

# Construction of dense linkage maps “on the fly” using early generation wheat breeding populations

J. T. Eckard · J. L. Gonzalez-Hernandez ·  
S. Chao · P. St Amand · G. Bai

Received: 4 March 2014 / Accepted: 15 May 2014  
© Springer Science+Business Media Dordrecht 2014

**Abstract** In plant species, construction of framework linkage maps to facilitate quantitative trait loci mapping and molecular breeding has been confined to experimental mapping populations. However, development and evaluation of these populations is detached from breeding efforts for cultivar development. In this study, we demonstrate that dense and reliable linkage maps can be constructed using extant breeding populations derived from a large number of crosses, thus eliminating the need for extraneous population development. Using 565 segregating F<sub>1</sub> progeny from 28 four-way cross breeding populations, a linkage map of the hexaploid wheat genome consisting of 3,785 single nucleotide polymorphism (SNP) loci and 22 simple sequence repeat loci was developed. Map estimation was facilitated by application of mapping algorithms for general pedigrees

implemented in the software package CRI-MAP. The developed linkage maps showed high rank-order concordance with a SNP consensus map developed from seven mapping studies. Therefore, the linkage mapping methodology presented here represents a resource efficient approach for plant breeding programs that enables development of dense linkage maps “on the fly” to support molecular breeding efforts.

**Keywords** Linkage mapping · Consensus map · Pedigree analysis · Wheat breeding · High-throughput genotyping

## Introduction

Genetic linkage maps, consisting of linked marker loci ordered along chromosomes, provide the essential framework for identifying genomic regions involved in trait expression and detection of marker-trait associations to enable molecular breeding. For plant species, linkage mapping has largely been confined to an experimental paradigm, in which a purpose-built population derived from a cross between two inbred lines is used for map construction. These experimental mapping populations have been attractive tools for genetic mapping due to their simplicity of development, power for quantitative trait loci (QTL) detection and applicability of available mapping algorithms and user-friendly software implementations (Doerge

---

**Electronic supplementary material** The online version of this article (doi:[10.1007/s11032-014-0116-1](https://doi.org/10.1007/s11032-014-0116-1)) contains supplementary material, which is available to authorized users.

---

J. T. Eckard · J. L. Gonzalez-Hernandez (✉)  
Department of Plant Science, South Dakota State  
University, Brookings, SD, USA  
e-mail: jose.gonzalez@sdstate.edu

S. Chao  
USDA, Agricultural Research Service, Fargo, ND, USA

P. St Amand · G. Bai  
USDA, Agricultural Research Service, Manhattan, KS,  
USA

2002). Despite these attractive attributes, experimental mapping populations have important limitations for the development of genetic linkage maps owing to their narrow genetic base and detachment from applied breeding efforts (Crepieux et al. 2004).

The increasing availability of high-throughput marker genotyping platforms (Kilian et al. 2003; Akbari et al. 2006; Akhunov et al. 2011; Allen et al. 2011; Deschamps et al. 2012) provides a means for the development of dense linkage maps (Bowers et al. 2012). However, the narrow genetic base of conventional mapping populations means that only a small fraction of these marker loci are polymorphic and thus informative for mapping in any given population. Development of a linkage map with dense marker coverage has therefore required integration of maps from several experimental populations to form a consensus map (Wu et al. 2008). However, the process of developing and genotyping several large populations that have no direct contribution to cultivar development is inefficient from the perspective of an applied breeding program. Furthermore, conventional mapping algorithms can only facilitate joint likelihood estimation of the linkage map if each of the constituent populations is of the same structure (Wu et al. 2008). Therefore, consensus mapping is computationally inefficient since it often relies on interpolation of disparate estimates from each constituent population.

To overcome the limitations of conventional mapping populations, researchers have proposed using broad-based populations derived from multi-parent crosses. For example, the maize nested association mapping (NAM) population consists of 5,000 recombinant inbred lines derived by crossing 25 diverse founders to a single common founder (Yu et al. 2008). Early generation four-way cross populations have been used for genetic mapping in cotton and wheat, facilitating the development of higher density genetic maps relative to biparental populations (Trebbi et al. 2008; Qin et al. 2008). This concept of developing “*n*-way” intercross mapping populations has been extended to recombinant inbred lines, with such populations referred to multi-parent advanced generation intercross (MAGIC) populations (Cavanagh et al. 2008). A MAGIC population of over 1,500 RILs from a four-way cross among elite wheat cultivars was used to map 1,162 simple sequence repeat (SSR), single nucleotide polymorphism (SNP) and DArT loci (Huang et al. 2012) and later used to map 4,300 SNPs

(Cavanagh et al. 2013). MAGIC populations have also been developed for rice using eight-way crosses among elite *indica* and *japonica* lines (Bandillo et al. 2013). These rice populations enabled the identification 17,387 polymorphic SNP loci using genotype-by-sequencing methods (Bandillo et al. 2013).

The preceding results indicate that multi-parent mapping populations, also referred to as “second generation” (Rakshit et al. 2012) or “next generation” (Morrell et al. 2012) mapping populations, provide a powerful resource for the construction of dense linkage maps. A major motivation for utilizing these multi-parent mapping populations is that they more closely resemble the broad genetic base and multi-allelic/multi-genic inheritance of breeding populations and thus provide more direct inference for QTL mapping applications (Holland 2007). However, considerable time and resources are required for the development and evaluation of these complex experimental populations, which detracts from applied breeding efforts for cultivar development. For example, development of an *n*-way cross for deriving a MAGIC population requires *n*/2 generations of intercrossing to generate the base population, followed by 6–7 generations of inbreeding to develop the RILs (Rakshit et al. 2012).

Plant breeders are continually developing a large number of segregating populations through carefully planned crosses among numerous elite parents. Primary segregating populations from these crosses consist of relatively small sibships that have yet to be subjected to intense selection by breeders. Collectively, these early generation breeding populations represent a substantial pool of informative genetic recombinations that can be used for the development of dense genetic maps. Linkage maps developed using a large number of early generation breeding populations should therefore provide comparable marker density and reliability to those developed using consensus mapping or multi-parent mapping populations. Thus, utilizing existing populations available in plant breeding programs for the purpose of genetic linkage mapping should provide an efficient alternative to development of experimental mapping populations. In combination with high-throughput and nondestructive genotyping technologies, this strategy would allow breeders to develop dense linkage maps “on the fly” to support their molecular breeding efforts.

Development of linkage maps using plant breeding populations requires mapping algorithms that can handle partially informative marker data, ambiguous linkage phases and simultaneous analysis of numerous small sibships of arbitrary population structure. Due to the prevailing experimental paradigm, such generalized mapping algorithms are not incorporated into software used for linkage mapping in plant populations (Cheema and Dicks 2009). However, these mapping algorithms are commonly used for linkage analysis in humans and animals populations, where purpose-built populations are not available. Therefore, mapping algorithms used for multi-point linkage analysis in general pedigrees (Lander and Green 1987; Weaver et al. 1992) can be adopted for the purpose of constructing linkage maps in plant breeding populations.

In this study, we apply multi-point linkage analysis of general pedigrees to develop a dense genetic linkage map from a 9,000 SNP array using early generation wheat breeding populations. Our objectives are to (1) confirm that dense linkage maps can be obtained using primary segregating populations from breeding programs and (2) assess the accuracy of such derived linkage maps by evaluating their concordance with a recently released SNP consensus map of the wheat genome.

## Materials and methods

### Plant materials

Segregating  $F_1$  populations were developed from 28 four-way crosses among 10 winter wheat founder lines (Table 1). These wheat breeding populations were developed for the purpose of pyramiding resistance loci for Fusarium head blight. Founders included two backcross derived lines Wesley-Fhb1-BC06 and Wesley-Fhb1-BC56 (Wesley/2\*ND2928), an experimental line AL-107-6106 (Alsen/NE00403//NE02583-107), hard winter wheat cultivars Lyman (KS93U134/Arapahoe), Overland (Millennium sib//Seward/Archer), NE06545 (KS92-946-B-I5-1/Alliance), NI08708 (CO980829/Wesley), McGill (NE92458/Ike), and soft winter wheat cultivars Ernie (Pike/MO9965) and Freedom (GR876/OH217). A total of 565 four-way  $F_1$  plants were derived from the 28 four-way crosses, with an average of 20 four-way  $F_1$  plants per cross

(Table 1). Founder lines and four-way  $F_1$  plants were vernalized and then transplanted as individual plants in 4 × 4 inch pots in a greenhouse.

### DNA extraction

Approximately 2 g of healthy leaf tissue was collected from each founder and four-way  $F_1$  plant. For founder lines, tissue samples from multiple plants were pooled into a single sample. Leaf tissue was transferred to liquid nitrogen immediately upon collection and subsequently stored at  $-80^\circ\text{C}$  to prevent degradation. DNA was isolated from the leaf tissue using a midiprep phenol/chloroform extraction protocol adapted from Karakousis and Langridge (2003). Briefly, leaf tissue was flash frozen in liquid nitrogen and then ground to a fine powder using a mortar and pestle. Ground leaf tissue was then mixed with 5 mL of DNA extraction buffer (1 % *n*-lauroylsarcosine, 100 mM Tris-base, 100 mM NaCl, 10 mM EDTA, 2 % polyvinyl-polypyrrolidone, pH 8.5) and 5 mL of phenol/chloroform/isoamyl alcohol 25:24:1 saturated with 10 mM Tris (pH 8.0) for nucleic acid separation. After mixing and centrifugation, the supernatant was transferred to a 10:1 solution of isopropanol and sodium acetate for overnight precipitation of nucleic acids. Precipitated nucleic acid was pelletized by centrifugation and washed with 70 % molecular grade ethanol to remove salts. After drying, the pellet was suspended in 10 mM Tris buffer (pH 8.0) containing 40 µg/mL of RNase A.

