Mapping translocation breakpoints using a wheat microarray

Prasanna R. Bhat1, Adam Lukaszewski1, Xinping Cui2, Jin Xu2,3, Jan T. Svensson1, Steve Wanamaker1, J. Giles Waines1 and Timothy J. Close1,*

1Department of Botany and Plant Sciences, University of California, Riverside, California, USA 92521-0124, 2Department of Statistics, University of California, Riverside, California, USA 92521-0124 and 3Department of Statistics, East China Normal University, Shanghai, China, 200062

Received December 7, 2006; Revised and Accepted February 23, 2007

ABSTRACT

We report mapping of translocation breakpoints using a microarray. We used complex RNA to compare normal hexaploid wheat (17 000 Mb genome) to a ditelosomic stock missing the short arm of chromosome 1B (1BS) and wheat-rye translocations that replace portions of 1BS with rye 1RS. Transcripts detected by a probe set can come from all three Triticeae genomes in ABD hexaploid wheat, and sequences of homoeologous genes on 1AS, 1BS and 1DS often differ from each other. Absence or replacement of 1BS therefore must sometimes result in patterns within a probe set that deviate from hexaploid wheat. We termed these ‘high variance probe sets’ (HVPs) and examined the extent to which HVPs associated with 1BS aneuploidy are related to rice genes on syntenic rice chromosome 5 short arm (5S). We observed an enrichment of such probe sets to 15–20% of all HVPs, while 1BS represents ~2% of the total genome. In total 257 HVPs constitute wheat 1BS markers. Two wheat-rye translocations subdivided 1BS HVPs into three groups, allocating translocation breakpoints to narrow intervals defined by rice 5S coordinates. This approach could be extended to the entire wheat genome or any organism with suitable aneuploid or translocation stocks.

INTRODUCTION

Interest in translocations and other chromosomal rearrangements stems in part from their association with novel phenotypes, or their usefulness in genetic mapping (1,2). A number of human diseases arise from rearrangements involving unstable regions of DNA (3). Several methods have been tried to map genomic rearrangements. Comparative genomic hybridization (CGH) to microarrays detects numeral (copy number) changes but provides no information on the structure of abnormalities (4), while multicolor spectral karyotyping detects both structural and numeral aberrations but has low resolution (5). End-sequence profiling (ESP) is well suited to the detection of genomic rearrangements and complements existing cytogenetic techniques by making use of a reference genome sequence to identify BAC clones whose ends span the junctions of rearranged positions (6). The speed of translocation breakpoint mapping also has been improved by array painting (7), a method in which chromosomes are flow-sorted and competitively hybridized onto a large insert clone array to reveal a shift in relative hybridization intensity ratios, thus mapping the breakpoint at a resolution defined by size of the clones on the array. Array painting and PCR to amplify the junctions have facilitated the sequencing of translocation breakpoints in humans (7). More recently a single nucleotide polymorphism (SNP) based CGH approach for the high-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping was described for humans (8).

Translocation breakpoint mapping in plants lags far behind that in humans. In situ hybridization with total genomic DNA (GISH) has been useful for the characterization of alien introgressions in plants (9). Translocation breakpoints have been used as physical landmarks in cytologically integrated molecular linkage maps of barley chromosomes (10). With recent simplifications, GISH can now be used in large-scale screening for new chromosome constructs. The level of resolution of GISH has been tested in many plant species, for example recombined wheat-rye chromosomes with genetically determined positions of the translocation breakpoints. Depending on the detection method used, the limits of resolution were at least 9.8 and 3.5 cM, and depending on the translocation configuration and the detection method used, some distally located breakpoints could not be visualized at all. These resolution limits of GISH may hamper isolation of critical translocation breakpoints in...
some chromosome engineering efforts (11). The use of microarrays in plants that lack a complete genome sequence provides an alternative tactic to improve the resolution of translocation breakpoint mapping.

