# A Multiparent Advanced Generation Inter-Cross for genetic mapping in wheat

Bevan Emma Huang[1], Andrew W George[1], Kerrie L Forrest[2], Andrzej Kilian[3], Matthew J Hayden[2], Matthew K Morell[4], Colin R Cavanagh[4]

[1]CSIRO Mathematics, Informatics and Statistics and Food Futures National Research Flagship, Queensland EcoSciences Precinct, Dutton Park QLD 4102 Australia

[2]Department of Primary Industries Victoria, Victorian AgriBioscience Centre, La Trobe Research and Development Park, Bundoora VIC 3083 Australia

[3]Diversity Arrays Technology, Yarralumla ACT 2600 Australia

[4]CSIRO Plant Industry and Food Futures National Research Flagship, Canberra ACT 2601

Corresponding Author:
Colin Cavanagh
CSIRO Plant Industry and Food Futures National Research Flagship
GPO Box 1600
CANBERRA ACT 2601
Australia
Phone: +61 2 6246 4974

Author E-mail addresses:
Emma.Huang@csiro.au
Andrew.George@csiro.au
Kerrie.Forrest@dpi.vic.gov.au
A.kilian@diversityarrays.com
Matthew.Hayden@dpi.vic.gov.au
Matthew.Morell@csiro.au
Colin.Cavanagh@csiro.au

Running title: Mapping a wheat MAGIC population

Keywords: MAGIC, linkage, wheat, RIL, QTL mapping

Manuscript word count (excluding cover page, supporting information and references): 6968

# Summary

We present the first results from a novel Multi-parent Advanced Generation Inter-Cross (MAGIC) population derived from four elite wheat cultivars. The large size of this MAGIC population (1579 progeny), its diverse genetic composition, and high levels of recombination all contribute to its value as a genetic resource. Applications of this resource include interrogation of the wheat genome and the analysis of gene-trait association in agronomically important wheat phenotypes. Here we report the utilization of a MAGIC population for the first time for linkage map construction. We have constructed a linkage map with 1162 DArT, SNP and SSR markers distributed across all 21 chromosomes. We benchmark this map against a high density DArT consensus map created by integrating more than 100 bi-parental populations. The linkage map forms the basis for further exploration of the genetic architecture within the population, including characterization of linkage disequilibrium, founder contribution and inclusion of an alien translocation into the genetic map. Finally we demonstrate the application of the resource for quantitative trait loci (QTL) mapping using the complex trait plant height as a proof of principle.

# Introduction

The demands being placed on food production globally require greater advances in genetic improvement of our food crops. A crucial step towards achieving these advances is the rapid identification of genetic elements that can be utilized in breeding to provide better control and delivery of agronomic traits. Traditional strategies for identification of such genetic elements in plants have relied heavily on data collected from experimental crosses based on two founders. The founders are chosen to be highly divergent for a trait of interest in order to have increased power for detection of genetic elements associated with this trait. To establish homozygous immortal populations, recombinant inbred lines (RILs) or doubled haploids are produced as progeny. Bi-parental populations are popular for mapping due to the ease of population development, high power for quantitative trait loci (QTL) detection and availability of a wide range of statistical analysis tools (Lynch and Walsh 1998; Doerge 2002). However, their weakness is that loci are mapped with low resolution (large support intervals) due to the limited opportunity for recombination.

To reduce linkage disequilibrium and improve mapping resolution, the advanced intercross (AIC), an extension of RILs, was proposed by Darvasi and Soller (1995). However, the narrow genetic base (two founders) of each cross limits the phenotypic diversity available and hence its usefulness for analyzing multiple traits. As the objective moves from dissecting individual traits to identifying the genetic control of complex multigenic traits (e.g., yield) and multiple traits (e.g., grain quality and

disease resistance) there is likewise a need to move from 'purpose-built' biparental populations to those with a broad genetic and phenotypic base. This, combined with the increased availability of high-throughput marker genotyping systems (Akbari et al. 2006; Landjeva et al. 2007; Berard et al. 2009; Akhunov et al. 2010; Chao et al. 2010; Allen et al. 2011) has inspired interest in more complex breeding designs.

The AIC approach was extended by Mott et al. (2000) to produce heterogeneous stock mouse populations derived from multiple parents. In plants, the Multi-parent Advanced Generation Inter-Cross (MAGIC) offers an alternative to heterogeneous stock populations with similar advantages for mapping (Cavanagh et al. 2008). MAGIC involves inter-crossing a number of parent lines ($2^n$) for *n* generations in a set mating design to combine the genomes of all ($2^n$) parents in the progeny lines. The mating design for a four parent population is depicted in Figure 1.

We have established two MAGIC populations in wheat – one with four founders and the other with eight founders. One of the benefits of creating the four-parent population in advance of the eight-way is in providing a platform for rigorous testing of analysis tools which is intermediate in complexity to a biparental population and the eight-parent population. As complexity increases, bringing together all eight founders coupled with advanced intercrossing of progeny lines provides a mechanism to substantially increase recombination frequency and facilitate further reduction in linkage disequilibrium. These features of MAGIC create an opportunity to identify gene-trait associations with greater resolution. The first results based on a MAGIC approach in the model species *Arabidopsis thaliana* support this by mapping several

known QTL with high precision and detecting novel QTL for germination date (Kover et al. 2009).

In an unsequenced polyploid crop species such as wheat, there are several challenges and corresponding opportunities that may not arise in model organisms. For example, construction of a linkage map requires accounting for issues such as multiple founders, polyploidy and incomplete marker information, but provides an opportunity to create a highly polymorphic informative marker map. The greater complexity and population size afford the ability to map alien translocations (common in wheat) typically ignored in smaller biparental populations. MAGIC also provides the opportunity to simultaneously assess the effect of multiple alleles at a locus and to probe interactions involving allelic diversity across the genome.

In this paper we explore the genetic architecture and practical utilization of the first wheat MAGIC population. First, we construct a linkage map from a four-parent MAGIC population and validate it against a comprehensive DArT consensus map drawing together maps from over 100 biparental populations. Second, we explore our ability to incorporate alien translocations into the linkage map. Third, we characterize the level of linkage disequilibrium (LD) across the genome and compare it with previous estimates for LD from previous studies. Not only is it important to understand the level of LD in order to calculate the marker density required to fully exploit MAGIC, but the LD itself provides a basis for comparison between the genetic structure of MAGIC and those of existing mapping populations. Fourth and finally we demonstrate the power and precision of MAGIC for QTL mapping for plant height, an important trait for yield potential.

# Results

## *MAGIC Map*

In total, 1162 markers were mapped in the population across all 21 chromosomes. As the majority of genotypes arose from DArT markers, the first framework map was composed solely of DArT markers genotyped across 871 lines. Note that in the MAGIC population both DArTs and SNPs provide equivalent levels of information since they are biallelic and all alleles are fixed in the founders and progeny RILs. SNPs were assigned to chromosomes based on their level of recombination with the DArT markers, and all markers were then reordered within groups.

Microsatellites and gene markers were primarily chosen for genotyping to fill gaps in the DArT framework map and to include important agronomic genes. Sixty-one such markers were genotyped, including the reduced height genes *RHT-B1* and *RHT-D1* (Peng et al. 1999) and known flowering genes for photoperiod sensitivity *PPD-D1* (Beales et al. 2007) and *VRN1* (Trevaskis et al. 2003; Yan et al. 2003; Danyluk 2003). Additionally, the STM773 marker was chosen for genotyping based on its position within the *Sr36* translocation (Bariana et al. 2001).

Chromosomes were oriented to match those in the consensus DArT map. Supporting Information 1 provides the full map, which consists of 826 DArT markers, 283 SNPs and 53 microsatellite/gene markers. On average 64 markers were mapped to

each A chromosome, 86 to each B chromosome, and 16 to each D chromosome. The coverage was unequally distributed among the genomes simply due to the availability of markers; the D genome has been reported to have much lower coverage than the A and B genomes (Akhunov et al. 2010). The total genome measured 3894 cM in length, distributed as 1481 cM for the A genome, 1424 cM for the B genome, and 989 cM for the D genome. Thus, on average the three genomes had one marker every 3.3, 2.4 and 8.7 cM, respectively. From the 1162 genetic markers, a total of 963 unique positions were identified. Table 1 summarizes the map coverage and content for each chromosome.

Since different RILs were genotyped on different platforms, we restricted the data to those with less than 50% missing data total (n=871) to consider recombination events. Across the genome, the number of recombination events had a median of 37 and values ranging from 1 to 80. Figure 2 shows the distribution of recombination events from simulation for RILs with two, four and eight parents. For a four-parent MAGIC design the observed median is approximately half that expected from the simulation.