### Genotyping

Founder lines and all 565 four-way  $F_1$  plants were genotyped at 26 polymorphic simple SSR marker loci. SSR genotyping was conducted at the USDA-ARS Hard Winter Wheat Genetics Research Unit, Manhattan, KS. PCR was conducted in 14 µL PCR, consisting of 40 ng of template DNA, 0.1 µM of each primer, 0.2 mM of each dNTP, 1× ammonium sulfide PCR buffer, 2.5 mM  $\text{MgCl}_2$  and 0.6 unit of *Taq* polymerase. A touchdown PCR program described by Zhang et al. (2012) was used for amplification. Differentially labeled primers were used for fluorescence detection of SSR amplicons from multiplex PCR as described by Zhang et al. (2012). PCR products were separated and detected using an ABI Prism 3730 Genetic Analyzer, and allele calls were made from resulting fluorescence

**Table 1** Summary of breeding populations used for linkage map development, including the pedigrees, number of progeny and number of plants used for SNP genotyping

Population	Pedigree	Four-way F1 plants	SNP genotyped
01	Wesley-Fhb1-BC56/NE06545//Ernie/Overland	20	19
03	Ernie/Wesley-Fhb1-BC06//Ernie/NE06545	26	–
05	Ernie/Wesley-Fhb1-BC06//Lyman/AL-107-6106	22	16
06	Ernie/Wesley-Fhb1-BC56//Ernie/Lyman	40	37
09	Ernie/Wesley-Fhb1-BC56//NI08708/Lyman	40	38
10	Ernie/Lyman//Ernie/Wesley-Fhb1-BC06	12	–
14	Ernie/Overland//Freedom/Wesley-Fhb1-BC56	5	–
16	Ernie/Overland//Overland/Wesley-Fhb1-BC56	24	23
17	Ernie/Overland//NI08708/Wesley-Fhb1-BC06	33	30
20	Ernie/NE06545//McGill/Wesley-Fhb1-BC56	28	–
23	Ernie/McGill//Lyman/Wesley-Fhb1-BC06	12	9
26	Freedom/Wesley-Fhb1-BC06//Ernie/Overland	12	9
27	Freedom/Wesley-Fhb1-BC06//Lyman/AL-107-6106	7	7
28	Freedom/Wesley-Fhb1-BC06//Overland/Wesley-Fhb1-BC56	11	9
30	Freedom/Wesley-Fhb1-BC56//Ernie/NE06545	4	–
35	Freedom/Ernie//Overland/Wesley-Fhb1-BC56	34	30
36	Freedom/Ernie//NI08708/Wesley-Fhb1-BC06	29	–
40	Freedom/Overland//Lyman/AL-107-6106	8	8
41	Freedom/NI08708//Wesley-Fhb1-BC56/NE06545	7	–
45	AL-107-6106/Overland//Lyman/Wesley-Fhb1-BC06	11	10
48	AL-107-6106/Overland//NI08708/Lyman	14	14
54	Lyman/Wesley-Fhb1-BC56//Ernie/Lyman	37	35
57	Lyman/Wesley-Fhb1-BC56//NI08708/Lyman	31	–
64	Overland/Wesley-Fhb1-BC56//Ernie/Lyman	44	41
65	Overland/Wesley-Fhb1-BC56//Ernie/NE06545	5	–
67	Overland/McGill//Lyman/Wesley-Fhb1-BC06	12	9
71	NI08708/Wesley-Fhb1-BC06//Ernie/NE06545	8	–
76	NI08708/Lyman//Overland/Wesley-Fhb1-BC56	29	28
Total population		565	372

peaks using GeneMarker version 1.6 (SoftGenetics, LLC). SSR marker genotypes were scored after visual assessment of fluorescence profiles to correct erroneous and ambiguous allele calls.

A subset of 18 populations consisting of 372 four-way F<sub>1</sub> plants (Table 1) were genotyped for approximately 9,000 SNP marker loci. SNP genotyping was conducted using an Infinium 9,000 SNP iSelect Beadchip assay developed for wheat (Cavanagh et al. 2013). The assay was performed using the Illumina BeadStation and iScan instruments at the USDA-ARS Biosciences Research Laboratory, Fargo, ND. GenomeStudio version 2011.1 (Illumina) was used for cluster analysis and SNP genotype calling.

The minimum “GenTrain” score (a measure of the reliability of SNP calling based on cluster distribution) was reduced to 0.05 in GenomeStudio to facilitate delineation of compressed but unambiguous SNP clusters. Genotype clusters were then visually assessed for each SNP and manually revised to improve genotype calling. SNP loci represented by more than three genotypic clusters, and those SNP loci with >20 % deviation from the expected heterozygote frequency under Mendelian segregation were excluded from the analysis to avoid complications of polyploid inheritance.

Mendelian inheritance errors for both SSR and SNP loci were detected using the “prepare” function of

CRI-MAP version 2.504 (Green et al. 1990). Genetic impurities in the founders were diagnosed as outliers (i.e., samples with a low “GenTrain” score) within the respective homozygote cluster and by unexpected segregation patterns. For those cases where genotyping errors could not be rectified, including cases where a founder line conferred more than one allele at a locus within the same population, the genotypic data were replaced with missing values.

### Linkage mapping

Linkage analysis was performed using the software package CRI-MAP version 2.504 (Green et al. 1990). CRI-MAP provides an interactive environment for multi-point maximum-likelihood estimation of linkage maps in general pedigrees. For the purpose of constructing the required pedigree data, each four-way cross was considered as a separate pedigree. The assumption of independence among pedigrees was made without any loss of generality, since the founder lines were phase known. The sex designation of founders and single-cross  $F_1$  hybrids was generally assigned corresponding to the actual crossing scheme, while four-way  $F_1$  plants were arbitrarily designated as females. All linkage analysis in CRI-MAP was conducted on an IBM  $\times 3755$  M2 server with 24 processor cores and 128 GB of RAM.

Maximum-likelihood estimates of pairwise recombination fractions among marker loci were obtained using the “twopoint” option of CRI-MAP. Pairwise recombination fractions and associated LOD scores were exported to JoinMap version 4.0 (van Ooijen 2006) for identification of linkage groups. Marker loci were hierarchically clustered in JoinMap based on independence test LOD scores (van Ooijen 2006). Linkage groups were designated as the hierarchical nodes beyond which no significant disaggregation occurred. This point was reached at LOD scores ranging from 10.0 to 35.0 for different linkage groups. Cross linkage statistics computed by JoinMap were used to combine fragmented linkage groups and unassigned marker loci. Chromosomal assignment of the linkage groups was enabled by cross referencing loci from each linkage group with available SSR and expressed sequence tag (EST) mapping data.

Pairwise recombination fractions were then used to cluster marker loci with recombination fractions  $< 0.001$  into genetic bins. From each genetic bin, the

locus with the greatest number of informative, phase-known meioses was identified and considered to be the “primary” locus representing the genetic bin. These primary loci were used to estimate a linkage map of uniquely ordered loci, prior to incorporating the remaining loci to derive the final linkage map. This strategy reduced the number of possible marker orders that had to be initially interrogated, thus increasing the efficiency of the mapping algorithm.

Linkage maps were constructed using the CRI-MAP “build” option. For each chromosome, several pairs of highly informative primary loci with a recombination fraction of 0.30–0.40 were selected to initialize the map order. A framework linkage map was then constructed from each selected pair of loci by sequentially incorporating the remaining primary loci in decreasing order of informativeness. For this first round of map development, only those loci that mapped to an interval with a likelihood ratio 1,000:1 compared to all other intervals were retained in the map. The resulting set of linkage maps was compared, and the most complete map with the highest likelihood was retained for further development. The stringent likelihood threshold and interrogation of multiple initial map orders provided a reliable framework of highly informative markers for subsequent rounds of map development.

Successive rounds of map development were performed to incorporate the primary loci that could not be uniquely ordered in the first round. The likelihood threshold for incorporation of loci was reduced between each round of map development. Specifically, the successive rounds of map development were conducted using likelihood thresholds of 100:1, 10:1 and 2:1. Between each round of map development, the CRI-MAP “flips4” option was used to test the 24 possible permutations for each set of 4 consecutive loci in the current map order. Any local rearrangements that increased the likelihood were used to revise the current map order prior next round of map development. Any loci that mapped to end of the chromosome with a distance of  $> 30$  cM to the nearest locus were removed from the map and were assumed to either be misclassified to the chromosome or belong to unlinked regions of the same chromosome.

After determining the most likely order of the primary loci, the CRI-MAP “chrompic” function was used to detect loci and individuals resulting in unlikely recombination patterns. Double recombination events

between a set of 3 consecutive loci were considered to be the result of genotyping errors, and these data were replaced with missing values. Double recombinations between loci separated by uninformative regions were retained. Unlikely crossover events that were prevalent at a locus within a specific pedigree were considered to be the result of genotyping errors on the founder lines, and the data for the entire pedigree were replaced with missing values for the locus in question when the errors could not be rectified.

Finally, the remaining loci within genetic bins were incorporated into the linkage map. Unmapped loci having a zero estimated recombination fraction with a primary locus were incorporated as a “haplotyped system,” with the primary locus. CRI-MAP only considers the primary locus in each “haplotyped system” when evaluating marker orders, whereas all loci were used in likelihood calculations (Green et al. 1990). Genetic distances were not forced to zero between markers within “haplotyped systems.” Unmapped loci with nonzero estimated recombination fractions with the primary locus were incorporated into the map order to the right of the primary locus, in decreasing order of informativeness. The CRI-MAP “flips4” option was then iteratively used to permute the local marker orders until no higher likelihood map order could be obtained. The final linkage maps were charted using MapChart version 2.2 (Voorrips 2002).