Wheat is one of the most important agricultural crops globally and has a 1x genome size of ~5700 Mb. Hexaploid wheat (6x) has seven homoeologous groups of chromosomes, each composed of a representative of the A, B and D genomes (12) for a total (3x) genome size of 17 000 Mb. The availability of aneuploids and translocation lines makes wheat an attractive system for mapping using methods other than those that depend on meiotic recombination frequencies among populations segregating marker alleles. The group 1 homoeologous chromosomes have been the most studied, primarily because these chromosomes have several agronomically important genes. This includes 22 genes or QTLs conferring resistance to stripe rust in wild emmer wheat (13,14). In a single gene-rich region ISO.8 of the short arm of chromosome 1, 41 RFLP markers were identified (15). One tally of markers for wheat group 1 chromosomes placed the total at ~1500, mainly RFLPs and SSRs (16). In a more recent effort, 2212 EST loci were assigned to group 1 chromosomes including 326 associated with 1BS (17) using wheat nullisomic-tetrasomic, ditelesomic and deletion lines.

Here, we report the use of a novel alternative method to develop markers of genes present in a single chromosome arm of hexaploid wheat and mapping of translocation breakpoints using a wheat genome array. We used RNA from aneuploid stocks, a commercially available wheat genome array, and a statistical method that took into consideration the polyploid nature of hexaploid wheat.

MATERIALS AND METHODS

Plant materials

The set of lines used in this study was developed in hexaploid wheat cultivar ‘Pavon 76’, here referred to as ‘Pavon’. Ditelesomic lines for the group-1 chromosomes were produced by centric misdivision of univalents. Centric translocation 1RS.1BL was transferred into Pavon from cv. Genaro, obtained from the International Center for the Improvement of Maize and Wheat (CIMMYT), Mexico. This translocation originated from the cv. Kavkaz (18–20). It was combined with the ph1b mutation and a series of homoeologous recombinants 1BS-1RS were produced (21). For this experiment, two recombinants were selected namely (i) 1B+5 has a 1B chromosome of Pavon with a terminal segment of the short arm replaced by a corresponding segment of 1RS. (ii) T – 1 is a 1RS.1BL translocation with the terminal segment of 1RS replaced by a corresponding segment of 1BS, as illustrated in Figure 1.

RNA extraction, labeling, hybridization to wheat genome array

Seed germination was done in sterile glass dishes. Seedlings were harvested in a well-hydrated state to provide unstressed tissue. Additional seedlings were exposed to 90% relative humidity in desiccator jars for two days to provide stressed tissue. Total RNA was initially isolated from frozen whole seedlings using TRIzol (Gibco BRL Life Technologies, Rockville, MD, USA) reagent. The RNA was further purified using an RNasea spin column (Qiagen, Chatsworth, CA, USA) and an on-column DNase treatment. For each genotype, a 50:50 mixture of stressed and unstressed whole seedling RNA was used in each of two replicate samples. The purpose of combining unstressed and stressed RNA from whole seedlings was to maximize the complexity of the RNA population in order to utilize RNA as a surrogate for genomic DNA, as described earlier (22,23). RNA integrity was assessed prior to target preparation using RNA Lab-On-A-Chip (Caliper Technologies Corp., Mountain View, CA, USA) evaluated on an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA).

Two separate RNA samples per genotype were used to make biotin-tagged cRNAs which were then hybridized to the Affymetrix wheat genome array (a total of 10 chips for five genotypes). Labeling, hybridization, washing and scanning methods were performed as recommended by Affymetrix (Affymetrix GeneChip Expression Analysis Technical Manual; Affymetrix, Inc., Santa Clara, CA, USA) by Sriti Misra at the DNA and Protein Microarray Facility at the University of California, Irvine. The chip images were scanned for visible defects and then extracted using default settings and tabulated as CEL files using Affymetrix GeneChip Operating Software (GCOS 1.2). We used a global scaling factor of 500, a normalization value of 1 (i.e. no normalization procedure is applied to the data), and default parameter settings for the wheat

![Diagram of wheat chromosomes](https://example.com/diagram.png)

**Figure 1. Translocation stocks used in this study.** Wheat chromosomes are colored red and introgressed rye segments are green from fluorescence in situ hybridization with a probe of the total genomic DNA of rye. Long arm 1BL is present in each case. (A) normal hexaploid Pavon (with normal1BS), (B) Pavon carrying a translocated chromosome in which the entire short arm of rye 1R has replaced the entire short arm of 1B (1RS.1BL), (C) Pavon carrying a translocated chromosome 1B with a proximal segment of rye 1R replacing a proximal region of 1BS (T–1), (D) Pavon carrying a translocated chromosome 1B with a terminal segment of rye 1RS replacing a terminal segment of 1BS (1B+5), and (E) Pavon carrying a ditelesomic chromosome 1B missing the short arm (Dt1BL).
more advanced nonlinear normalization methods, such as quantile normalization, which may smooth out the probe level differentiation between two lines. Such differentiation is the primary target of our detection method.