As a measure of map informativeness, we explored the contribution of parental genotypes to the RIL progeny in Figure 3. Similar to the above analysis of recombination events, we considered only RILs with less than 50% missing data overall. For these, we identified the relative contributions of the parents across the genome as well as regions of the genome where individual founder contributions could not be distinguished (low total contributions). Regions such as Chromosome 6D

indicate poor representation in terms of the number of polymorphic markers available, consistent with previous studies (Akhunov et al. 2010).

## *Validation*

A consensus DArT map containing 4606 markers (Supporting Information 2) served as a benchmark for the results from this mapping population. On average 223 markers were mapped to each A chromosome, 313 to each B chromosome, and 122 to each D chromosome. The total consensus map length was 3097 cM. There were 2309 unique positions, with the three genomes having on average one position every 1.29, 0.90 and 2.86 cM, respectively. As might be expected from a map based on over 100 populations, the consensus map contained more DArT markers and had better coverage than the MAGIC map. In total there were 714 DArT markers in common between the maps. Ten markers were assigned different linkage groups in the two maps; however, this does not necessarily indicate misassignments in the MAGIC map - within the DArT consensus map itself there were 49 markers assigned to multiple linkage groups. Omitting these markers left 655 markers for which we compared map positions within chromosomes.

Marker positions are compared between the two maps in Figure 6, where points along the diagonal indicate exact agreement. Three chromosomes (4D, 5D, and 6D) had insufficient (zero or one) markers in common for comparison and were omitted. For the remaining chromosomes, the number of common markers ranged from 7 (Chr 3D) to 98 (Chr 3B). Typically the map length was longer for the MAGIC map, but as shown in Figure 6, the order was in good agreement with the consensus

map. Correlation between map positions for each chromosome ranged from 0.54 (Chr 3D) to 0.998 (Chr 4B). Indeed, apart from Chr 3D the correlations were uniformly larger than 0.938; inspecting Figure 6 we see that the low correlation for Chr 3D is due in part to the small number of common markers, and in part due to the fact that we observed recombination in the MAGIC population between five markers on the chromosome that was not evident in the consensus map.

## *SR36 Translocation*

The distorted inheritance of the Baxter *Sr36* translocation can be seen in Figure 2 on Chromosome 2B. It encompasses a large region of the chromosome near the centromere and generates segregation distortion throughout the chromosome. Prior to removing any markers or lines associated with the *Sr36* translocation, we created a draft order of markers on Chromosome 2B and considered the heatmap for LD between these markers (Figure 4A). Rather than exhibiting compact blocks of LD among tightly linked markers, the LD heatmap displayed long-ranging LD between markers at opposite ends of the chromosome. This indicated that the order was incorrect; indeed, the long-ranging LD turns out to result from the presence of the translocation. In Figure 4B the markers on/near the translocation have been identified and grouped together; now markers outside this region exhibit high LD with neighboring markers but relatively low LD with more distant markers.

The procedure for identifying markers in the translocation highlighted 19 markers which isolated the Baxter allele and displayed the same pattern of distorted linkage and segregation. Some markers, such as BQ172173B_Ta_2_2, had the

genotype BB for Yitpi, AA for Chara and Westonia, and NULL for Baxter, possibly indicating regions of the wheat chromosome replaced with the alien translocation. Other markers in the block included the starch branching enzyme genes *SBEIIa* and *SBEIIb* and microsatellites CFA2278 and STM773. The inclusion of the STM773 microsatellite in the map confirmed that this block contains or surrounds the translocation (Bariana et al. 2001).

Using the markers listed above, we identified 630 out of 1579 lines which are likely to carry the translocation. Note that this represents nearly 40% of the population, indicative of the segregation distortion accompanying the translocation. These lines carry the Baxter allele at every marker where a founder allele for the line can be imputed, and have at least 10 (out of 18) imputed founder alleles present. We compared this set with those identified by the STM773 microsatellite known to be on the translocation. There is a 0.89 correlation between the results; the two classifications agreed for 1222 lines and disagreed for 72.

## *Linkage Disequilibrium*

As the extent of LD affects both mapping resolution and the effectiveness of QTL mapping, understanding LD across the genome is of critical importance in the MAGIC population. It will help to ensure adequate marker coverage is achieved, provide a guide to resolution achievable across the genome and characterize the genetic structure of the RILs. We investigated the extent of LD between pairs of markers within and between chromosomes.

Within chromosomes, the mean LD dropped off quickly, while it was uniformly low between chromosomes. Within chromosomes, the mean LD tended to fall to <0.8 within ~5 cM and to < 0.2 within 40 cM (Figure 5, Supporting Information 3). Between markers on separate chromosomes, the mean LD was 0.0037, while the 95$^{th}$ percentile was 0.017. Typically the furthest extent of LD due to linkage occurred between 50 and 100 cM, where the loess fits intersected this baseline LD value.

LD block structure for all chromosomes can be seen in Supporting Information 4. The number of haplotype blocks per chromosome ranged from 1 (Chrs 3D, 5D, 6D) to 23 (Chr 3B) with a median of 12. The lowest number of blocks, and blocks containing the fewest markers, occurred on the relatively sparse D chromosomes. The number of markers in each block ranged from 2 to 29 (Chr 6B) with a median of 3. Blocks ranged in length from 0 to 11.6 cM, with a median of 2 cM.

## *Plant Height QTL*

QTL analysis of plant height detected nine QTL at a significance level of 0.001. The genomewide QTL profile for this analysis is shown in Figure 7 with 1-LOD support intervals shown in green. The width of these intervals ranged from 10 cM to nearly 60 cM, with the largest intervals occurring on Chr 5B and Chr 2D where the profile was quite flat near the QTL.

All of the QTL were fit together in a full model and summarized in Table 2. The QTL on chromosomes 1B, 5B and 5D were reduced in significance after accounting for other genetic effects. The most significant QTL were detected on chromosomes 2B, 2D, 4B, and 4D. The latter two QTL represent the known effects on plant height of the

reduced height genes *Rht-B1* and *Rht-D1*. In addition, it is expected that the effect detected on 2D may be due to the flowering gene for photo-period sensitivity *PPD-D1*. The two QTL corresponding to the *Rht* genes were located directly at the gene positions (which are at 35 and 104 cM). As the name of the genes indicates, the effects on plant height of founders which carry the genes (Chara and Baxter for *Rht-B1* and Yitpi and Westonia for *Rht-D1*) is negative.

# Discussion

The creation of highly informative, high density linkage maps utilizing current, relevant germplasm in wheat has been hampered due to its polyploidy and low polymorphism. Additionally, standard bi-parental approaches can access only limited amounts of the genetic diversity within elite cultivars. Multi-parent populations provide a mechanism to overcome some of these issues.

This study has highlighted the potential contained in the novel genetic platform of MAGIC for thorough investigation of the largest (16GB) cereal genome and for its use in identifying associations with traits of interest. In this paper we have described the first linkage map created within a MAGIC population in any species. It provides the wheat research community with a relevant and highly informative linkage map for comparative genomic studies. It is evident that in some regions along the genome there are relatively large gaps in the marker map; this is most likely due to lower polymorphism level, particularly on the D genome (Akhunov et al. 2010). On chromosome 3D these gaps are large enough that some markers behave as if unlinked;

however, based on previous studies these markers have been demonstrated to be linked and hence it is likely that population specific recombination may occur.

The MAGIC map compares favorably with the marker order achieved in a consensus map incorporating over 100 bi-parental populations. The observed differences in marker order between the MAGIC map and the DArT consensus map occur most frequently when there is very little spacing between markers, which is not surprising given the low numbers of recombinants. In the consensus map, there are several regions where markers cannot be separated by recombination, but are spread out over 10-20 cM in the MAGIC map. This fits with the hypothesis that MAGIC populations allow greater map resolution due to more generations of breeding. This increase in resolution within the four-parent MAGIC population is expected to be more apparent with a greater number of founders and higher marker density.

The size of the MAGIC population permits the investigation of translocations in more depth than possible in previous mapping populations. We have identified nearly twenty markers surrounding (or in) the *Sr36* translocation on Chr 2B and isolated a set of over 400 lines which almost certainly carry the translocation. Further, we have been able to include the translocation within the map. This has not yet been possible in other mapping populations such as Sunco/Tasman, where Chromosome 2B is discarded due to the presence of the same translocation (Chalmers et al. 2001).

The ability to achieve these gains arises from two factors: the size of the population, and because it is based on four founders (three of which do not carry the translocation). As the population is large, individuals carrying the translocation can be separated from those that do not. By examining the group containing the

translocation, we are able to order all markers. Although some distortion still exists in the map, this was overcome by adding lines without the translocation to improve estimates of marker order in regions outside the translocation.