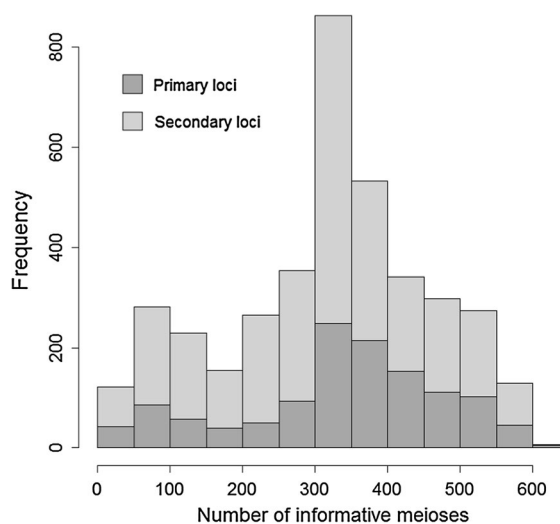
### Concordance analysis

A consensus map of the wheat genome has been developed for the iSelect 9,000 SNP assay through the Triticeae Coordinated Agricultural Project (TCAP) as described by Cavanagh et al. (2013). The TCAP consensus map incorporates a four-way MAGIC population and 6 biparental mapping populations with a combined population size of 2,486 fixed lines. The TCAP consensus map was used as the standard for evaluating the accuracy of linkage maps developed in this study. For each chromosome, Spearman’s rank-order correlation coefficient was computed as a measure of concordance for locus ordering between the TCAP consensus map and the linkage map developed in this study. For each linkage map, relative genetic distances were computed as the locus position in cM divided by the total map length in cM. The relative genetic distances estimated from the breeding populations in this study were plotted against those

from the TCAP consensus map to visually compare patterns of recombination and locus ordering, as well as to diagnose causes of poor concordance.

### Results

Of the approximately 9,000 SNPs assayed, 3,977 were polymorphic and produced clusters that facilitated reliable scoring of SNP genotypes. Additionally, 22 of the 26 SSR loci amplified a product that could be reliably scored, resulting in a total of 3,999 informative loci for subsequent linkage analysis. The average number of informative meioses per SNP locus was 320 and ranged from 20 to 604 (Fig. 1). Hierarchical clustering of these loci in JoinMap resulted in the identification of 31 linkage groups. Cross linkage statistics provided by JoinMap combined with previous SSR and EST mapping data enabled the combination of these linkage groups and unassigned marker loci into 21 groups, putatively representing the 21 wheat chromosomes. Evaluation of recombination fractions identified a total of 1,269 unique genetic bins, with an average bin size of 5 loci. Therefore, 67 % of the interrogated loci were found to be co-localized, resulting from tight linkage among the marker loci as well as markers that interrogated the same locus. Comparatively, only 38 % of these SNP



**Fig. 1** Distribution of the number of informative meioses for polymorphic SNP loci. Primary loci are the most informative loci from each genetic bin used for initial map development, whereas secondary loci are the remaining loci in the genetic bins

**Table 2** Summary of the estimated genetic maps for each chromosome

Chromosomes	Loci mapped		Centimorgans		Recombination		
	Total	Genetic bins	Total	Mean interval	Informative meioses	Observed crossovers	Singletons (<20 cM)
1A	350	88	174.3	2.0	29,086	1,074	19
1B	193	59	160.5	2.8	20,150	919	18
1D	110	30	92.1	3.2	10,532	464	2
2A	214	92	225.3	2.5	30,299	1,450	13
2B	364	72	142.8	2.0	25,932	917	12
2D	50	25	103.2	4.3	9,613	596	7
3A	231	76	171.6	2.3	27,134	1,107	8
3B	288	116	156.9	1.4	40,886	1,245	17
3D	30	11	78.3	7.8	2,762	205	2
4A	104	56	151.8	2.8	20,438	994	7
4B	104	47	140.6	3.1	18,667	901	18
4D	9	7	39.9	6.6	2,325	58	1
5A	296	104	255.5	2.5	41,824	1,750	15
5B	335	113	212.5	1.9	32,417	1,284	12
5D	39	22	190.1	9.1	7,337	747	1
6A	242	72	176.8	2.5	25,239	926	13
6B	338	85	133.7	1.6	29,899	840	7
6D <sup>a</sup>	44	16	45.4	3.0	5,618	104	0
7A	316	105	206.6	1.9	32,964	1,263	4
7B	189	58	186.8	3.3	20,273	1,161	7
7D <sup>a</sup>	29	15	36.2	2.7	3,748	189	1
Overall	3,875	1,269	3,080.9	2.5	437,143	18,194	184

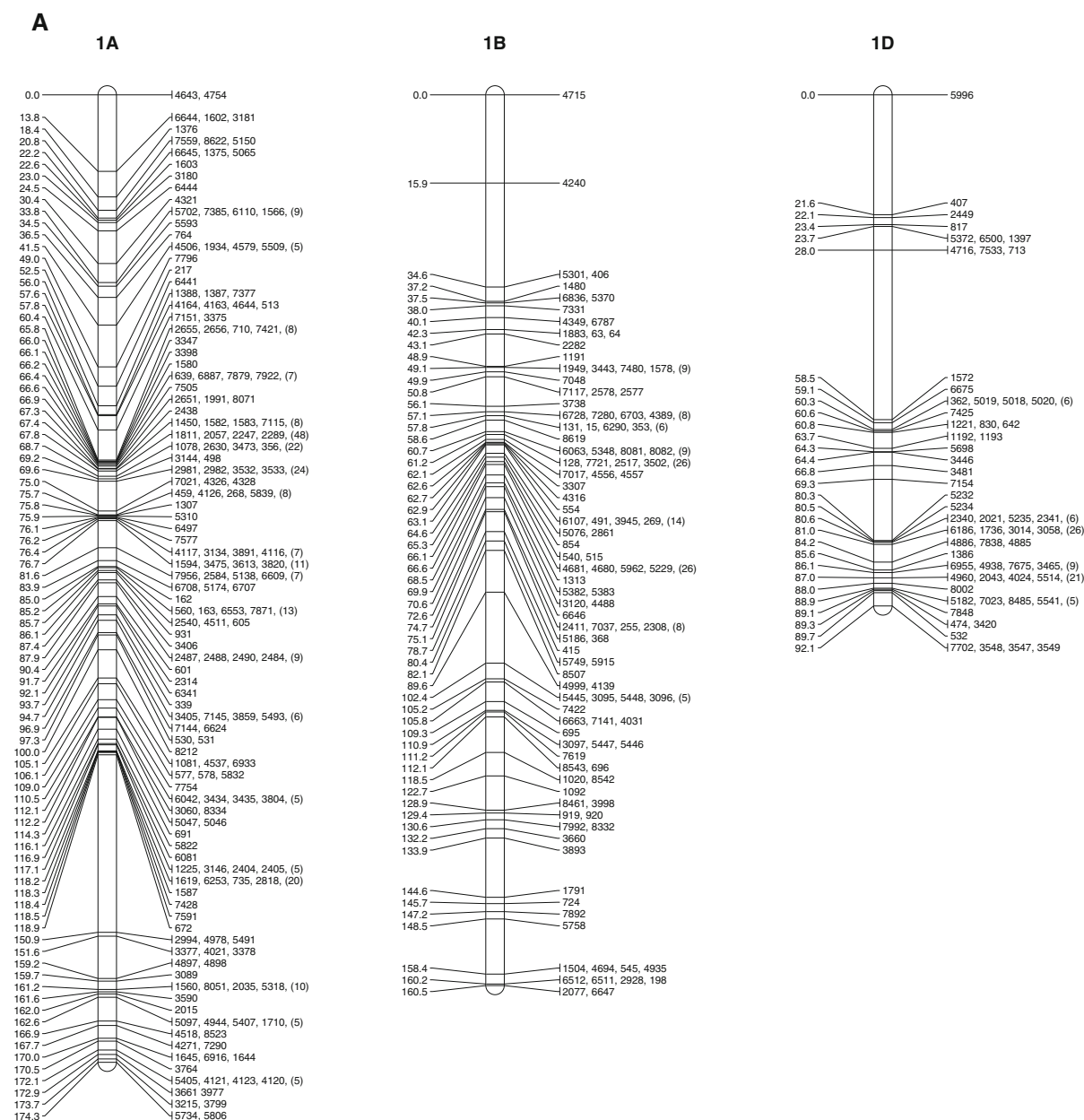
<sup>a</sup> The chromosome was represented by multiple linkage groups

markers were colocalized on the TCAP consensus map. Primary loci from each of the genetic bins provided approximately 437,000 of the uniquely informative data points, from which over 18,000 recombination events could be observed (Table 2).

Linkage maps estimated from the breeding populations are summarized in Table 2 and depicted in Fig. 2a–g. The estimated linkage maps included 3,875 loci and covered a total genetic distance of 3,080 cM, with an average interval of 2.5 cM between genetic bins. Marker coverage was relatively poor for the D genome. After curating the data to remove double-crossover events among consecutive marker trios, a total of 184 singletons remained within partially informative regions of 20 cM or less. This number of observed singletons represents a double-crossover rate of 0.04 % within a span of 20 cM, which is consistent with the expected maximum recombination rate ( $0.2^2 = 0.04$ ). Therefore, the majority singletons

remaining in the data were assumed to result from genuine recombination events.

