

## Analysis of Expressed Sequence Tag Loci on Wheat Chromosome Group 4

Miftahudin,<sup>\*,†</sup> K. Ross,<sup>†</sup> X.-F. Ma,<sup>\*</sup> A. A. Mahmoud,<sup>\*</sup> J. Layton,<sup>\*</sup> M. A. Rodriguez Milla,<sup>\*</sup>  
T. Chikmawati,<sup>\*,†</sup> J. Ramalingam,<sup>\*</sup> O. Feril,<sup>\*</sup> M. S. Pathan,<sup>\*</sup> G. Surlan Momirovic,<sup>\*,§</sup>  
S. Kim,<sup>\*</sup> K. Chema,<sup>\*</sup> P. Fang,<sup>\*</sup> L. Haule,<sup>\*</sup> H. Struxness,<sup>\*</sup> J. Birkes,<sup>\*</sup> C. Yaghoubian,<sup>\*</sup>  
R. Skinner,<sup>\*</sup> J. McAllister,<sup>\*</sup> V. Nguyen,<sup>\*</sup> L. L. Qi,<sup>1</sup> B. Echalié,<sup>1</sup> B. S. Gill,<sup>1</sup>  
A. M. Linkiewicz,<sup>\*,1</sup> J. Dubcovsky,<sup>\*,&</sup> E. D. Akhunov,<sup>\*,&</sup> J. Dvořák,<sup>\*,&</sup> M. Dilbirligi,<sup>\*,‡</sup>  
K. S. Gill,<sup>\*\*,‡</sup> J. H. Peng,<sup>††</sup> N. L. V. Lapitan,<sup>††</sup> C. E. Bermudez-Kandianis,<sup>††</sup>  
M. E. Sorrells,<sup>††</sup> K. G. Hossain,<sup>§§</sup> V. Kalavacharla,<sup>§§</sup> S. F. Kianian,<sup>§§</sup>  
G. R. Lazo,<sup>¶¶</sup> S. Chao,<sup>¶¶,2</sup> O. D. Anderson,<sup>¶¶</sup> J. Gonzalez-Hernandez,<sup>\*\*\*,3</sup>  
E. J. Conley,<sup>\*\*\*</sup> J. A. Anderson,<sup>\*\*\*</sup> D.-W. Choi,<sup>†††,4</sup> R. D. Fenton,<sup>†††</sup>  
T. J. Close,<sup>†††</sup> P. E. McGuire,<sup>\*,&</sup> C. O. Qualset,<sup>\*,&</sup>  
H. T. Nguyen<sup>\*</sup> and J. P. Gustafson<sup>†,5</sup>

<sup>\*</sup>Department of Agronomy, University of Missouri, Columbia, Missouri 65211, <sup>†</sup>Department of Biology, Bogor Agricultural University, Bogor 16144, Indonesia, <sup>‡</sup>USDA-ARS Plant Genetics Research Unit, Department of Agronomy, University of Missouri, Columbia, Missouri 65211, <sup>§</sup>Department of Plant Genetics and Breeding, University of Belgrade, 11080 Belgrade, Yugoslavia, <sup>1</sup>Department of Plant Pathology, Wheat Genetics Resource Center, Kansas State University, Manhattan, Kansas 66506-5502, <sup>2</sup>Department of Agronomy and Range Science, University of California, Davis, California 95616, <sup>\*\*</sup>Department of Crop and Soil Sciences, Washington State University, Pullman, Washington 99164-6420, <sup>††</sup>Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado 80523-1170, <sup>†††</sup>Department of Plant Breeding, Cornell University, Ithaca, New York 14853, <sup>§§</sup>Department of Plant Sciences, North Dakota State University, Fargo, North Dakota 58105, <sup>¶¶</sup>USDA-ARS Western Regional Research Center, Albany, California 94710-1105, <sup>\*,&</sup>Genetic Resources Conservation Program, University of California, Davis, California 95616, <sup>\*\*\*</sup>Department of Agronomy and Plant Genetics, University of Minnesota, St Paul, Minnesota 55108 and <sup>†††</sup>Department of Botany and Plant Sciences, University of California, Riverside, California 92521