The process by which we have identified the *Sr36* translocation shows promise for identifying general regions of the genome with modified recombination. The alien translocation exhibits a distinctive pattern due to both a lack of recombination and distorted segregation (Tsilo et al. 2008). These features result in points near/in the translocation appearing anomalous for most visualization tools, including the loess plots of LD against distance, the founder probabilities and heatmaps of recombination and LD. While prior knowledge about the existence and location of this translocation aided in our identification of *Sr36*, in other datasets these distinctive features should still be recognizable. Further development and refinement of the translocation analysis will thus be of particular interest for the eight-parent population where a higher prevalence of translocations is expected. Greater marker density will assist in identifying smaller translocations.

As well as identifying unusual genomic structures such as translocations, MAGIC provides a rich resource for characterizing general genomic structure. Compared to other studies of linkage disequilibrium in diverse germplasm (Crossa et al. 2007; Chao et al. 2010) LD in the MAGIC population extends much further, most likely due to linkage. Chao et al. (2010) found that LD extends further in the D genome than the A and B genomes; this does not appear to be true in MAGIC but is difficult to assess given the relatively low density of markers on the D genome. Chao et al. (2010) found significant LD extending 10-20 cM, while Crossa et al. (2007) found significant LD up to

40 cM. While LD in MAGIC has typically dropped below 0.2 by 40 cM, significant LD still exists at larger distances. This difference in extent of LD will impact the number of markers required for detecting association. Chao et al. (2010) estimated that placing markers at 0.2 cM will ensure LD between them of approximately 0.8; this will require at least 17,500 markers in a map. Since MAGIC comprises a set of closely related lines and hence exhibits greater stretches of LD, it should require fewer markers to ensure the same level of LD between markers used to detect association. The identification of specific regions of high and low LD will help to focus future genotyping efforts and to ensure that regions are adequately covered for fine mapping.

We are currently undertaking assessment of a diverse set of phenotypic traits, including disease resistance, yield potential and end product quality. Due to the genetic diversity captured within the MAGIC population all traits assessed have displayed large genetic variation. As a proof of principle, we have analyzed the important agronomic trait plant height to demonstrate that QTL mapping in MAGIC populations is possible. We illustrate the accuracy and precision of this process using known genes *Rht-B1* and *Rht-D1*. The region surrounding these genes was identified very accurately, but the analysis was unable to accurately identify the effect size for the QTL on chromosome 4D. As highlighted above, this lack of accuracy in estimating effect size is most likely due to limited haplotype information for this particular region, which is unusual compared to the remainder of the genome (Figure 3). To reduce the percentage of areas where the founders are uninformative, we are currently making an effort to increase marker density and informativeness.

The full power of the MAGIC population will be realized when SNP and other high-throughput marker platforms are available to significantly increase marker density. This will impact all of the analyses considered here, including identification of translocations, estimation of LD, and precision and power of QTL mapping. This impact stems from the fact that in areas of low marker density we cannot identify the origin of alleles or recombinations. Indeed, the population exhibits a median of 37 recombinations per line, which is less than the 51 we expect for a biparental RIL and approximately half that indicated by simulations for a four-parent MAGIC. This difference from the estimates is most likely due to the inability to identify recombinations which occur in gaps in the map. Such regions are evident from the "Total" line in Figure 3, which frequently indicates that the founder allele is ambiguous for more than 20% of the lines, with this percentage increasing at times to greater than 50%.

Our eventual goal is to identify a 'pipeline' to efficiently clone underlying genes. This will allow researchers to fully exploit the power and precision of mapping large numbers of QTL in populations such as MAGIC. With MAGIC RILs, it is possible to generate near isogenic lines (NILs) based on QTL mapping results by identifying heterozygous lines underlying QTL for fine mapping. After six generations of selfing, we expect ~2% of the progeny lines to be heterozygous for any locus; therefore the development of NILs can be conducted to include a broad diversity of genetic backgrounds within the population without the need to set up specific backcross populations in a range of backgrounds.

There is currently an explosion in available genetic markers through large SNP assays and genotype by sequencing approaches (GBS) (Sansaloni et al. 2011, Elshire et al. 2011). This, in conjunction with the generation of large, more complex crossing designs (such as eight parent MAGIC populations) will see a dramatic increase in the number of recombination events being assessed. We have extended our analysis software to run on Graphics Processing Units (GPUs) in order to keep pace with marker discovery. This results in over 100-fold increases in speed for estimating recombination and is an essential tool for creating high-resolution linkage maps. Such high-resolution linkage maps are a critical component to facilitate the anchoring of physical maps to the meiotic map. These resources offer an important role in the international effort to sequence the wheat genome (www.wheatgenome.org) which will transform genomic knowledge of wheat.

This paper has reported on the genetic structure of the first wheat MAGIC population as represented by its genetic linkage map, and its application for mapping QTL and estimating the effects of individual founders on a trait of interest. The ability to efficiently screen large numbers of traits within a single population allows us for the first time to seriously develop a systems biology approach to understanding the wheat genome and the consequences of environmental interactions. This will provide insight into the relationships among important, but quite often isolated, domains of wheat research. Such an understanding is critical if we are to meet the challenges being placed on food production.

# Experimental Procedures

## *Plant Material*

Four Australian wheat cultivars (Yitpi, Baxter, Chara and Westonia) were selected as founders of a 4-way MAGIC population based on three broad criteria: (i) genetic diversity based on a genetic survey of international wheat samples (Raman et al. 2010) (ii) diverse geographic distribution to ensure the representation of all major wheat growing regions in Australia (iii) phenotypic diversity for a range of traits based on breeders' recommendations (*e.g.*, agronomic performance). The four cultivars were crossed in two pairs to create F1 seed Yitpi x Baxter (AB) and Chara x Westonia (CD). These F1 seeds were then inter-crossed (AB x CD) via 70 independent crosses to generate 850 "4-way" (ABCD) F1 seed. Each ABCD F1 seed was grown, harvested and had two F2 seeds progressed without vernalization. Hence only spring lines were maintained to the F6 generation by single seed descent (Goulden 1939; Brim 1966) to create the 1579 recombinant inbred lines (RILs) as shown in Figure 1.

*DNA isolation and molecular marker genotyping*

DNA was isolated from leaf material of single plants of the F6 derived RIL lines of the 4-way MAGIC population using Machery-Nagel NucleoSpin® 96 Plant II kits supplied by Scientifix. Genotyping followed and used three types of molecular markers: Diversity Array Technology (DArT) markers, simple sequence repeat (SSR) markers, and single nucleotide polymorphism (SNP) markers. In total 1670 markers were genotyped across 1579 RILs.

Initially, 1285 DArT markers were genotyped using DArT's Wheat PstI (TaqI) 2.0 service (Akbari et al. 2006). To increase density in regions of low polymorphism, 57 simple sequence repeat (SSR) markers were genotyped. This was performed using the M13 tailing system as described by Oetting et al. (1995) with recommended PCR conditions (http://wheat.pw.usda.gov/GG2/index.shtml). PCR product was run on a 3130x1 Genetic Analyser and the output analyzed using GeneMapper® v4.0 software.

Finally, 1440 RILs were genotyped using a two-stage process. First, the parental lines of the 4-way MAGIC population were genotyped with the 1536 wheat SNPs described by Chao et al. (2010). Genotyping was performed by S. Chao at the USDA-ARS genotyping laboratory in Fargo, North Dakota using an Illumina Bead Station and according to standard Illumina GoldenGate™ assay protocols (Illumina, San Diego CA). SNP allele clustering and genotype calling for the resulting data were performed using BeadStudio v3.0 software (Illumina, San Diego CA). Correct genotype calling was ensured by visual inspection of SNP allele clusters. Next, 384 SNPs observed to be polymorphic among the parental lines were selected for genome-wide coverage, based on physical deletion-bin mapping data available from the Wheat SNP Database (http://probes.pw.usda.gov:8080/snpworld/Search). Progeny genotyping was performed on the Illumina BeadXpress platform and according to standard Illumina GoldenGate™ assay protocols (Illumina, San Diego CA). SNP genotype calling was performed using BeadStudio v3.0 software. Errors in allele calling by the BeadStudio software were detected by visual inspection of SNP allele clusters and manually corrected.

*Phenotypic data*

In 2009, a field trial was conducted at Yanco, NSW to assess a range of agronomic traits. The field trial consisted of 1100 F6 derived F8 RILs from the 4-way MAGIC population and check cultivars including the parental lines. The trial was based on a partially replicated (Smith et al. 2006) spatially optimized design that was formulated using the software DiGGer (Coombes 2002). Forty percent of the RILs were replicated and check cultivars were sown in triplicate. In total 1620 (2m x 5m) plots (10 rows per plot) were sown across three irrigation bays and grown to maturity. Irrigation was utilized to reduce water stress. Plant height, the trait of interest, was recorded prior to machine harvest.