Poor marker coverage on the D genome resulted in multiple linkage groups per chromosome for both the TCAP consensus map and the map estimated in this study. Therefore, only the A and B genomes were used for analysis of concordance with the consensus map. Linkage maps of the A and B genomes developed from the breeding populations in this study exhibited high concordance with the TCAP consensus maps (Fig. 3a–c). Excluding chromosomes 2B and 6B, the average rank-order correlation with the consensus maps was 0.98, indicating a high level of agreement regarding the locus ordering between the two sets of linkage maps. Chromosome 6B had the lowest rank-order correlation with the consensus map (0.52), due to a large centromeric inversion of the locus order (Fig. 3b). The linkage map for chromosome 2B had a region of highly suppressed centromeric recombination compared to

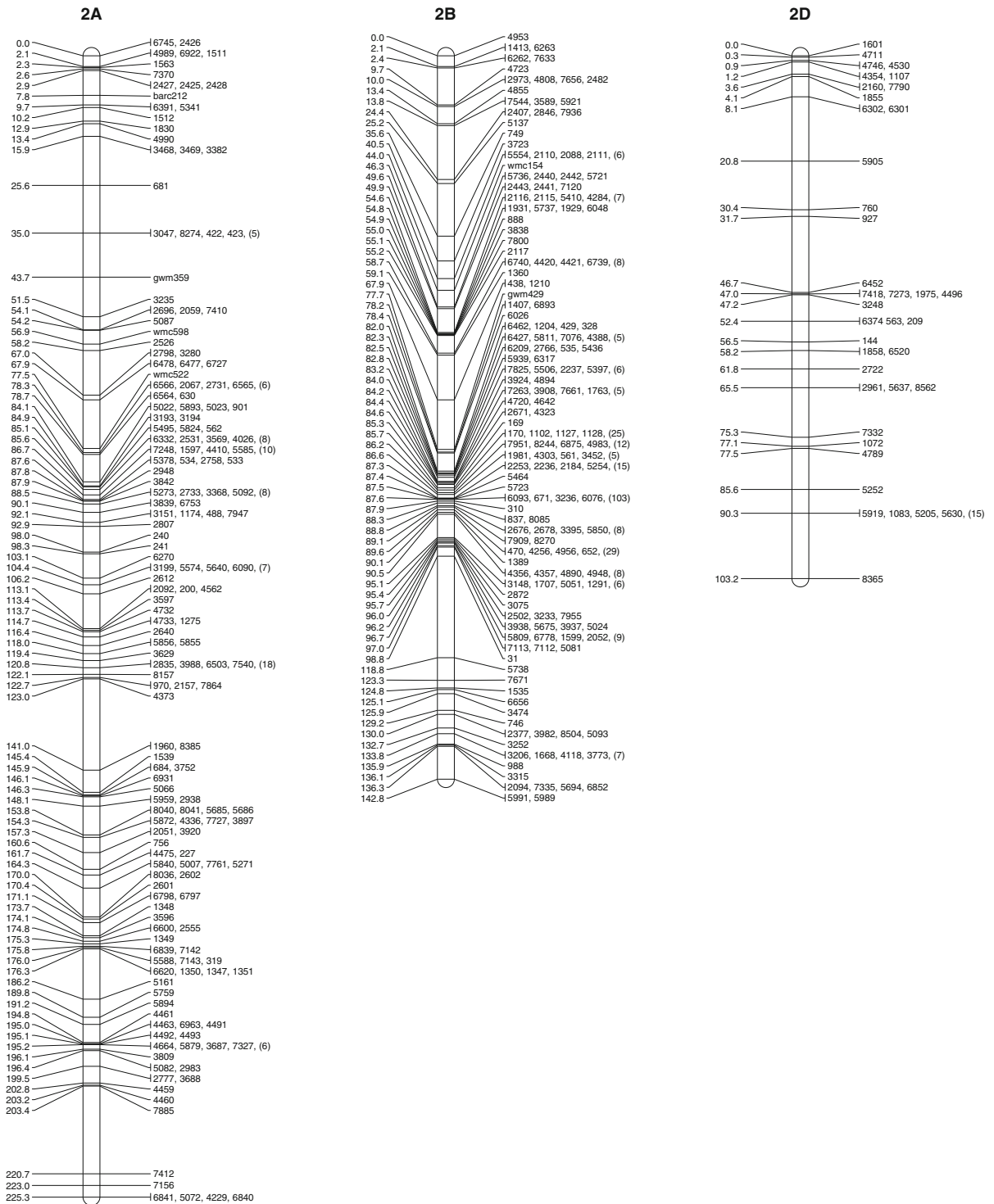


**Fig. 2** **a** Estimated genetic maps for homeologous group 1 chromosomes. Genetic bins are followed by the number of constituent markers in *parentheses* if  $>4$ . See supplemental data for a complete list. **b** Estimated genetic maps for homeologous group 2 chromosomes. Genetic bins are followed by the number of constituent markers in *parentheses* if  $>4$ . See supplemental data for a complete list. **c** Estimated genetic maps for homeologous group 3 chromosomes. Genetic bins are followed by the number of constituent markers in *parentheses* if  $>4$ . See supplemental data for a complete list. **d** Estimated genetic maps for homeologous group 4 chromosomes. Genetic bins are followed by the number of constituent markers in *parentheses* if

$>4$ . See supplemental data for a complete list. **e** Estimated genetic maps for homeologous group 5 chromosomes. Genetic bins are followed by the number of constituent markers in *parentheses* if  $>4$ . See supplemental data for a complete list. **f** Estimated genetic maps for homeologous group 6 chromosomes. Genetic bins are followed by the number of constituent markers in *parentheses* if  $>4$ . See supplemental data for a complete list. **g** Estimated genetic maps for homeologous group 7 chromosomes. Genetic bins are followed by the number of constituent markers in *parentheses* if  $>4$ . See supplemental data for a complete list



**B**



**Fig. 2** continued

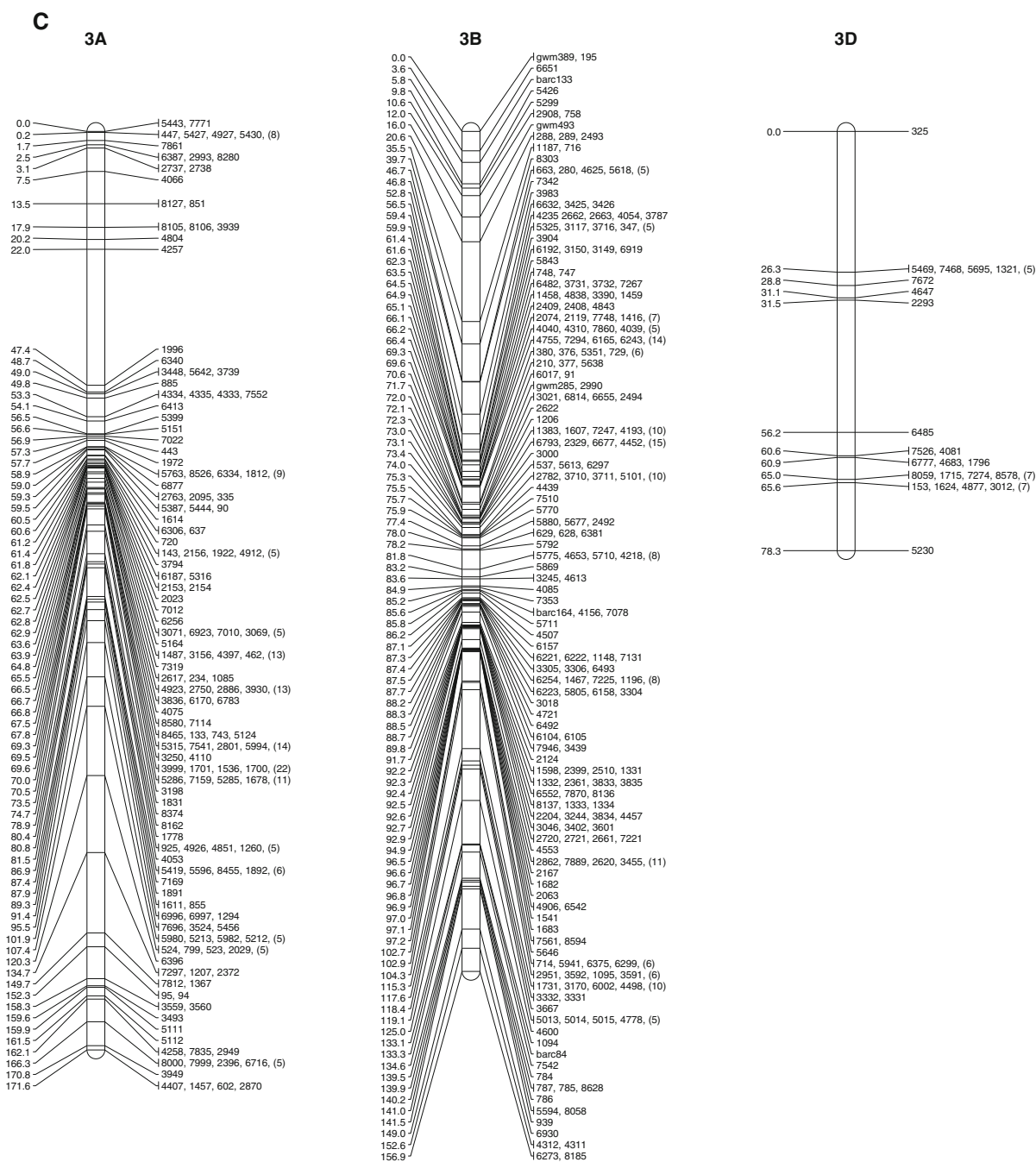


Fig. 2 continued

the consensus map, which resulted in shuffling of locus orders and a reduced rank-order correlation with the consensus map (0.72).

