al. 2005; Yang et al. 2006).

The virtually constant genotyping error rate for genomes with a threefold difference in ploidy level and an up to 150-fold variation in size, can largely be attributed to the way DArT markers are scored. The '*P*'-value used by DArTsoft as a measure of marker quality represents the proportion of the relative target signal variance that is due to variation between allelic states. Both the *P*-value and the call rate are correlated with scoring reproducibility (see Results). The application of similar *P* and call rate thresholds in different experiments and genomes, therefore, ensures a similarly low level of scoring discordance. Relaxing these thresholds results in the identification of additional markers, but these markers have an increased rate of scoring inconsistencies.

DArT fingerprints reflect genetic relationships

Principal coordinate analysis revealed an interesting pattern of genetic relationships among the materials studied with the DArT array. The European wheat lines (mostly from western Europe) clustered together and were close to cultivars Norwin and Winalta, two Canadian hard red winter wheats. This is consistent with the low level of diversity in the modern western European germplasm reported by Roussel et al.

(2005). The Australian spring wheat lines were more diverse and separated into two groups. The first group comprised cultivars of the 'Condor' family based on the CIMMYT introduction WW15: Janz, Sunbri, Sunco, Sunvale and Braewood. These cultivars have one or more of the following alien translocations responsible for improved rust resistance: VPM-1 (*Sr38/Lr37/Yr17*) from *Triticum ventricosum*, *Sr26* from *Thinopyrum intermedium*, *Sr36* from *Triticum timopheevii* or *Sr24/Lr24* from *Thinopyrum ponticum* (Harbans Bariana, personal communication; McIntosh et al. 1995). The second group was more diverse and had a broader range of adaptation. Cultivars Frame, Yitpi, Excalibur, Kalgarin and Wyalkatchem are in the "Spear" family and are generally adapted to the drier regions of southern and western Australia. They have different sets of rust resistance genes (Harbans Bariana, personal communication). Close to this group was the 'Hartog/Suneca' family including cultivars Sunlin, Suneca, Sunstate, Diamondbird, Sunbrook, Marombi, Rowan and H45, which are based on the CIMMYT lines Pavon S, Ciano 67 and Sonora 64. These lines often have rust genes *Sr2*, *Sr9g*, *Sr30* and some also have the alien translocations *Sr26* or VPM-1, but not *Sr36* (Harbans Bariana, personal communication; McIntosh et al. 1995). Cultivar Baxter with *Sr2*, *Sr30* and *Sr36* was an exception.

Although there are only a few cultivars in common between this study and the RFLP-based diversity study of Australian material by Paull et al. (1998), both studies detected the major families in Australian wheat germplasm and a fairly high level of diversity. A more recent RFLP/SSR study of the same sample of Australian germplasm (Parker et al. 2002) showed a similar picture. The genetic distance matrices for the two marker types were correlated, but the 19 SSR loci (with 160 scorable bands) failed to improve the significance of cultivar groupings obtained with 90 RFLP probes × five restriction enzymes. The authors concluded that a larger number of SSR loci would be required to determine robust genetic relationships among a large number of accessions. Since the accuracy of genetic distance measurements depends on the number of markers and their distribution in the genome (Schut et al. 1997) one would expect the DArT-based diversity pattern to be at least as precise as the one derived from the combined RFLP and SSR data.

Two of three cultivars represented by two DNA samples extracted from different plants showed variation for one (Sunbri) to six (Sunco) of the 411 DArT markers. We interpret these differences as intravarietal heterogeneity. We have observed similar levels of heterogeneity (approximately 1%) in barley cultivars analysed with DArT (Wenzl et al. 2004). The high resolution of DArT arrays enables applications in seed purity and genetic ID testing, including the discrimination among closely related cultivars.

Genetic mapping

Our ability to build an integrated map comprising both DArT and FW markers has demonstrated that DArT markers behave in a Mendelian fashion and can be scored as single-locus markers despite the hexaploid nature of wheat. All measures of marker quality used in this study point to a good performance of DArT arrays for genetic mapping applications. The information content of the DArT dataset was comparable to that of the combined SSR/RFLP/AFLP dataset of the FW map. DArT markers showed somewhat less clustering around centromeres when compared to FW markers, even after the removal of (mostly AFLP) markers with redundant segregation patterns during the FW map curation process (Lehmensiek et al. 2005). It seems that DArT markers have a stronger tendency than SSR and AFLP markers in particular, to map to gene-rich telomeric regions (Vuylsteke et al. 1999). We have observed a similar tendency in a barley consensus map with approximately 2,000 DArT and 1,000 SSR, RFLP and STS markers, despite some degree of clustering around centromeres observed with all types of marker as a result of centromeric suppression of recombination (Tanksley et al. 1992; Peter Wenzl et al. submitted).