## *Map Construction*

The MAGIC map was constructed with the R package mpMap (Huang and George 2011) via the following steps. First, we filtered out monomorphic markers, markers with missing data in the founders, markers with > 20% missing data, and markers with a segregation distortion p-value < 1e-10. These loci were removed and later re-assessed manually for inclusion.

Second, we estimated the recombination fraction (*r*) between all pairs of loci with the function 'mpestrf'. This function maximizes the likelihood of observing data for a pair of markers over discrete values of recombination fractions. The default values for this function were used, ranging from 0 to 0.1 in steps of 0.005, and from 0.1 to 0.5 in steps of 0.01.

Third, we grouped the markers based on the estimated recombination fractions and LOD scores with the function 'mpgroup'. A marker *j* was added to a linkage group if, for one or more markers *k* within the group, $r_{jk} < 0.15$ and $LOD_{jk} > 5.0$. These are the default values for the function and ensure high confidence in linkage group assignments. Within linkage groups we ordered markers with the function 'mporder'. An overall order was constructed that minimizes the total chromosome length based on the maximum likelihood estimates of recombination fractions. We compared local permutations of this order on the basis of minimal crossovers with the function 'compare.orders' to further confirm the validity of our overall ordering. Finally, linkage groups were combined into chromosomes in a framework map based on knowledge of previously mapped markers (http://probes.pw.usda.gov:8080/snpworld/Search; http://www.triticarte.com.au; Somers et al. 2003). Map positions were computed as the sum of adjacent recombination fractions transformed by the Haldane map function.

As a measure of map resolution, we computed recombination events for all lines using the function 'mpprob'. This calculates the multipoint probability at each locus that the observed genotype is inherited from each of the four founders. Founder alleles were imputed if the probability of inheritance exceeded a threshold of 0.7. Recombination events were then imputed at locations where the founder allele changed along the genome. We compared this estimated number of recombination events for each line against that expected from simulation. For biparental RILs, and four and eight parent MAGIC designs, we generated 10,000 datasets based on the

wheat consensus map of length ~2500 cM (Somers et al. 2004) using the package R/ricalc (Broman 2005).

## *Validation*

To assess the quality of the MAGIC map, we performed both internal and external comparisons. Internally we examined a series of diagnostic plots to propose changes to ordering which were then tested through the 'compare.orders' function in R/mpMap. We examined heatmaps based on both recombination fractions and linkage disequilibrium. These provided a tool for visualization of the relationships between all pairs of markers. The color of each point in the heatmap represents the strength of relationship between a pair of markers. In particular, red points indicate two markers with similar genetic patterns, likely to be located close together on the genome. Hence we expect to see red blocks of markers clustered around the diagonal, with lighter colours radiating out from there. Red points far from the diagonal may indicate incorrect order, as this would otherwise suggest tight linkage between physically distant markers. Examples of LD diagnostics are shown in Figure 4, where Fig 4A shows the LD plot for markers following initial ordering on Chromosome 2B while Fig 4B is the updated plot showing the final ordering.

Once the map satisfied all diagnostic checks, we compared it to an external DArT consensus wheat map. A consensus map based on DArT technology was generated from a selection of over 100 bi-parental mapping populations geographically distributed in 21 countries over five continents. Each individual map was based on genotyping involving the analysis of between 206 and 1525 markers,

with an average density of 582 (Detering et al. 2010). The consensus map was previously shown to accurately reflect the physical map for chromosome 3B (Paux et al. 2008).

## *Identification of Translocations*

In wheat, there are a large number of known translocations (Kammholz et al. 2001) which can cause suppression in recombination and differences in recombination fractions between populations. One objective of this research was to test the ability to identify translocations within the MAGIC population and to account for them during the map construction process. We outline our procedure for a region known to contain a translocation and demonstrate the effect of its removal on the resultant map. This serves as a proof of principle; a similar procedure may be followed in other regions where known translocations are located.

*Sr36* is a translocation introduced from *Triticum timopheevii* (Nyquist 1962) for stem rust resistance. It is known to be carried by the variety Baxter on chromosome 2B (Akbari et al. 2006) and cause segregation distortion (Tsilo et al. 2008). Given that Baxter is one of the four founders in our MAGIC population and that we observed segregation distortion consistent with this translocation, we assumed that *Sr36* was present in the MAGIC progeny. There were two aims regarding this translocation: first, to identify markers associated with the translocation; and second, to identify lines containing the translocation. To accomplish the first aim, we computed the degree of segregation distortion across the chromosome. Markers with high segregation distortion for which a) the Baxter allele differed from all other founder alleles, and b)

mutual recombination of less than 0.05 was observed, were tagged as potential markers in the translocation. To accomplish the second aim we estimated the probability of a line having inherited an allele from the founder Baxter for the identified markers (using the function 'mpprob'). Lines uniformly inheriting genetic material from Baxter across the region were considered likely to carry the translocation.

Due to the extended breeding process, final RILs which carry the translocation must have carried it through all previous generations. However, those which do not carry the translocation may have ancestors within the pedigree that do have the translocation. This potential presence of the translocation in earlier generations can cause distortion in the estimation of recombination fractions due to the recombination suppression in these earlier generations. Hence we focused on the subset of RILs which have always carried the translocation to construct an initial order for chromosome 2B. In some regions, such as near the translocation, there were insufficient recombination events in this subset to finalize the order. For these regions we refined the order using recombination fractions estimated from all lines.

## *Linkage Disequilibrium*

We generalize methods for estimating multi-allelic linkage disequilibrium to accommodate the structure of MAGIC. Several methods have been proposed to estimate linkage disequilibrium in the situation of multi-allelic markers (Zhao et al. 2005). These are all based on the $2^n$ x $2^n$ table indicating which two-marker haplotypes are inherited from which of the $2^n$ founders. We focused on two methods here as the

most easily interpretable equivalents of biallelic LD measures. The other methods can be similarly generalized. Let $p_i$ and $q_j$ be the allele frequencies associated with two marker loci where the four founders are labeled A, B, C and D, and *i,j* ϵ *{A, B, C, D}*. Then we compute the quantities

$$r^2 = \sum_{i,j} p_{ij} \frac{\delta_{ij}^{\ 2}}{p_i q_j (1-p_i)(1-q_j)}$$

$$D' = \sum_{i,j} \frac{\delta_{ij}}{D_{\max}}$$

where $\delta_{ij} = \sum_{i,j} p_{ij} - p_i q_j$ , $D_{max}$ is the minimum of $p_i q_j$ and $(1-p_i)(1-q_j)$ when $\delta_{ij} < 0$ and the minimum of $p_i(1-q_j)$ and $(1-p_i)q_j$ when $\delta_{ij} \geq 0$, $r^2$ indicates a generalization of the usual correlation between markers to the multi-allelic case, and D' is a generalization of the usual measurement of historical recombination. Note that in Zhao et al. (2005), $r^2$ is referred to as $r^2_{hap}$; we have simplified the notation here. Both of these quantities range from 0 to 1, but $r^2$ was found by Zhao et al. (2005) to have better behaviour in terms of quantifying usable LD in a population for QTL mapping and marker-assisted selection.

If all observed markers were fully informative, we could apply these measures directly to the data. However, in practice it is rare that the inheritance of the founder allele is directly observed in the finals. Thus, we modify the table of haplotype counts by replacing each count with the probability of a line inheriting each particular founder allele. Multipoint probabilities were computed using the function 'mpprob' in R/mpMap and LD was computed using the function 'mpcalcld'. For interpretation of genomic structure in this study we focus on $r^2$.

We compared the values for $r^2$ at various distances to provide an indication of how it varied with distance across the genome. These distributions were visualized with boxplots, which are a convenient way of graphically displaying the median (line in box), the 25$^{th}$ and 75$^{th}$ percentiles (top and bottom of box), and outliers (distinct points). Boxplots for $r^2$ for all linked markers within certain distance ranges were displayed side-by-side to facilitate comparison of their medians and variation.

Similar to Breseghello and Sorrells (2006), we determined critical values for evidence of linkage from the distribution of LD for unlinked loci. Breseghello and Sorrells took the parametric 95$^{th}$ percentile after fitting a normal distribution to the square-root transformed values; we took the 95$^{th}$ percentile directly since the parametric fit did not appear appropriate. Above this value we expect observed LD to be caused by linkage. Thus we measured the extent of LD in a chromosome by where the weighted least squares regression (loess) curve fit intersected this baseline value.

We determined haplotype blocks by using the basic criterion of minimal LD range. That is, a block is defined as a region in which all pair-wise D' values exceeded a given threshold. We selected a threshold of 0.9 as intermediate in stringency to the criteria of Wang et al. (2002) and of Gabriel et al. (2002). Wang et al. required markers in blocks to exhibit no historical recombination (D'=1), while Gabriel et al. defined pairs of markers to have "strong evidence for historical recombination" if the upper confidence bound on D' was less than 0.9, and defined blocks as regions over which <5% of comparisons showed strong evidence of historical recombination.