The majority of the linkage maps estimated from the breeding populations showed regions with reduced

recombination relative to the TCAP consensus maps. These lower estimates of recombination were typically centromeric, which resulted in “s-shaped” trends when plotted against the consensus map positions (Fig. 3a–c). However, the differences in estimated

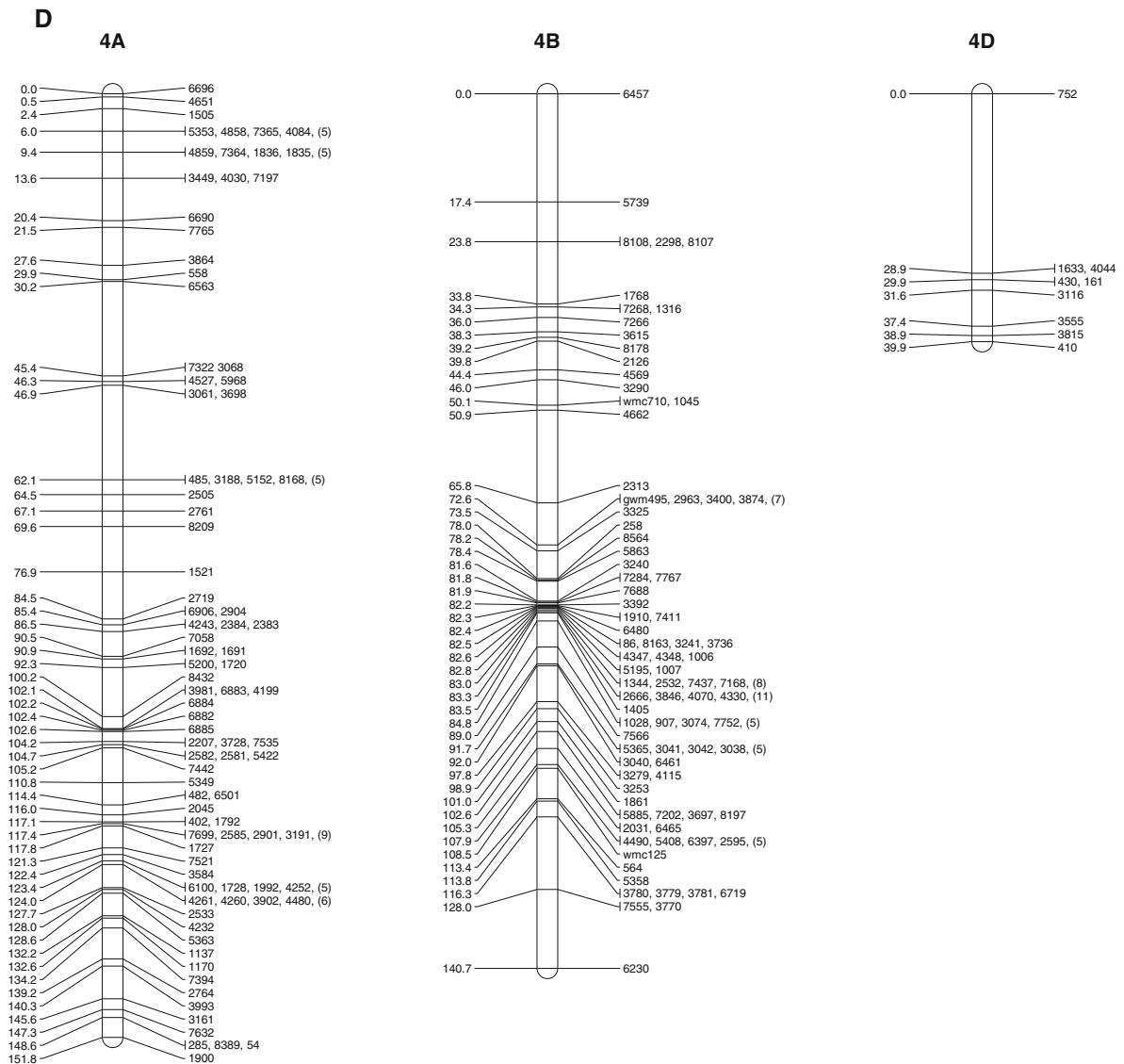


Fig. 2 continued

recombination frequencies did not greatly affect the relative ordering of the loci.

## Discussion

The application of mapping algorithms developed for general pedigrees to existing breeding populations facilitated the development of a 3,875 locus linkage map of the wheat genome without the need for extraneous population development. High-throughput SNP genotyping of 18 crosses comprised of a

total of 372 four-way  $F_1$  individuals enabled the mapping of over 43 % of the interrogated loci. These results indicate that a collection of breeding populations, derived from crosses among numerous parents, provide a highly polymorphic and informative genetic resource for the development of linkage maps. Comparatively, Cavanagh et al. (2013) were able to map roughly 46 % of the SNP loci from the same 9,000 SNP assay using a four-way cross MAGIC population consisting of 1,579 recombinant inbred lines. Biparental mapping populations used for consensus map individually enabled mapping of

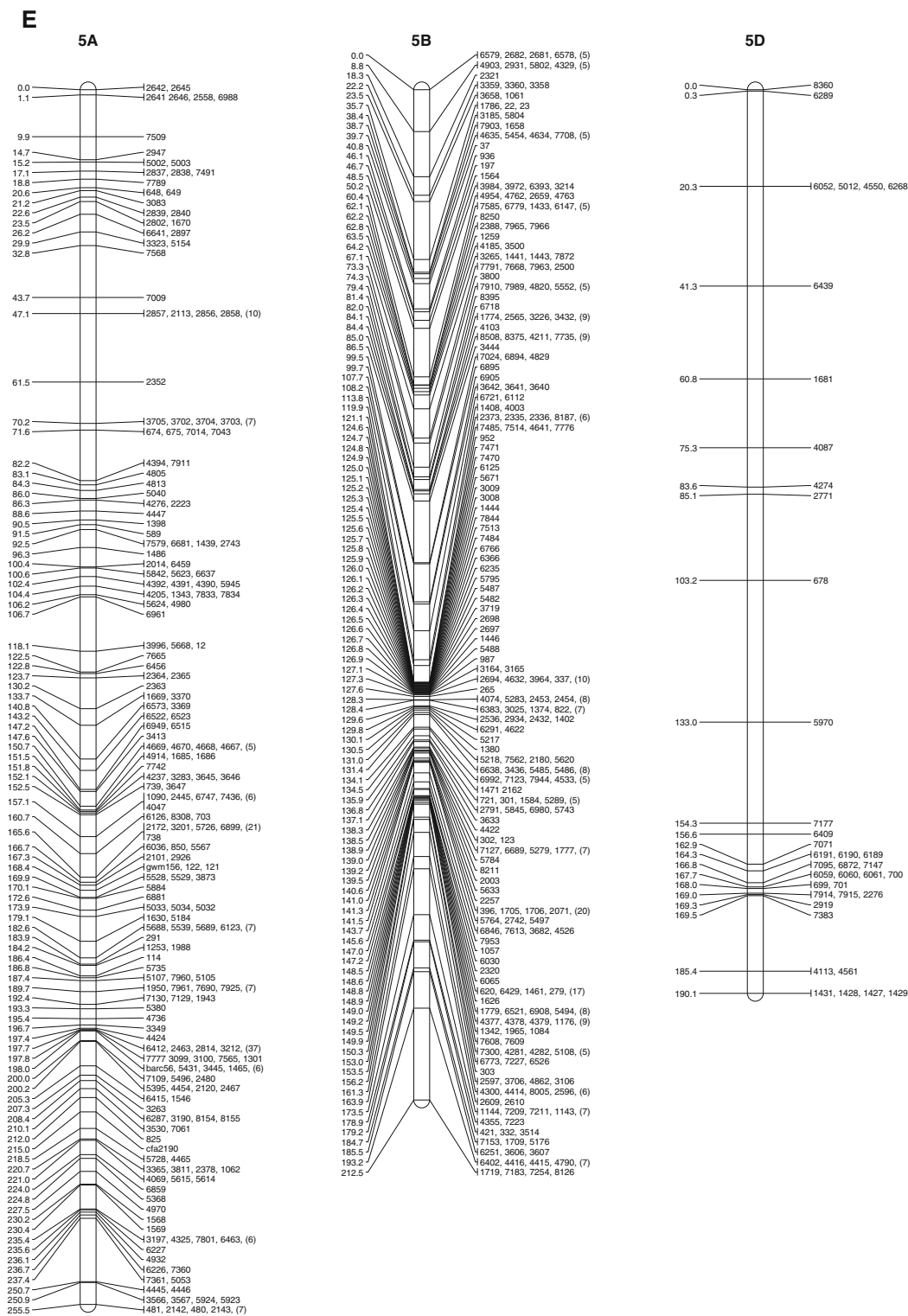
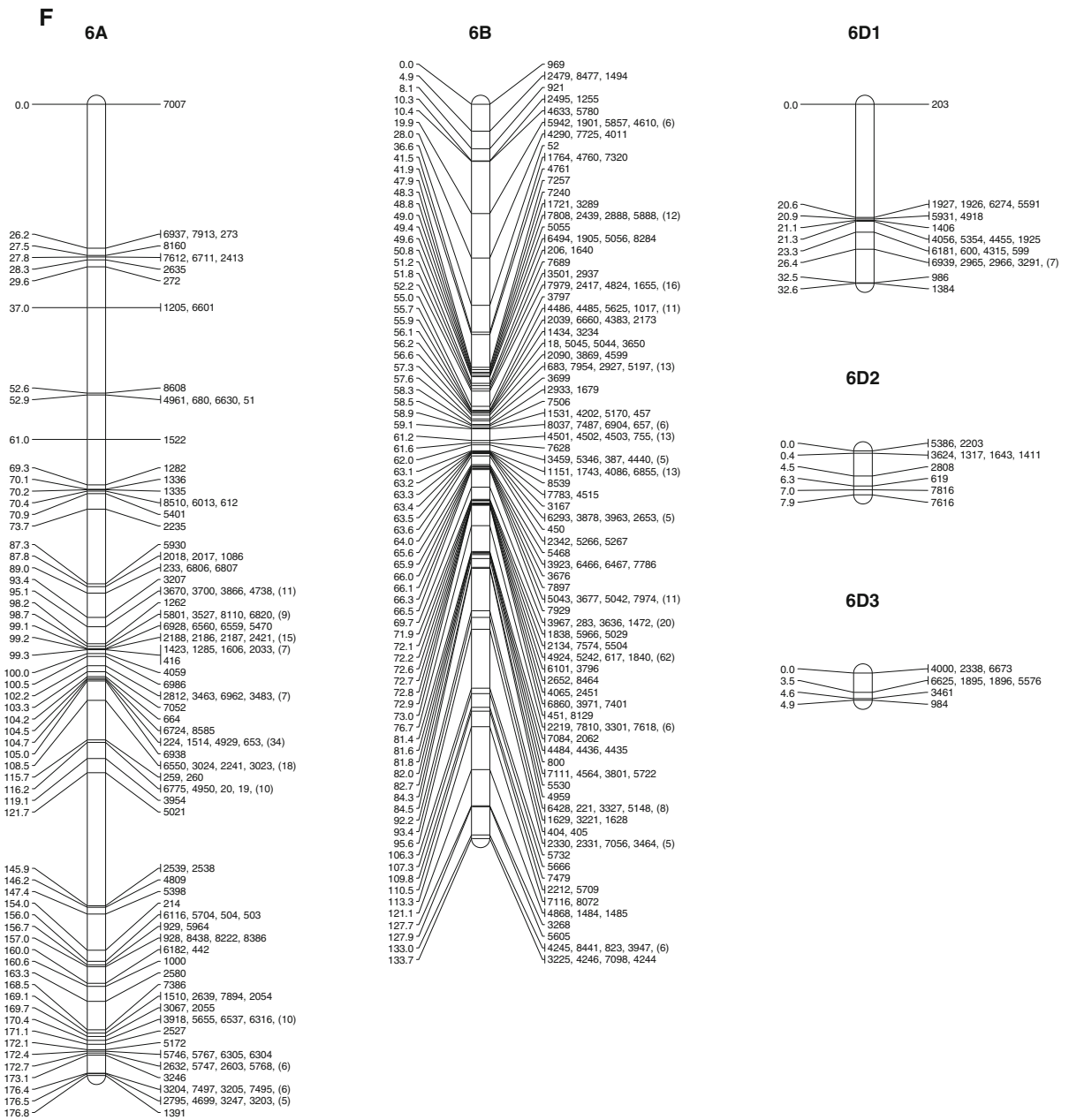