First, we claim that for genes that are present equivalently in both of two genotypes the signal intensities of the probes from the corresponding probe sets shall display a parallel pattern between the two genotypes (see Figure 2A and B), while genes that are not present equivalently shall display two non-parallel patterns. These non-parallel patterns then constitute high variation in probe intensity differences between the two genotypes (see Figure 2C and D). In our case, a probe set for a gene that exists in three homoeologous versions on the A, B and D genomes will give a characteristic pattern of probe intensities that is different from the pattern of probe intensities that would be produced if the B genome contribution is missing. This leads us to use the following algorithm.

(S1) Use only probe sets with ‘present’ call by GCOS in both replicates of Pavon (complete ABD) samples. This step filters out probe sets with low quality values.

(S2) Let \( S_{ij} \) be the log (base 2) scaled observed signal intensity of the \( i \)th probe in the \( i \)th probe set of genotype \( t \). (e.g. \( t=1,2 \), representing Pavon and Pavon (Dt1BL), respectively.)

(S3) Let \( \bar{S}_{ij} \) denote the sample mean of \( S_{ij} \) taken over duplicate samples.

(S4) For every probe set \( i \), let the probe difference between two lines be \( d_{ij} = \bar{S}_{ij} - \bar{S}_{2ij} \) and compute the within probe set variances, denoted by \( \nu_i \).

(S5) The candidate probe sets are selected from the right tail of the distribution of \( \nu \) which corresponds to high within probe sets variance. The actual cutoff may be chosen subjectively by the user based on the distribution of \( \nu \). In the case of 1BS, it was not difficult to find the value (0.2) with largest descendant rate from the distribution of \( \nu \) (see Supplementary Figure 1).

RESULTS

The ‘high variance probe set’ hypothesis

Each probe set on an Affymetrix wheat genome array is composed of 11 pairs of 25-mer oligonucleotides, each pair including one perfect match 25-mer and one center-nucleotide-mismatch to the target gene sequence. In the present work, we used only the perfect match probe data (see Materials and methods section), i.e. 11 data points per probe set. Because of the high similarity of genes on homoeologous chromosomes in hexaploid wheat, a given probe set must often measure intensities from all three genomes (A, B and D). But given also that homoeologous genes usually do not have identical sequences, and that 25-mer hybridization is affected by mismatches, then it follows that the relative signal intensities of each individual probe in a probe set must not always be
identical in their contribution to the aggregate signal. Our intention was to exploit this non-identical contribution of individual homoeologous genes to the aggregate signal to assign probe sets to individual chromosome arms using aneuploid stocks of wheat. For this purpose, if we compare the signals that come from all three genomes in normal hexaploid wheat with the signals from an aneuploid missing one chromosome arm, then the differences in signal intensity, when considering each probe in the probe set, will have generally higher variance when a probe set represents a gene that is absent from the aneuploid (resides on the missing arm) than when it is present in the aneuploid. In contrast, when the probe set variance is low, we cannot infer that the gene is located on the missing arm; this could be the result of genes located elsewhere in the genome.

The net consequence of the above consideration is that probe sets which mark genes on the missing chromosome arm should be enriched among ‘high variance probe sets’ (HVPs). Similarly, we can expect that if a segment of a wheat chromosome arm is replaced by a corresponding segment of a rye homoeologue, then different probe set signal patterns will be observed for the genes involved. Furthermore, it follows that the HVPs, when consistent across a series of recombinant lines, can provide genetic markers that delimit the position of translocation breakpoints. Screening of a set of 1BS-1RS recombinant lines was therefore expected to identify genes present in each bin defined by translocation breakpoints.