Manuscript received December 5, 2003

Accepted for publication June 1, 2004

### ABSTRACT

A total of 1918 loci, detected by the hybridization of 938 expressed sequence tag unigenes (ESTs) from 26 Triticeae cDNA libraries, were mapped to wheat (*Triticum aestivum* L.) homoeologous group 4 chromosomes using a set of deletion, ditelosomic, and nulli-tetrasomic lines. The 1918 EST loci were not distributed uniformly among the three group 4 chromosomes; 41, 28, and 31% mapped to chromosomes 4A, 4B, and 4D, respectively. This pattern is in contrast to the cumulative results of EST mapping in all homoeologous groups, as reported elsewhere, that found the highest proportion of loci mapped to the B genome. Sixty-five percent of these 1918 loci mapped to the long arms of homoeologous group 4 chromosomes, while 35% mapped to the short arms. The distal regions of chromosome arms showed higher numbers of loci than the proximal regions, with the exception of 4DL. This study confirmed the complex structure of chromosome 4A that contains two reciprocal translocations and two inversions, previously identified. An additional inversion in the centromeric region of 4A was revealed. A consensus map for homoeologous group 4 was developed from 119 ESTs unique to group 4. Forty-nine percent of these ESTs were found to be homologous to sequences on rice chromosome 3, 12% had matches with sequences on other rice chromosomes, and 39% had no matches with rice sequences at all. Limited homology (only 26 of the 119 consensus ESTs) was found between wheat ESTs on homoeologous group 4 and the Arabidopsis genome. Forty-two percent of the homoeologous group 4 ESTs could be classified into functional categories on the basis of blastX searches against all protein databases.

**G**ENOME analysis has been used to establish the evolutionary and homoeologous relationships of the three genomes (AA, BB, and DD) that make up

hexaploid wheat (*Triticum aestivum* L.). Each of the 21 chromosomes has been identified and characterized by SEARS (1954, 1966) with respect to genomic and homoeologous relationships. There is a high degree of colin-

<sup>1</sup>Present address: Plant Breeding and Acclimatization Institute, Radzikow 05-870 Blonie, Poland.

<sup>2</sup>Present address: USDA-ARS Biosciences Research Laboratory, Fargo, ND 58105-5674.

<sup>3</sup>Present address: Department of Plant Sciences, North Dakota State University, Fargo, ND 58105-5051.

<sup>4</sup>Present address: Eugentech, 52 Oun-Dong, Yusong, Taeson, 305-333, Republic of Korea.

<sup>5</sup>Corresponding author: USDA-ARS, 206 Curtis Hall, University of Missouri, Columbia, MO 65211. E-mail: pgus@missouri.edu

earity among hexaploid wheat chromosomes within each of the seven homoeologous groups (VAN DEYNZE *et al.* 1995). This relationship among homoeologous chromosomes suggests that most genes, especially those coding for morphological characteristics and molecular markers, may be found in all three genomes in a well-conserved order (VAN DEYNZE *et al.* 1995).

The syntenic relationship among the three genomes composing hexaploid wheat has been altered by inter- and intragroup translocations, inversions, and deletions (AKHUNOV *et al.* 2003a). A previous study has shown that synteny tends to decrease in the more distal regions of the chromosome arms (AKHUNOV *et al.* 2003b). Several of the 21 chromosomes of hexaploid wheat contain translocations of considerable size (GALE 1990). NARANJO *et al.* (1987) proposed translocations involving chromosomes 4A, 5A, and 7B, which have been found in *T. turgidum* L. and *T. aestivum*. Using cytogenetic, isozyme, RFLP marker, and *in situ* hybridization analyses of chromosome 4A, the evolutionary evidence for translocations involving chromosome arms 4AL, 5AL, and 7BS has been firmly established (CHAO *et al.* 1989; NARANJO 1990; LIU *et al.* 1992; CHEN and GUSTAFSON 1994, 1997; DEVOS *et al.* 1995; MICKELSON-YOUNG *et al.* 1995; NELSON *et al.* 1995). A pericentric inversion within chromosome 4A changed the native chromosome arm homoeologues of group 4. As a result, the short arm of 4A (4AS) is homoeologous to 4BL and 4DL, while 4AL is homoeologous to 4BS and 4DS. In addition, a paracentric inversion within 4AL was also reported (DEVOS *et al.* 1995; MICKELSON-YOUNG *et al.* 1995). Finally, a pericentric inversion has been observed within chromosome 4B (ENDO and GILL 1984; GILL *et al.* 1991; FRIEBE *et al.* 1994; MICKELSON-YOUNG *et al.* 1995).