Fig. 2 continued



**Fig. 2** continued

21–41 % of the SNP loci with population sizes ranging from 96 to 250 fixed lines (Cavanagh et al. 2013). Given the time and resources required for the development of such purpose-built mapping populations, it seems clear that exploiting existing segregating populations in breeding programs provides a cost-efficient alternative for development of dense linkage maps. Furthermore, it should be noted that

the breeding populations used in this study were highly interrelated by common founders and the two Wesley-*Fhbl* backcross lines were nearly isogenic. A collection of breeding populations derived from a larger number of founders or from a more diverse set of founders would therefore be informative for more loci and thus allow for more extensive linkage mapping than possible in this study.

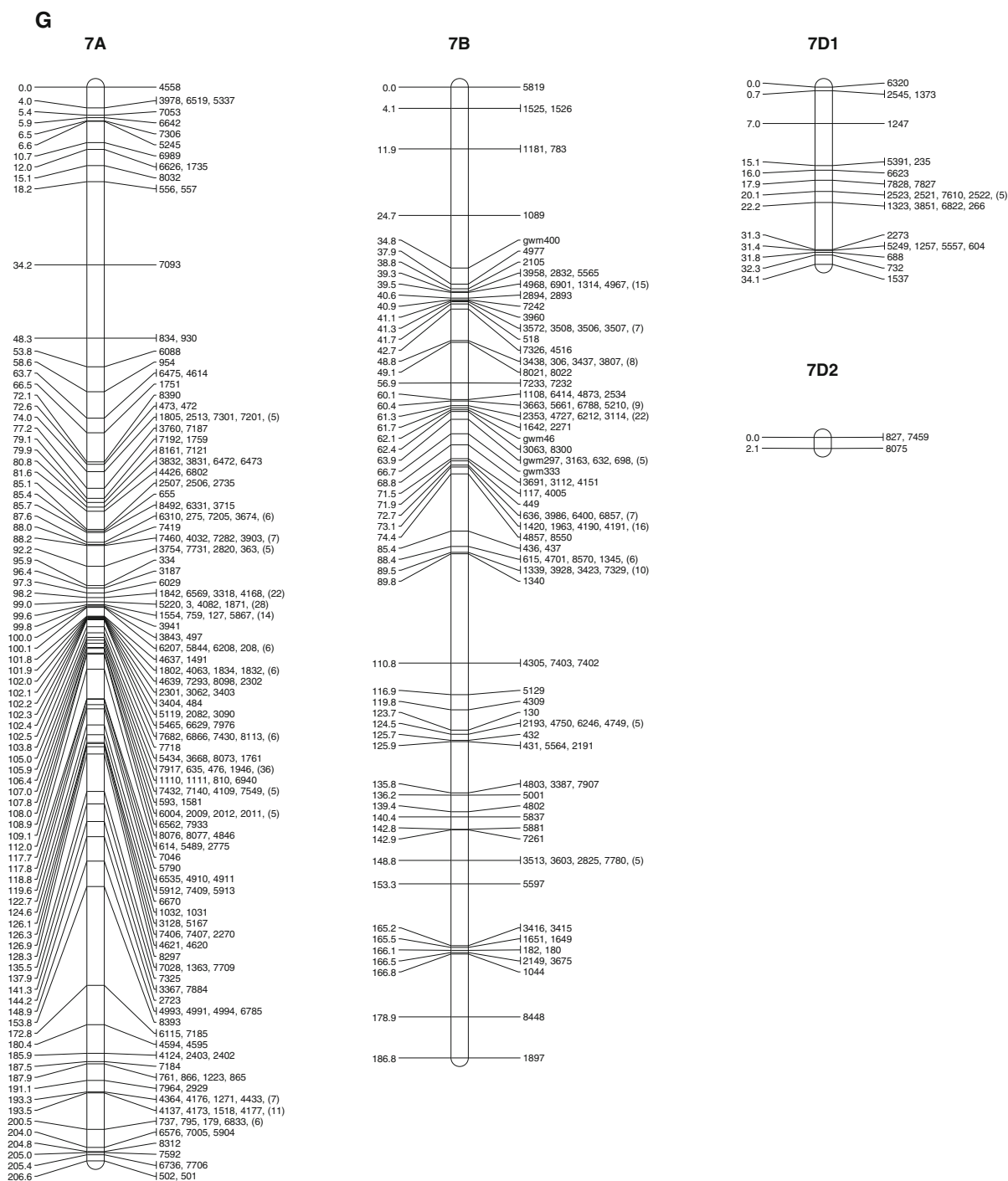
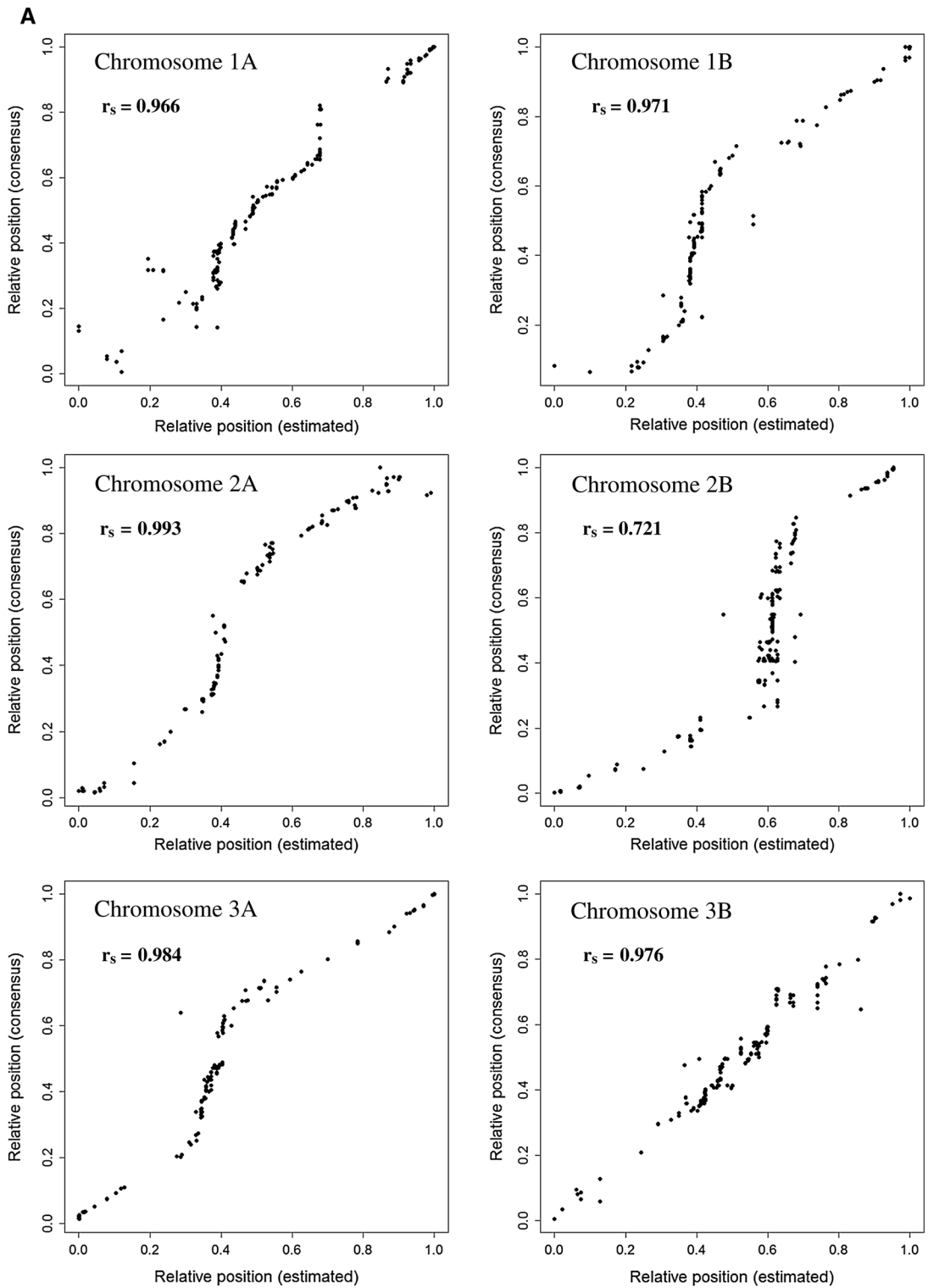