The resolution of the DArT map of wheat was not as high as the resolution of the barley DArT map reported by Wenzl et al. (2004). This was mainly due to the hexaploid nature of wheat, which translates to the wheat genetic map being roughly three times as long as the map for barley. An initial survey of DArT markers in seven additional crosses has resulted in the tentatively assignment of more than 1,100 markers to the 21 chromosomes. This analysis has also shown that true marker redundancy (due to multiple copies of the same clone on the array rather than closely linked loci) may be as low as 10% (Neil Howes, unpublished observations). The redundancy of the current wheat array is therefore substantially lower than for the barley arrays used by Wenzl et al. (2004). Assuming that the frequency distribution of polymorphic markers in libraries prepared from genomic representations follows a Poisson distribution, we extrapolate that it may be feasible to develop a *Pst*I/*Taq*I array with 2,000–4,000 unique DArT markers polymorphic for a broad spectrum of wheat cultivars (Cohen 1960). Such an array

would provide unprecedented genome coverage for the routine genotyping of hexaploid wheat.

Conclusions

The results of this study demonstrate that DArT can be effectively deployed to genotype polyploid species with large genomes such as wheat. The data quality for wheat was similar to the quality of DArT data previously generated for barley and several other species. A single DArT assay, which takes a maximum of three working days to complete from DNA to data, generates a reproducible medium-density scan of the hexaploid wheat genome that is useful for a range of molecular breeding and genomics applications.

Acknowledgments We thank the Australian Grains Research and Development Cooperation (GRDC; www.grdc.com.au) for financial support. Triticarte P/L (www.triticarte.com) is a joint venture of Diversity Arrays Technology P/L (DArT P/L; www.DiversityArrays.com) and the Value Added Wheat Cooperative Research Centre (VAWCRC; www.wheat-research.com.au). The Triticarte/DArT team thank their colleagues at CAMBIA (www.cambia.org) for their friendship and collaborative spirit as well as many interesting discussions during the period when CAMBIA was sharing laboratory facilities with Triticarte/DArT. We also thank two anonymous reviewers for their detailed comments, which have helped to improve the manuscript.

References

Anderson MJ (2003) PCO: a FORTRAN computer program for principal component analysis. Department of Statistics, University of Auckland, Auckland, New Zealand

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. Mol Ecol 13:3261–3273

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

Bryan GJ, Collins AJ, Stephenson P, Orry A, Smith JB, Gale MD (1997) Isolation and characterisation of microsatellites from hexaploid bread wheat. Theor Appl Genet 94:557–563

Caldwell KS, Dvorak J, Lagudah ES, Akhunov E, Luo MC, Wolters P, Powell W (2004) Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. Genetics 167:941–947

Chalmers KJ, Campbell AW, Kretschmer J, Karakousis A, Henschke PH, Pierens S, Harker N, Pallotta M, Cornish GB, Shariflou MR, Rampling LR, McLauchlan A, Daggard G, Sharp PJ, Holton TA, Sutherland MW, Appels R, Langridge P (2001) Construction of three linkage maps in bread wheat. *Triticum aestivum*. Aust J Agric Res 52:1089–1119

Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SPA (1996) Accessing genetic information with high-density DNA arrays. Science 274:610–614

Cohen AC Jr (1960) Estimating the parameter in a conditional Poisson distribution. Biometrics 16:203–211

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:1–15

Dreisigacker S, Zhang P, Warburton ML, Van Ginkel M, Hoisington D, Bohn M, Melchinger AE (2004) SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different megaenvironments. Crop Sci 44:381–388

Feuillet C, Keller B (2002) Comparative genomics in the grass family: molecular characterisation of grass genome structure and evolution. Ann Bot (Lond) 89:3–10