Estimates of $r^2$ were based on all markers due to the minimum minor allele frequency (MAF) observed being 0.11. While low MAF values have previously been

shown to significantly affect mean $r^2$ (Yan et al. 2009), this effect should be small since all markers had MAF greater than 0.1.

## *QTL Mapping*

QTL mapping was tested on plant height measurements - derived from plots in the field trial described above. A phenotypic base model for the field grown material was established that included a fixed effect for progeny lines (versus checks); random effects for genotype, irrigation bay and an autoregressive (AR1) field row and field column effect.

For all analyses we used the 'mpIMmm' function in the R package mpMap, which performs interval mapping in the context of a linear mixed model. The base model described above is augmented with separate fixed effects for each founder computed at hypothesized positions of a QTL along the genome. Founder effects were based on multi-point probabilities computed using R/qtl (Broman et al. 2003) at a step size of 2 cM. All mixed models were fitted using ASReml R v3.0 (Butler 2009; Gilmour et al. 2009). A joint Wald test for the significance of all founder effects was then formed at each hypothesized position to determine likely positions of QTL. Identification of QTL peaks occurred for p-values less than 0.001.

# References

Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, A., et al. (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theor Appl Genet,* **113**, 1409-1420.

Akhunov, E.D., Akhunova, A.R., Anderson, O.D., Anderson, J.A., Blake, N., Clegg, M.T., Coleman-Derr, D., Conley, E.J., Crossman, C.C., Deal, K.R., et al. (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics,* **11**, 702.

Allen, A.M., Barker, G.L.A., Berry, S.T., Coghill, J.A., Gwilliam, R., Kirby, S., Robinson, P., Brenchley, R.C., D'Amore, R., McKenzie, N., et al. (2011) Transcript-specific single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (Triticum aestivum L.) *Plant Biotechnology Journal* doi: 10.1111/j.1467-7652.2011.00628.x.

Bariana, H.S., Hayden, M.J., Ahmed, N.U., Bell, J.A., Sharp, P.J., and McIntosh, R.A. (2001) Mapping of durable adult plant and seedling resistances to stripe rust and stem rust diseases in wheat. *Aust J Agric Res,* **52,** 1247-1255.

Beales, J., Turner, A., Griffiths, S., Snape, J.W., and Laurie, D.A. (2007) A pseudo-response regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). *Theor Appl Genet,* **115,** 721–733.

Berard, A., Le Paslier, M.C., Dardevet, M., Exbravat-Vinson, F., Bonnin, I., Cenci, A., Haudry, A., Brunel, D., and Ravel, C. (2009) High-throughput single nucleotide polymorphism genotyping in wheat (Triticum spp.) *Plant Biotechnology Journal,* **7,** 364-374.

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

Brim, C.A. (1966) A modified pedigree method of selection in soybeans. *Crop Science,* **6,** 220.

Broman, K. (2005) The genomes of recombinant inbred lines. *Genetics,* **169,** 1133-1146.

Butler, D. (2009) asreml: asreml() fits the linear mixed model. R package version 3.00.

Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (20080 From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol,* **11,** 215-221.

Chao, S., Dubcovsky, J., Dvorak, J., Luo, M.-C., Baenziger, S.P., Matnyazov, R., Clark, D.R., Talbert, L.E., Anderson, J.A., Dreisigacker, S., et al. (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (Triticium aestivum L.). *BMC Genomics,* **11,** 727.

Chalmers, K.J., Campbell, A.W., Kretschmer, J., Karakousis, A., Henschke, P.H., Pierens, S., Harker, N., Pallotta, M., Cornish, G.B., Shariflou, M.R., et al. (2001) Construction of three linkage maps in bread wheat, Triticum aestivum. *Aust J Agric Res,* **52,** 1089-1119.

Coombes, N.E. (2002) The reactive tabu search for efficient correlated experimental designs. Liverpool John Moores University, Liverpool.

Crossa, J., Burgueno, J., Dreisigacker, S., Vargas, M., Herrera-Foessel, S.A., Lillemo, M., Singh, R.P., Trethowan, R., Warburton, M., Franco, J., et al. (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics,* **177,** 1889-1913.

Danyluk, J., Kane, N.A., Breton, G., Limin, A.E., Fowler, D.B., and Sarhan, F. (2003) TaVRT-1, a putative transcription factor associated with vegetative to reproductive transition in cereals. *Plant Physiology,* **132,** 1849–1860.

Darvasi, A. and Soller, M. (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics,* **141,** 1199-1207.

Detering, F., Huttner, E., Uszynski, G., Wenzl, P., and Kilian, A. (2010) A consensus genetic map of wheat: ordering 5000 wheat DArT markers. 20th International Triticeae Mapping Initiative, Bejing, China.

Doerge, R.W. (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet,* **3**, 43-52.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE,* **6,** e19379 doi: 10.1371/journal.pone.0019379.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., et al. (2002) The structure of haplotype blocks in the human genome. *Science,* **296,** 2225-2229.

Ganal, M.W., Altmann, T., and Röder, M.S. (2009) SNP identification in crop plants. *Curr Opin Plant Biol,* **12,** 211-217.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. (2009) ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES UK.

Goulden, C.H. (1939) Problems in plant selection. In Proceedings of the Seventh International Genetics Congress. Cambridge University Press, pp. 132-133.

Gupta, P.K., Roy, J.K., and Prasad, M. (2001) Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science,* **80,** 524-535.

Huang, B.E. and George, A.W. (2011) R/mpMap: A computational platform for the genetic analysis of multi-parent recombinant inbred lines. *Bioinformatics,* **27,** 727-729.

Jaccoud, D., Peng, K., Feinstein, D., and Kilian, A. (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res,* **29,** e25 doi: 10.1093/nar/29.4.e25.

Kammholz, S.J., Campbell, A.W., Sutherland, M.W., Hollamby, G.J., Martin, P.J., Eastwood, R.F., Barclay, I., Wilson, R.E., Brennan, P.S., and Sheppard, J.A. (2001) Establishment and characterization of wheat genetic mapping populations. *Aust J Agric Res,* **52,** 1079-1088.

Keim, P., Diers, B.W., Olson, T.C., and Shoemaker, R.C. (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. *Genetics,* **126,** 735-742.

Kover, P.X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I.M., Purugganan, M.C., Durrant, C., and Mott, R. (2009) A multiparent advanced generation inter-cross to fine map quantitative traits in Arabidopsis thaliana. *PLoS Genetics,* **5,** e1000551 doi:10.1371/journal.pgen.1000551.

Landjeva, S., Korzun, V., and Borner, A. (2007) Molecular markers: actual and potential contributions to wheat genome characterization and breeding. *Euphytica,* **156,** 271-296.

Lynch, M. and Walsh, B. (1998) Genetics and Analysis of Quantitative Traits. Sunderland, MA: Sinauer. pp. 980.

Mott, R., Talbot, C.J., Turri, M.G., Collins, A.C., and Flint, J. (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc Natl Acad Sci USA,* **97,** 12649-12654.

Nyquist, W.E. (1962) Differential fertilization in the inheritance of stem rust resistance in hybrids involving a common wheat strain derived from Triticum timopheevii. *Genetics,* **47,** 1109-1124.

Oetting, W.S., Lee, H.K., Flanders, D.J., Wiesner, G.L., Sellers, T.A., and King, R.A. (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics,* **30,** 450-458.

Paux, E., Sourdille, P., Salse, J., Saintenac, C., Choulet, F., Leroy, P., Korol, A., Michalak, M., Kianian, S., Spielmeyer, W., et al. (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. *Science,* **322,** 101-104.

Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., et al. (1999) Green revolution genes encode mutant gibberellin response modulators. *Nature,* **400,** 256-261.

Raman, H., Stodart, B.J., Cavanagh, C., Mackay, M., Morell, M., Milgate, A., and Martin, P. (2010) Molecular diversity and genetic structure of modern and traditional landrace cultivars of wheat (Triticum aestivum L.) *Crop and Pasture Science,* **61,** 222-229.

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., and Kilian, A. (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings,* **5,** P54 doi:10.1186/1753-6561-5-S7-P54

Shin, J.-H., Blay, S., McNeney, B., and Graham, J. (2006) LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Soft,* **16,** Code Snippet 3.

Somers, D.J., Kirkpatrick, R., Moniwa, M., and Walsh, A. (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. *Genome,* **46,** 431-437.

Somers, D.J., Isaac, P., and Edwards, K. (2004) A high-density microsatellite consensus map for bread wheat (Triticum aestivum L.) *Theor Appl Genet,* **109,** 1105-1114.