Fig. 2 continued

Correct ordering of marker loci during linkage mapping is a critical factor determining the accuracy and power of subsequent QTL mapping applications (Collard et al. 2009). Therefore, an assortment of

breeding populations must not only enable the mapping of a large number of marker loci, but also the accurate ordering of those loci from large genotypic data sets. Evaluation of rank-order correlations



**Fig. 3** a–c Comparison of genetic maps estimated for A and B genomes with the TCAP consensus maps

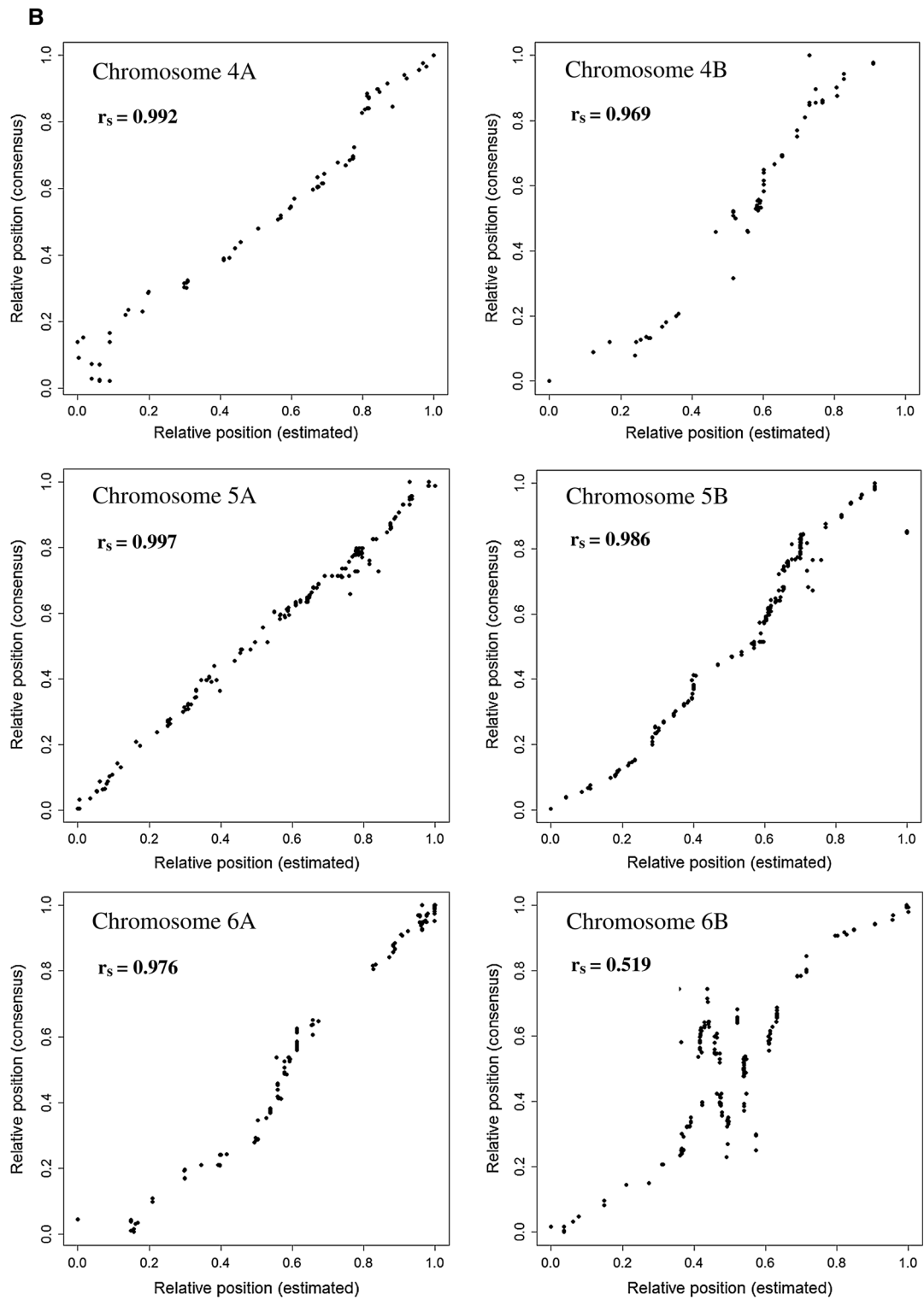
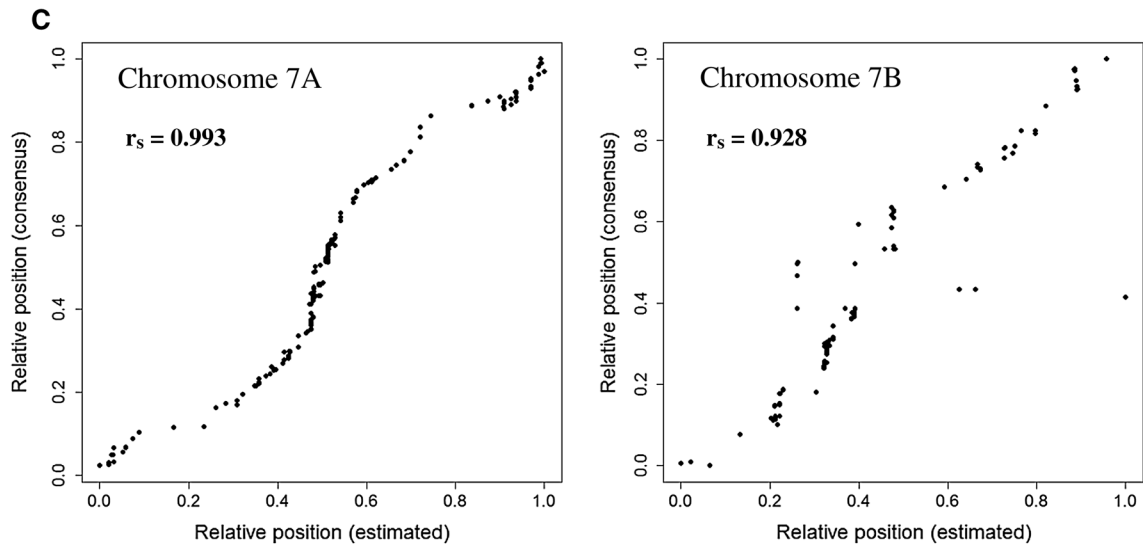


Fig. 3 continued





**Fig. 3** continued

between the linkage maps developed in this study and the extensive TCAP consensus map ( $>0.97$  for most chromosomes) revealed that marker orders were largely conserved. Local rearrangements of tightly linked markers ( $<5$  cM) relative to the consensus map are of little practical concern, since the marker density could be thinned for most QTL mapping applications (Collard et al. 2009). There were a few significant disparities, most notably the large inversion of locus order on chromosome 6B, that resulted in poor rank-order correlations with the consensus maps. Such disparities can result from true inversions, genotyping errors or convergence of the mapping algorithm to a local optimum during the map building process. Local optima can occur with any linkage mapping approach, since mapping algorithms can interrogate only a subset of potential locus orders, and thus may inadvertently discard the true map order if there is more statistical support for an alternative order at any stage in the mapping process (Green et al. 1990). Therefore, the discrepancies in locus orders observed with the TCAP consensus order are not necessarily indicative of issues specific to the mapping algorithms or populations employed in this study, but more likely reflect the effects of genotyping errors and initial map orders on convergence. Overall, the high degree of concordance with the consensus maps suggests that an amalgamation of relatively small breeding populations can provide reliable ordering of marker loci when those orders are jointly estimated over all populations.

While the ordering of loci in this study was generally consistent with the TCAP consensus map, there were notable differences in the estimated genetic distances between loci for many of the chromosomes. There were several large centromeric regions on the consensus maps that exhibited low rates of recombination in the current study (e.g., chromosomes 1B, 2B and 3A). Furthermore, considerably fewer marker loci could be mapped to unique genetic positions in this study compared to the consensus map. There are several possible explanations for these disparities in estimated genetic distances. First, the population of 372 four-way  $F_1$  individuals (565 for SSR markers) provided far fewer recombination events than the 2,294 recombinant inbred lines and 194 doubled haploid lines employed for development of the consensus map. Due to the smaller sample of recombination events in this study, fewer recombinations would be expected to be observed between tightly linked loci, thus resulting in the underestimation of genetic distances and the co-localization of tightly linked loci. Therefore, simply utilizing a larger number of breeding populations should reduce many of the discrepancies in the estimated recombination rates. Second, depending on allele frequencies and crossing schemes, markers can have a wide range of informativeness when interrogated over a large number of relatively small breeding populations (Fig. 1). Markers providing few informative meioses would present few opportunities to observe recombination

and would thus tend to co-localize with nearby marker loci. Lastly, in certain situations, a consensus mapping approach can overestimate genetic distances. For example, if duplicate loci are not jointly mapped in the same population, then those loci could map to unique positions when interpolating genetic distances among populations. The larger amount of data collected over the consensus mapping populations could also result in a greater number of genotyping errors, which can result in overestimated genetic distances (Goldstein et al. 1997).