It is reasonable to expect that undetected chromosome structural variations exist in hexaploid wheat. Detailed analyses at the molecular level will reveal the evolution of chromosome structure and the distribution and physical location of additional breakpoints and rearrangements within homoeologous group 4 chromosomes of hexaploid wheat.

ENDO (1988, 1990) ascertained that certain *Aegilops* chromosomes, particularly from *Aegilops cylindrica* host, when present as monosomic additions to wheat, caused random, mostly terminal, deletions in wheat chromosomes. Broken chromosome ends tend to be healed by the addition of new telomeres (WERNER *et al.* 1992; TSUJIMOTO 1993; FRIEBE *et al.* 2001). ENDO and GILL (1996) isolated some 350 homozygous deletion lines derived from *T. aestivum* cv. Chinese Spring. Adjacent deletion breakpoints for a given chromosome arm define a physical region on that arm called a bin. A set of lines containing deletions for each of the three genomes and seven groups of hexaploid wheat chromosomes has been used to create a deletion map of the wheat genomes.

A project funded by the National Science Foundation

(NSF) proposed to identify and characterize expressed sequence tag unigenes (ESTs) and to map them to specific chromosome regions (bins) using a representative subset of deletion, nullisomic-tetrasomic, and ditelosomic lines (<http://wheat.pw.usda.gov/NSF>). QI *et al.* (2003) reported a molecular characterization of the deletion stocks used in this project. The present study examines the ESTs that mapped to group 4 chromosomes, to detect homoeologous chromosome structural variations, study the spatial relationships among these ESTs at the bin level, and investigate homologous relationships with rice (*Oryza sativa* L.) and *Arabidopsis thaliana* (L.) Heynh.

## MATERIALS AND METHODS

**Plant materials:** The aneuploid and deletion stocks used in this analysis were developed in or derived from Chinese Spring wheat and included 21 nullisomic-tetrasomic (NT) lines and 24 ditelocentric lines obtained from the U.S. Department of Agriculture (USDA)-Sears collection of wheat genetic stocks [USDA-Agricultural Research Service (ARS), University of Missouri, Columbia, MO] and 101 deletion stocks obtained from B. S. Gill (Wheat Genetics Resource Center, Kansas State University).

The stocks relevant to homoeologous group 4 defined centromeric regions and nine, seven, and eight deletion bins in chromosomes 4A, 4B, and 4D, respectively (Figure 1). Deletion breakpoint and bin designations were based on the fractional length (FL) values of breakpoints established by ENDO and GILL (1996). The proximal bin in each arm was delimited by the most proximal deletion breakpoint and the centromere of the relevant Chinese Spring nulli-tetrasomic and ditelosomic aneuploid stocks.

**Library materials:** Twenty-six of the project's 42 cDNA libraries (LAZO *et al.* 2004; ZHANG *et al.* 2004), representing a variety of tissues, developmental stages, and stress treatments of wheat and related species, produced ESTs mapping to homoeologous group 4 chromosomes. Twenty-four of these libraries were used for the functional analysis reported here. See the project website ([http://wheat.pw.usda.gov/cgi-bin/nsf/nsf\\_library.cgi](http://wheat.pw.usda.gov/cgi-bin/nsf/nsf_library.cgi)) for details on library identity, development, and characterization.