Smith, A.B., Liw, P., and Cullis, B.R. (2006) The design and analysis of multi-phase plant breeding experiments. *J Agric Sci,* **144,** 393-409.

Trevaskis, B., Bagnall, D.J., Ellis, M.H., Peacock, W.J., and Dennis, E.S. (2003) MADS box genes control vernalization-induced flowering in cereals. *Proc Natl Acad Sci USA,* **100,** 13099–13104.

Tsilo, T.J., Jin, Y., and Anderson, J.A. (2008) Diagnostic microsatellite markers for the detection of stem rust resistance gene Sr36 in diverse genetic backgrounds of wheat. *Crop Science,* **48,** 253-261.

Wang, N., Akey, J.M., Zhang, K., Chakraborty, R., and Jin, L. (2002) Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination and mutation. *Am J Hum Genet,* **71,** 1227-1234.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., and Dubcovsky, J. (2003) Positional cloning of the wheat vernalization gene VRN1. *Proc Natl Acad Sci USA,* **100,** 6263–62687.

Zhao, H., Nettleton, D., Soller, M., and Dekkers, J.C.M. (2005) Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genet Res Camb,* **86,** 77-87.

# Supporting Information

**Supporting Information 1.** Integrated MAGIC wheat map

**Supporting Information 2.** Consensus DArT map

**Supporting Information 3.** LD as a function of distance (cM) for all chromosomes. Points indicate $r^2$ measure of LD; orange curve indicates the loess fit; blue horizontal line indicates the baseline value of LD (=0.017). Loess curves are not fit for D chromosomes due to the lower marker density.

**Supporting Information 4.** LD heatmaps ($r^2$) for all chromosomes. Red blocks along the diagonal typically indicate haplotype blocks. Map positions of markers are indicated by black lines along the diagonal.

# Tables

**Table 1.** Summary of MAGIC map chromosomes. Columns denote the chromosome name, the number of DArTs, SNPs, SSRs and total number of markers mapped and the length in centiMorgans.

| Chromosome | No. DArTs | No. SNPs | No. SSRs | No. Markers | Length (cM) |
|---|---|---|---|---|---|
| **1A** | 63 | 14 | 2 | 79 | 185 |
| **1B** | 62 | 18 | 2 | 82 | 213 |
| **1D** | 12 | 4 | 2 | 18 | 90 |
| **2A** | 24 | 17 | 1 | 42 | 184 |
| **2B** | 66 | 31 | 5 | 102 | 225 |
| **2D** | 23 | 6 | 4 | 33 | 163 |
| **3A** | 27 | 18 | 1 | 46 | 218 |
| **3B** | 117 | 9 | 2 | 128 | 331 |
| **3D** | 11 | 0 | 2 | 13 | 104 |
| **4A** | 43 | 25 | 4 | 72 | 257 |
| **4B** | 13 | 6 | 2 | 21 | 128 |
| **4D** | 2 | 3 | 2 | 7 | 104 |
| **5A** | 22 | 21 | 4 | 47 | 210 |
| **5B** | 68 | 17 | 3 | 88 | 204 |
| **5D** | 3 | 1 | 8 | 12 | 239 |
| **6A** | 51 | 21 | 0 | 72 | 210 |
| **6B** | 81 | 21 | 0 | 102 | 185 |
| **6D** | 0 | 2 | 1 | 3 | 5 |
| **7A** | 69 | 18 | 0 | 87 | 216 |
| **7B** | 52 | 28 | 0 | 80 | 138 |
| **7D** | 17 | 3 | 8 | 28 | 285 |
| **All** | **826** | **283** | **53** | **1162** | **3894** |

**Table 2.** QTL estimates from fit of full model of plant height, including chromosome (Chr), 1-LOD support interval lower (SI L) and upper (SI U) bounds, position (Pos), left and right flanking markers (LMrk and RMrk), Wald statistic for test that no parents have effects on trait (Wald, 3 df), p-value (p), and effects and standard errors for founder effects for Yitpi (Y), Chara (C) and Baxter (B) relative to Westonia.

| Chr | SI L | Pos | SI U | LMrk | RMrk |
|---|---|---|---|---|---|
| 1B | 194 | 198.4 | 204 | wPt-3765 | wPt-7746 |
| 1B | 126 | 132.6 | 144 | BF485168B_Ta_2_1 | wPt-8280 |
| 2B | 72 | 80.6 | 91.9 | BE404601B_Ta_2_1 | BE445278B_Ta_2_3 |
| 2D | 0 | 14.0 | 58 | *PPD-D1* | GWM102 |
| 4A | 201.5 | 214.0 | 226 | wPt-3655 | wPt-2903 |
| 4B | 30 | 35.45 | 40 | *RHT-B1* | wPt-1708 |
| 4D | 94 | 103.7 | 103.7 | wPt-0941 | *RHT-D1* |
| 5B | 56.0 | 64.67 | 104 | wPt-3289 | BE497820B_Ta_2_3 |
| 5D | 0 | 0 | 28 | CFA2104_1 | BE445181D_Ta_1_1 |

| Chr | Wald | p | effY | seY | effC | seC | effB | seB |
|---|---|---|---|---|---|---|---|---|
| 1B | 11.35 | 0.010 | 1.49 | 2.81 | 3.85 | 1.19 | 5.22 | 2.90 |
| 1B | 6.04 | 0.11 | 0.28 | 1.08 | 0.49 | 0.97 | -1.77 | 0.96 |
| 2B | 79.98 | 0 | 1.76 | 1.02 | 2.82 | 1.26 | -5.26 | 0.92 |
| 2D | 53.01 | 1.82e-11 | 18.5 | 3.16 | 7.87 | 3.04 | 4.90 | 2.58 |
| 4A | 21.19 | 9.63e-5 | -8.08 | 2.29 | -9.06 | 1.99 | -8.02 | 2.15 |
| 4B | 281.59 | 0 | 0.16 | 0.93 | -12.17 | 1.53 | -12.42 | 1.33 |
| 4D | 316.06 | 0 | -6.06 | 13.2 | 10.46 | 6.73 | 12.48 | 6.81 |
| 5B | 10.16 | 0.017 | 0.80 | 1.05 | 2.87 | 1.13 | 3.19 | 1.34 |
| 5D | 8.66 | 0.034 | -0.27 | 1.03 | -3.81 | 1.38 | 1.38 | 1.31 |

**Note:** effY, effC, and effB denote the estimated effect size of the Yitpi, Chara and Baxter alleles respectively, relative to the Westonia effect size. seY, seC and seB denote the standard error of the corresponding estimated effect size.

# Figure Legends

**Figure 1.** Diagram of 4-way MAGIC mapping population A=Yitpi; B=Baxter; C=Chara; D=Westonia

**Figure 2.** Distribution of recombination events for biparental RILs (red line), four-parent MAGIC (blue line) and eight-parent MAGIC (black line). Histograms are based on 10,000 simulations of datasets using a wheat consensus map of total length ~2500 cM.

**Figure 3.** Genomewide founder and total assignment percentages for all chromosomes except 6D, calculated on 871 RILs with low missing data. Orange lines indicate boundaries of chromosomes

**Figure 4.** Heatmaps for linkage disequilibrium ($r^2$) for (A) Chromosome 2B before accounting for SR36 translocation (B) Chromosome 2B after accounting for SR36 translocation. The position of the genetic marker stm773 is labeled since it is known to be in the translocation

**Figure 5.** Boxplots comparing LD measure $r^2$ across a range of distances between linked markers.

**Figure 6.** Comparison of MAGIC and DArT maps with regard to marker position for all chromosomes except 4D, 5D and 6D

**Figure 7.** QTL profile for plant height on chromosomes with QTL detected at $p < 0.001$. Red line indicates significance threshold; green regions indicate 1-LOD support intervals