Compilation of HVPs

The data from five genotypes were used to compile a list of HVPs by comparing the genotypes in nine of ten possible pairwise combinations (Supplementary Table 1). In order to find an appropriate statistical cutoff point for HVPs (see Materials and methods section), the density of HVPs related to rice gene models on the short arm of rice chromosome 5 was examined as a function of the variance cutoff. The Pavon versus Pavon Dt1BL comparison provided the most extreme case of genetic difference (presence versus absence of the complete 1BS arm) and is shown in Figure 3. The highest variance cutoff level (1.0), ~32% of HVPs matched rice chromosome 5 genes above a fairly constant background level of 5–10% for each of the remaining 11 rice chromosomes. As the variance cutoff point was reduced, the percentage of rice chromosome 5 genes declined accordingly, indicating that the higher the variance the greater the confidence that HVPs truly match rice chromosome 5 genes. We took...
432 distinct rice gene models. Among these, 68 gene models (15.7% of the total) were clustered on the terminal portion of the short arm of rice chromosome 5 (5S), which is syntenic to wheat chromosome arm 1BS (17,24). Figure 5 illustrates the physical position of the rice orthologs for these HVPs, notably the high concentration of rice orthologs of the 1BS probe sets (variance \(\geq 0.1\)) which are clustered on the short arm of rice chromosome 5 (5S), 42 have a best hit elsewhere in the rice genome, 148 unique genes. The variance scores and rice gene models are listed in Supplementary Table 2.

**Wheat 1BS translocation breakpoint mapping on rice 5S**

Rice chromosome 5 short arm is 12.5 Mb in length (http://www.tigr.org/tdb/e2k1/osa1/). To determine the positions of wheat 1BS translocation breakpoints, we made further use of the physical map coordinates of the rice gene models that match the HVPs. We noted the outermost rice gene model within each HVP marker bin derived from the two breakpoints (T \(1\) and 1B \(+5\)) (Figure 6). The T \(-1\) breakpoint falls within the interval of 0.2–0.3 Mb on rice 5S between gene models LOC_Os05g01110.1 and LOC_Os05g01610.1, whereas the 1B \(+5\) breakpoint was positioned within the 1.6–2.1 Mb interval between rice gene models LOC_Os05g03780.1 and LOC_Os05g04610.1. Therefore, the syntenic region of rice 5S that defines both of these breakpoints is not more than 2.1 Mb from the telomere end.

**DISCUSSION**

Alternative estimations of HVPs for wheat 1BS

As noted in the Introduction section, a total of 326 ESTs previously were associated with loci on 1BS by DNA hybridization of cDNAs to genomic DNA (17). But these were just a subset of 769 ESTs that were associated with one or more 1S arm. We searched these 769 ESTs against the content of the Affymetrix wheat GeneChip and then compared the best blastn hits to the HVP list. A total of 2983 HVPs from nine informative pairwise genotype comparisons are compiled in Supplementary Table 1. Of the 769 ‘mapped’ ESTs, 101 had a match to an HVP wheat source sequence. Among these 101, a total of 52 have a best hit to a gene model on rice 5S, 42 have a best hit elsewhere in the rice genome,
and seven have no rice best hit of e−4 or better. Therefore, we can conclude that the true percentage of HVPs associated with 1BS is approximately double (101/52 × 100% = 194%) the number that results from the more conservative view of only counting those with rice 5S best hits. Among all 2983 HVPs (not just those matching mapped ESTs), 208 HVPs are associated with wheat 1BS based on wheat 1BS-rice 5S synteny. When adding 49 HVPs that do not have rice 5S best hits but match ESTs mapped on wheat 1BS, a total of 257 HVPs could be used preferentially as wheat 1BS markers. Given the homoeologous nature of wheat arms these HVPs must equally mark all three homoeologous arms.