Hayden MJ, Good G, Sharp PJ (2002) Sequence tagged microsatellite profiling (STMP): improved isolation of DNA sequence flanking target SSRs. Nucleic Acids Res 30:e129

Hayden MJ, Stephenson P, Logojan AM, Khatkar D, Rogers C, Koebner RMD, Snape JW, Sharp PJ (2004) A new approach to extending the wheat marker pool by anchored PCR amplification of compound SSRs. Theor Appl Genet 108:733–742

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

Kammholz SJ, Campbell AW, Sutherland MW, Hollamby GJ, Martin PJ, Eastwood RF, Barclay I, Wilson RE, Brennan PS, Sheppard JA (2001) Establishment and characterisation of wheat genetic mapping populations. Aust J Agric Res 52:1079–1088

Kilian A, Huttner E, Wenzl P, Jaccoud D, Carling J, Caig V, Evers M, Heller-Uszynska, Cayla C, Patarapuwadol S, Xia L, Yang S, Thomson B (2005) The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In: Tuberosa R, Phillips RL, Gale M (eds) Proceedings of the international congress "In the wake of the double helix: from the green revolution to the gene revolution", 27–31 May, 2003, Avenue Media, Bologna, Italy, pp 443–461

Lalouel JM (1977) Linkage mapping from pair-wise recombination data. Heredity 38:61–77

Langridge P, Lagudah ES, Holton TA, Appels R, Sharp PJ, Chalmers KJ (2001) Trends in genetic and genome analysis in wheat: a review. Aust J Agric Res 52:1043–1077

Lehmensiek A, Eckermann PJ, Verbyla AP, Appels R, Sutherland MW, Daggard GE (2005) Curation of wheat maps to improve map accuracy and QTL detection. Aust J Agric Res 56:1347–1354

Lezar S, Myburg AA, Berger DK, Wingfield MJ, Wingfield BD (2004) Development and assessment of microarray-based DNA fingerprinting in *Eucalyptus grandis*. Theor Appl Genet 109:1329–1336

Liu BH, Knapp SJ (1990) GMENDEL: a program for Mendelian segregation and linkage analysis of individual or multiple progeny populations using log-likelihood ratios. J Hered 81:407–418

McIntosh RA, Wellings CR, Park RF (1995) Wheat rusts: an atlas of resistance genes. CSIRO, Melbourne, Vic., Australia

Mochida K, Yamazaki Y, Ogihara Y (2004) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Mol Genet Genomics 270:371–377

van Ooijen JW, Voorrips RE (2001) Joinmap 3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands

van Os H, Stam P, Visser RGF, van Eck HJ (2005) RECORD: a novel method for ordering loci on a genetic linkage map. Theor Appl Genet 112:30–40

Parker GD, Fox PN, Langridge P, Chalmers K, Whan B, Ganter PF (2002) Genetic diversity within Australian wheat breeding

programs based on molecular and pedigree data. Euphytica 124:293–306
Paull JG, Chalmers KJ, Karakousis A, Kretschmer JM, Manning S, Langridge P (1998) Genetic diversity in Australian wheat varieties and breeding material based on RFLP data. Theor Appl Genet 96:435–446
Pestsova E, Ganal MW, Röder MS (2000) Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 43:689–697
Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F (2005) SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor Appl Genet 111:162–170
Schut JW, Qi X, Stam P (1997) Association between relationship measures based on AFLP markers, pedigree data and morphological traits in barley. Theor Appl Genet 95:1161–1168
Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739–744
Tanksley SD, Ganal MW, Prince JP, de Vicente MC (1992) High density molecular linkage maps of the tomato and potato genomes: biological inferences and practical applications. Genetics 132:1141–1160
Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78
Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Fritjers A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414
Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lübberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921–935
Weber J, May P (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396
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 101:9915–9920
Williams J, Kubelik A, Livak K, Rafalski J, Tingey S (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535
Wittenberg AHJ, van der Lee T, Cayla C, Kilian A, Visser RGF, Schouten HJ (2005) Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. Mol Genet Genomics 274:30–39
Xia L, Peng K, Yang S, Wenzl P, de Vicente C, Fregene M, Kilian A (2005) DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. Theor Appl Genet 110:1092–1098
Yang S, Pang W, Harper J, Carling J, Wenzl P, Huttner E, Zong X, Kilian A (2006) Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology (DArT). Theor Appl Genet 113:585–595