The finding that marker loci can be accurately ordered by using a collection of existing breeding populations has important implications for molecular breeding efforts. Breeding programs can use this mapping approach to support molecular breeding efforts without expending resources on the development of purpose-built mapping populations. Newly developed marker resources can be directly applied to molecular breeding efforts by *de novo* linkage mapping in breeding populations rather than awaiting results of consensus mapping studies. This is an important consideration given the increasing interest in statistical methods that integrate QTL mapping efforts into plant breeding populations (Jansen et al. 2003; Crepieux et al. 2005; Rosyara et al. 2009; Würschum 2012). The divide between linkage mapping studies and marker assisted selection has been cited as a primary reason for the limited impact of marker assisted selection in crop breeding (Beavis 1998). The ability to develop reliable linkage maps directly in breeding populations represents an important step toward integrating these molecular breeding activates. The approach to linkage map development presented in this study should allow breeders to make the leap from having no prior genotypic information to conducting marker assisted selection within the same cohort of breeding populations, thus supporting a completely integrated molecular breeding approach.

Mapping algorithms for general pedigrees enable the construction of linkage maps using any number of disjoint populations of arbitrary size, structure and complexity. The four-way crosses used in this study resembled three generation pedigrees (grandparent, parent and offspring) commonly studied in human and animal populations, with the exception that the founders are completely inbred and thus generally provided phase-known haplotypes. However, the exclusive use of four-way crosses among inbred lines

in this study was completely arbitrary. The methods presented here can be equivalently applied to inbred and outbred mapping populations as well as to any assortment of crossing schemes. In fact, the mapping algorithms used in this study are routinely used for linkage map development in human and animal populations. Mapping algorithms typically used for plant populations have diverged from these generalized mapping algorithms as a result of the more simplistic population structures afforded by the availability of inbred founders and large single-cross populations. However, it has become apparent that these simplified populations, although analytically convenient, do not provide an accurate depiction of the complex genetic systems that exist within breeding programs. Therefore, this research serves as a reminder that the plant research community is not confined to the prevailing experimental paradigm for linkage analysis in plant species. Rather, a more generalized framework can be adopted to exploit the extant resources in breeding programs and facilitate the next generation of molecular breeding activities.

## Conclusions

Early generation plant breeding populations provide an existing source for genetic recombinations that can be used for the development of dense linkage maps. Application of a 9,000 SNP assay and 26 SSR markers to early generation breeding populations derived from 28 four-way crosses resulted in the development of a 3,875 locus linkage map of the wheat genome, with an average inter-marker distance of 2.5 cM. The developed linkage map had high rank-order concordance with an extensive TCAP consensus map developed from the same SNP assay, indicating that marker ordering was quite accurate. The mapping algorithms adopted from general pedigree analysis can handle any number of related or unrelated populations of arbitrary size, structure and complexity. Therefore, in combination with high-throughput genotyping platforms, the mapping approach used in this study should allow breeders to develop dense linkage maps “on the fly” to support their molecular breeding efforts.

**Acknowledgments** The authors acknowledge support from the US Wheat and Barley Scab Initiative under ARS Agreement No: 59-0200-3-005 to J.L.G.H. and by the South Dakota Agricultural Experimental Station.

## References

- Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Hettner E, Kilian A (2006) Diversity arrays technology (DART) for high-throughput profiling of the hexaploid wheat genome. *Theor Appl Genet* 113(8):1409–1420
- Akhunov E, Chao S, Catana V, See D, Brown-Guedira G, Akhunova A, Dubcovsky J, Cavanagh C, Hayden M (2011) New tools for wheat genetics and breeding: genome-wide analysis of SNP variation. In: McIntosh R (ed) Proceedings of the BGRI 2011 tech. work. Borlaug Global Rust Initiative, St. Paul, MN, pp 92–97
- Allen AM, Barker GLA, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brechley RC, D'Amore R, McKenzie N, Waite D, Hall A, Bevan M, Hall N, Edwards KJ (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotechnol J* 9(9):1086–1099
- Bandillo N, Raghavan C, Muyco P, Sevilla MAL, Lobina IT, Dilla-Ermita CJ, Tung C-W, McCouch S, Thomson M, Mauleon R, Singh RK, Gregorio G, Rodoña E, Leung H (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6:11
- Beavis WD (1998) QTL analyses: power, precision, and accuracy. In: Paterson AH (ed) Molecular dissection of complex trait. CRC Press, Boca Raton, pp 145–162
- Bowers JE, Bachlava E, Brunick RL, Rieseberg LH, Knapp SJ, Burke JM (2012) Development of a 10,000 locus genetic map of the sunflower genome based on multiple crosses. *G3* 2(7):721–729
- Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol* 11(2):215–221
- Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kianian S, Forrest K, Sainetnac C, Brown-Guedira GL, Akhunova A, See D, Bai G, Pumphrey M, Tomar L, Wong D, Kong S, Reynolds M, Lopez da Silva M, Bockelman H, Talber L, Anderson J, Dreisigacker S, Baenziger S, Carter A, Korzun V, Morrell PL, Dubcovsky J, Morell M, Sorrells ME, Hayden MJ, Akhunov E (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc Natl Acad Sci USA* 110(20):8057–8062
- Cheema J, Dicks J (2009) Computational approaches and software tools for genetic linkage map estimation in plants. *Brief Bioinform* 10(6):595–608
- Collard B, Mace E, McPhail M et al (2009) How accurate are the marker orders in crop linkage maps generated from large marker datasets? *Crop Pasure Sci* 52:362–372
- Crepieux S, Lebreton C, Servin B, Charvet G (2004) Quantitative trait loci (QTL) detection in multicross inbred designs: recovering QTL identical-by-descent status information from marker data. *Genetics* 168(3):1737–1749
- Crepieux S, Lebreton C, Flament P, Charvet G (2005) Application of a new IBD-based QTL mapping method to common wheat breeding population: analysis of kernel hardness and dough strength. *Theor Appl Genet* 111(7):1409–1419
- Deschamps S, Llaca V, May GD (2012) Genotyping-by-sequencing in plants. *Biology (Basel)* 1(3):460–483
- Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet* 3(1):43–52
- Goldstein DR, Zhao H, Speed TP (1997) The effects of genotyping errors and interference on estimation of genetic distance. *Hum Hered* 47:86–100
- Green P, Falls K, Crooks S (1990) Documentation for CRIMAP, version 2.4. [www.animalgenome.org](http://www.animalgenome.org). (Accessed 7 Jan 2014)
- Holland JB (2007) Genetic architecture of complex traits in plants. *Curr Opin Plant Biol* 10(2):156–161
- Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK, Cavanagh CR (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol J* 10(7):826–839
- Jansen RC, Jannink JL, Beavis WD (2003) Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. *Crop Sci* 43(3):829–834
- Karakousis A, Langridge P (2003) A high-throughput plant DNA extraction method for marker analysis. *Plant Mol Biol Rep* 21(1):95
- Kilian A, Huttner E, Wenzl P, Jaccoud D, Carling J, Caig V, Evers M, Heller-Uszynska K, Uszynski G, Cayla C, Patarapuwadol S, Xia L, Yang S, Thomson B (2003) The fast and the cheap: SNP and DART-based whole genome profiling for crop improvement. In: Proceedings of the international congress in wake double helix from green revolution to gene revolution, Avenue Media, Bologna, Italy, pp 443–461
- Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci USA* 84(8):2363–2367
- Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. *Nat Rev Genet* 13(2):85–96
- Qin H, Guo W, Zhang Y-M, Zhang T (2008) QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum* L. *Theor Appl Genet* 117(6):883–894
- Rakshit S, Rakshit A, Patil JV (2012) Multiparent intercross populations in analysis of quantitative traits. *J Genet* 91(1):111–117
- Rosyara UR, Gonzalez-Hernandez JL, Glover KD, Gedye KR, Stein JM (2009) Family-based mapping of quantitative trait loci in plant breeding populations with resistance to Fusarium head blight in wheat as an illustration. *Theor Appl Genet* 118(8):1617–1631
- Trebbi D, Maccaferri M, Giuliani S, Sorensen A, Sanguineti MC, Massi A, Tuberosa R (2008) Development of a multiparental (four-way cross) mapping population for multi-allelic QTL analysis in durum wheat. In: Appels R, Eastwood R, Lagudah E et al (eds) 11th proceedings of the international wheat genetics symposium. Sydney University Press, Brisbane, pp 14–15
- van Ooijen JW (2006) JoinMap<sup>®</sup> 4.0: software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen

- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93(1):77–78
- Weaver R, Helms C, Mishra SK, Donis-Keller H (1992) Software for analysis and manipulation of genetic linkage data. *Am J Hum Genet* 50(6):1267–1274
- Wu Y, Close TJ, Lonardi S (2008) On the accurate construction of consensus genetic maps. *Comput Syst Bioinform Conf* 7:285–296
- Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. *Theor Appl Genet* 125(2):201–210
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178(1):539–551
- Zhang X, Bai G, Bockus W, Ji X, Pan H (2012) Quantitative trait loci for fusarium head blight resistance in U.S. Hard winter wheat cultivar Heyne. *Crop Sci* 52(3):1187–1194