# Figures

**Figure 1.**

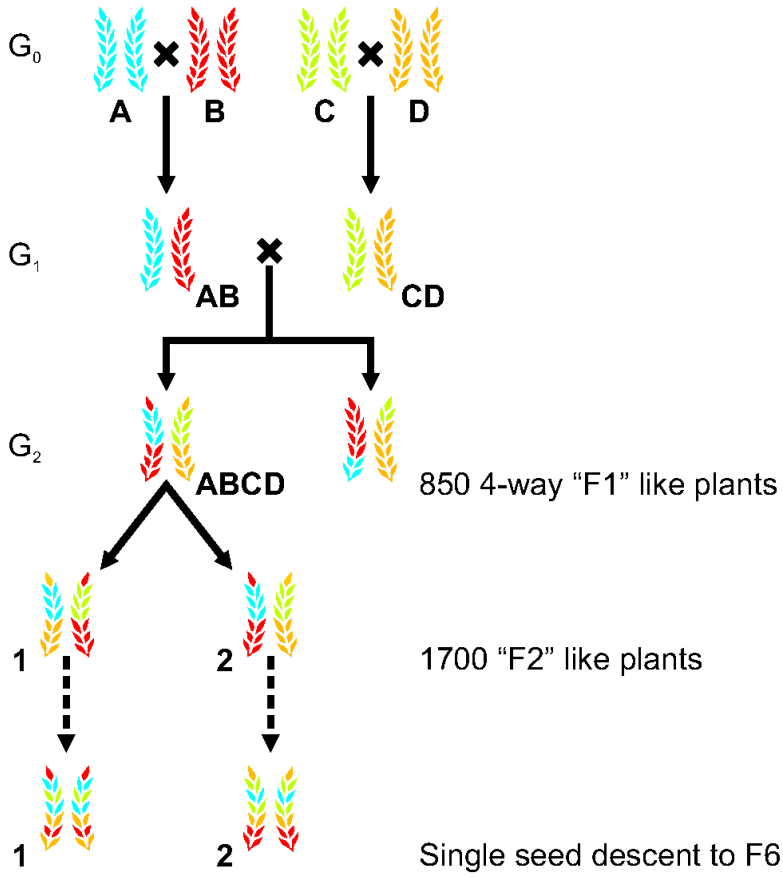

**Figure 2.**

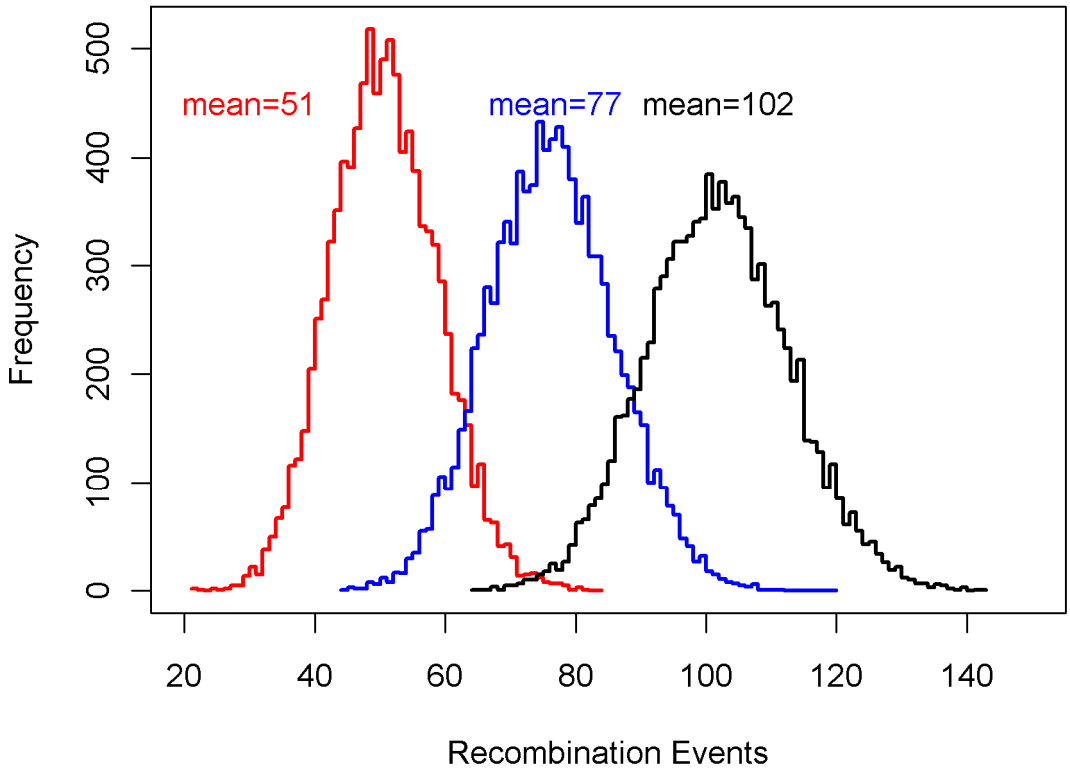
mean=51
mean=77
mean=102
Frequency
0
100
200
300
400
500
20
40
60
80
100
120
140
Recombination Events

**Figure 3.**

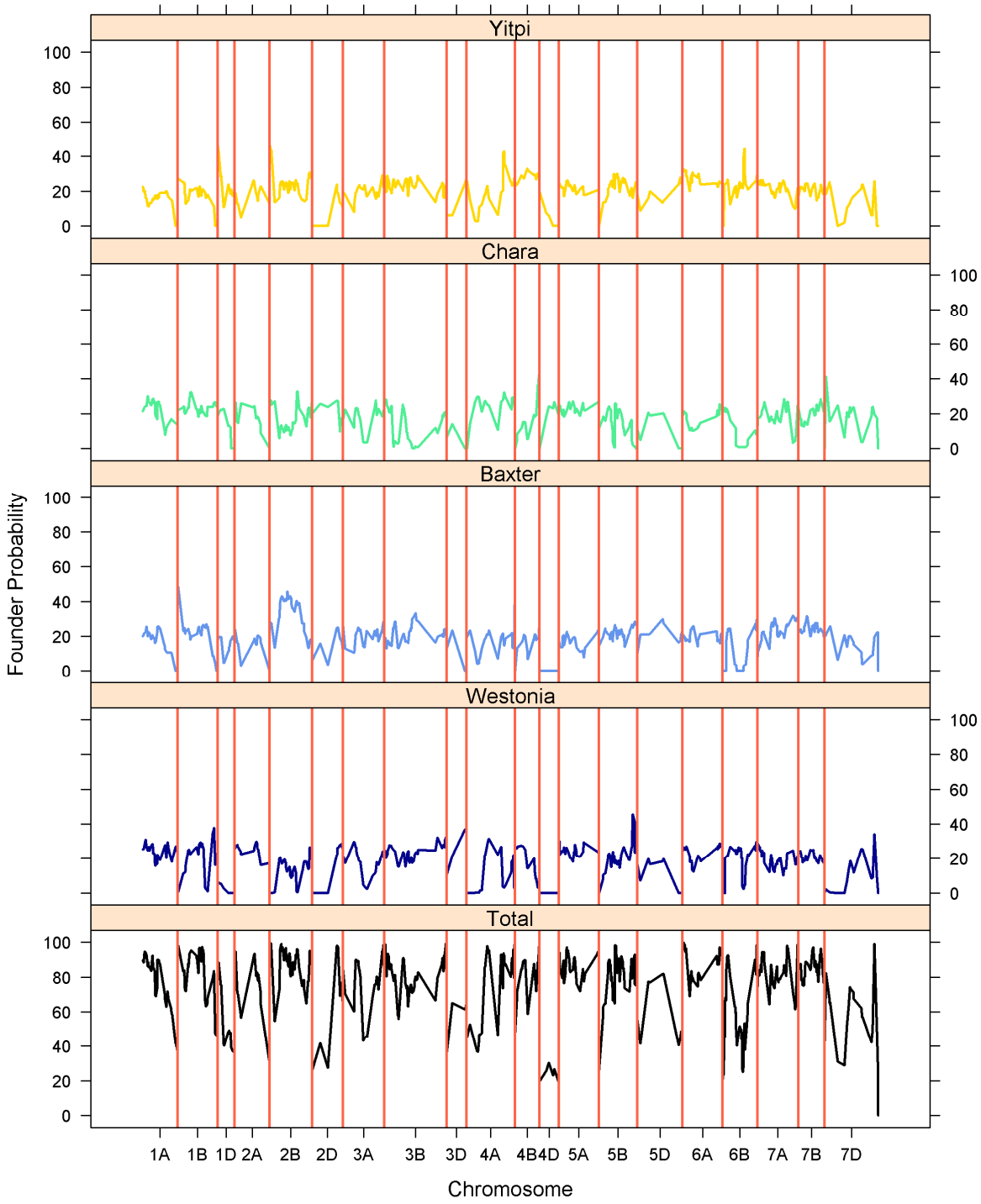
Yitpi
Chara
Baxter
Westonia
Total
Founder Probability
0
20
40
60
80
100
1A
1B
1D
2A
2B
2D
3A
3B
3D
4A
4B
4D
5A
5B
5D
6A
6B
7A
7B
7D
Chromosome