Figure 4. HVP-synteny-based mapping of translocation breakpoints in Pavon (1BS). 1RS.1BL is a translocation of the entire rye short arm of chromosome 1; 1B + 5 is a terminal translocation of rye 1R onto wheat chromosome 1B. Arrows indicate the concentration of rice 5S gene models corresponding to high variance probe sets at >0.1 variance from the three pairwise wheat genotype comparisons shown: (A) concentration of HVPs on entire rice 5S in Pavon versus 1RS.1BL, (B) concentration of HVPs on proximal rice 5S in 1B + 5 versus 1RS.1BL, (C) concentration of HVPs on terminal rice 5S in 1B versus 1B + 5.

Figure 5. Map position of HVPs between wheat 1BS versus Dt1BL on 12 chromosomes of rice. Inset: Pavon has the 1BS arm while the Pavon Dt1BL is missing 1BS. Rice chromosome 5 short arm is known to be syntenic to wheat chromosome 1BS. Vertical lines indicate rice gene model representing the wheat HVPs and the filled circles are the centromere positions. Numbers indicate which rice chromosomes; values on x-axis indicate base pairs. Variance cutoff = 0.1.

Why do most HVPs not have a blast hit on Rice 5S?
Most of the HVPs (~80%) in any given pairwise comparison do not have a blast hit on rice 5S. Many might arise from unequal consequences of aneuploidy on gene expression across the entire genome. In other words, deletions of 1BS might cause changes in the relative proportion of A, B and D genome transcripts from genes located anywhere else in the genome. Such changes in relative expression levels would be detected as HVPs. Addition or subtraction of a single chromosome arm altered gene expression patterns extensively in maize (25).

Use of microarrays for genotyping and mapping
The use of microarrays for genotyping by single feature polymorphisms (SFPs) has been reported using DNA in small-genome (140 Mb) Arabidopsis (26). Similarly, RNA was used as a surrogate for DNA to identify SFPs in large-genome (5300 Mb) barley using the Affymetrix barley genome array (22,27). The concept of genotyping with expression microarrays by hybridizing with cRNA instead of genomic DNA for simultaneous genotyping and gene expression analyses was also demonstrated in yeast (28) and subsequently Arabidopsis (29). The possibility to use SFPs for genome-wide association mapping or linkage disequilibrium has also been discussed (30). Here we report another variant of SFP analysis, one that seeks HVPs for assignment of genes to regions of one chromosome arm and translocation breakpoint mapping in polyploid wheat.

The results of this study establish a general approach that could be extended to the entire wheat genome, or to other polyploid organisms for which suitable cytogenetic stocks are available. Specifically, this approach provides a sensitive method to detect deletions, duplications or substitutions within wheat-alien recombinant chromosomes and translocation stocks, including those that carry agronomically desirable genes. Because HVPs are well suited for analyses of cytogenetic stocks, they are...
attractive for the allocation of genes (represented by probe sets) to chromosome arm bins defined by deletions or translocation junctions, and thus their wider application could accelerate and simplify physical mapping of large genomes such as wheat and other species.

Economical gene chip technology for mapping

Our analysis of the cost of arm-specific genetic marker development comparing this method with conventional PCR amplicon methods indicated that this array-based method is more economical. We designed 91 primer pairs from EST sequences found by others using DNA hybridization to be markers of genes in the wheat 1BS region (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi). We screened the five genotypes used in the current study and detected amplicon presence/absence polymorphisms for only eight primer pairs (results not shown). When weighing the modest up-front costs of the genome array and RNA processing against outlays for PCR reagents and personnel time, it appears that the array-based method is the more economical approach to the development of wheat 1S specific markers.

SUPPLEMENTARY DATA

Supplementary Data is available at NAR online.

ACKNOWLEDGEMENTS

Funding was provided by Southwest Consortium on Plant Genetics and Water Resources. The authors wish to thank Dr Christopher Davies and Brant Wong (Bioinformatics Division, Affymetrix Inc Emeryville, CA 94608, USA) for early access to the details of the wheat genome array. Authors thank Rehana Asghar for initial help with RNA isolation, Sundrish Sharma, Harkamal Walia and Bahman Ehdaie for useful discussion. We thank Sriti Misra and Dr Denis J. Heck from the DNA and Protein Microarray Facility, University of California, Irvine for microarray hybridizations. Funding to pay the Open Access publication charge was provided by the National Science Foundation Plant Genome Research Program DBI-0321756.

Conflict of interest statement. None declared.

REFERENCES