**Figure 4.**

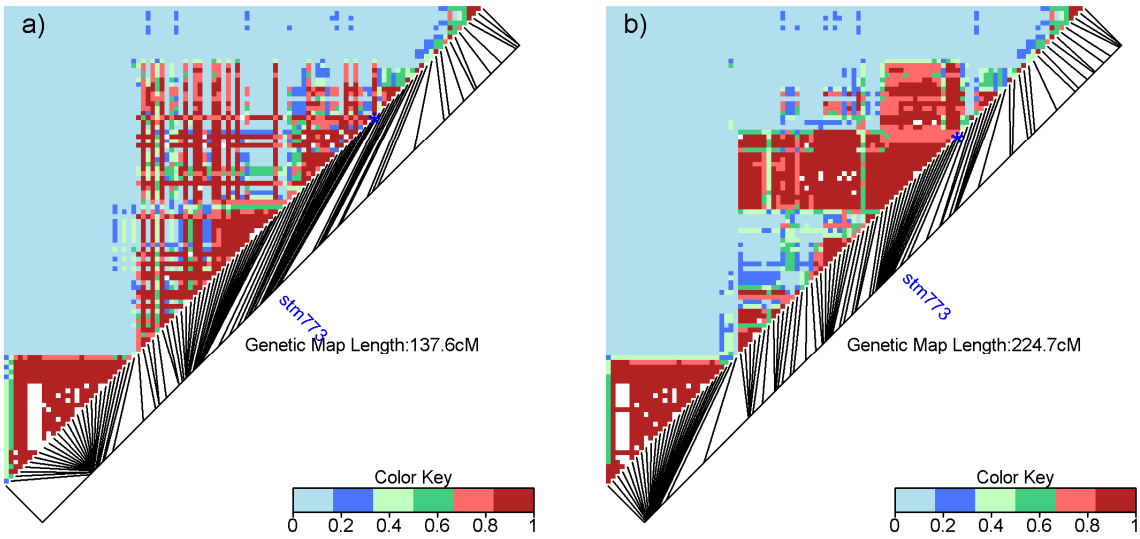
a)
stm773
Genetic Map Length:137.6cM
Color Key
0 0.2 0.4 0.6 0.8 1
b)
stm773
Genetic Map Length:224.7cM
Color Key
0 0.2 0.4 0.6 0.8 1

**Figure 5.**

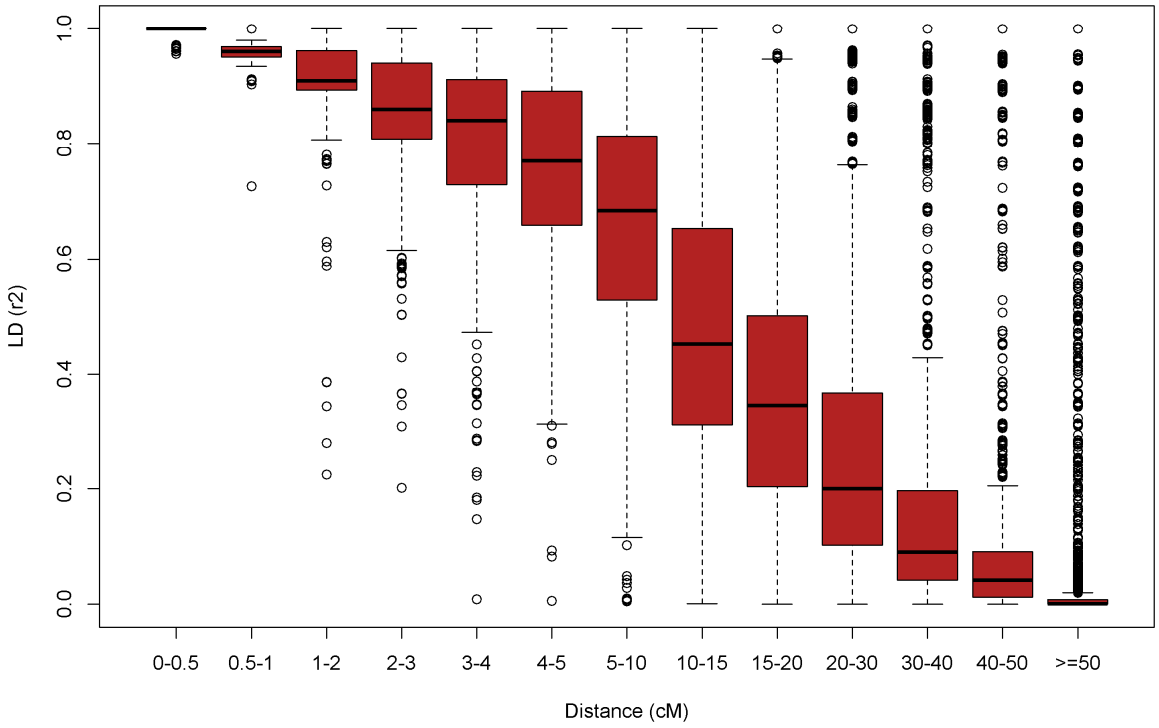
LD (r2)
0.0
0.2
0.4
0.6
0.8
1.0
0-0.5
0.5-1
1-2
2-3
3-4
4-5
5-10
10-15
15-20
20-30
30-40
40-50
>=50
Distance (cM)

**Figure 6.**

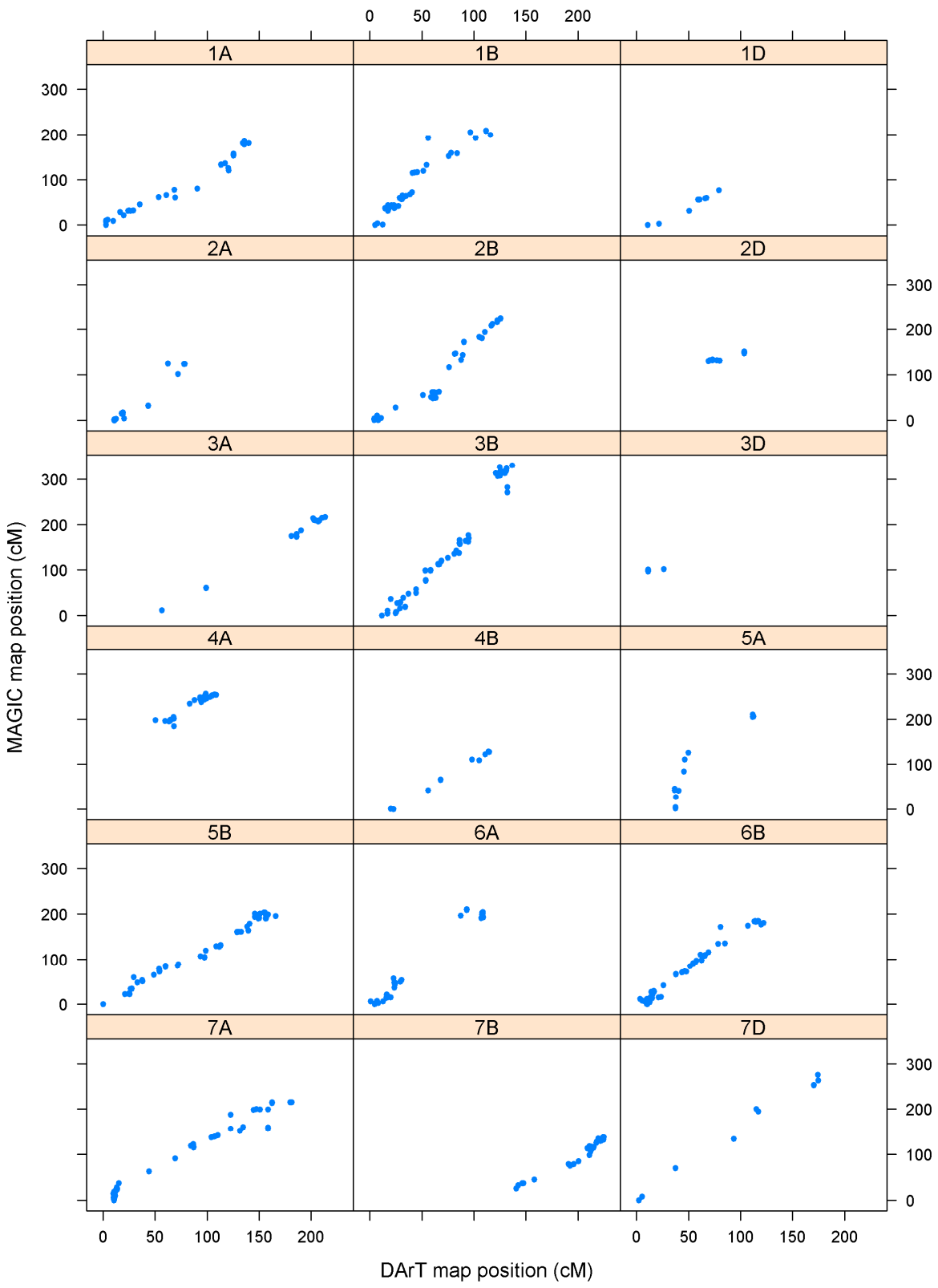
1A
1B
1D
2A
2B
2D
3A
3B
3D
4A
4B
5A
5B
6A
6B
7A
7B
7D
0
50
100
150
200
300
MAGIC map position (cM)
DArT map position (cM)

**Figure 7.**