**Deletion mapping:** All procedures used for genomic DNA isolation from the Chinese Spring deletion and aneuploid lines, restriction endonuclease digestions, gel electrophoresis, DNA gel blot hybridization, and EST analyses were uniformly carried out in 10 mapping laboratories (AKHUNOV *et al.* 2003a,b; LAZO *et al.* 2004). Procedures are available on line at <http://wheat.pw.usda.gov/NSF/project/>. For Southern analysis, genomic DNA from each aneuploid and deletion stock was digested with *EcoRI*. Lambda DNA, digested with *HindIII* and *BstEII*, was used as a size marker.

Deletion mapping was performed by hybridizing a cDNA clone, corresponding to an EST, to a set of Southern blots made from the panel of wheat aneuploid and deletion stocks described above (QI *et al.* 2003; LAZO *et al.* 2004). Absence of a particular restriction fragment in the lane for a genetic stock indicated that the locus was distal to the corresponding deletion breakpoint or within the chromosome or arm missing in the aneuploid. The analysis presented here was carried out on the subset of 4485 mapped and verified cDNA probes accumulated as of March 17, 2003. Each hybridization profile was analyzed twice in the mapping laboratory where it was

TABLE 1  
 $\chi^2$  analysis of the total and duplicated EST loci within homoeologous group 4 chromosomes

Chromosome/arm	All EST loci			Duplicated EST loci		
	Observed loci	Expected loci <sup>a</sup>	<i>P</i> -value <sup>b</sup>	Observed loci	Expected loci <sup>a</sup>	<i>P</i> -value <sup>b</sup>
4A	786	639.3		268	210.3	
4B	529	639.3	<0.001	157	210.3	<0.001
4D	603	639.3		206	210.3	
Total	1918			631		
4AS	246	291.1	<0.001	63	99.3	<0.001
4AL	540	494.9		205	168.7	
4BS	181	251.9	<0.001	47	74.8	<0.001
4BL	348	277.1		110	82.2	
4DS	238	215.3	>0.05	96	73.6	0.001
4DL	365	387.7		110	132.4	

<sup>a</sup> The expected loci for each chromosome were based on the assumption of equal EST content among the genomes. The expected loci for chromosome arms were based on the arm ratio (GILL *et al.* 1991).

<sup>b</sup> Chi-square test indicates the probability of a departure from the expected loci.

produced. These images were then uploaded to the NSF website and scored again by a homoeologous group coordinator. Once a consensus was reached, the mapping data were made available to the public ([http://wheat.pw.usda.gov/cgi-bin/westsql/map\\_locus.cgi](http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi)) and used for further analysis.

**Locus distribution and duplication analysis:** Duplication of an EST locus means that it was present in more than one copy either within a single chromosome (an intrachromosomal duplication) or among the three homoeologous group 4 chromosomes (an interchromosomal duplication). The number and distribution of intra- and interchromosomal duplicated EST loci within homoeologous chromosome group 4 were established using only loci clearly assigned to a specific bin. Loci were excluded from analysis if they could not be mapped to a specific bin but were mapped only to a chromosome, chromosome arm, the centromere, or an ambiguous bin whose FL values overlapped with adjacent bins. The expected numbers of EST loci for the chromosome arms were calculated using the arm ratio values for 4A, 4B, and 4D from the physical measurements of mitotic chromosomes (GILL *et al.* 1991). This density per arm was further broken down into the expected numbers of EST loci for each deletion bin located on that particular arm. The physical size of the deletion bins was based on the FL values of each deletion bin for that particular arm. A chi-square test with  $P < 0.01$  was used to test for significance between observed and expected numbers of mapped loci. The expected values for whole chromosomes were based on the assumption that the homoeologous chromosomes contain equal numbers of loci, whereas the expected values for chromosome arms were based on arm ratios (GILL *et al.* 1991; Table 1).

**Group 4 consensus:** A consensus map of group 4 chromosomes was constructed on the basis of 119 ESTs having loci that mapped to two or three of the homoeologous group 4 chromosomes. ESTs were assigned to consensus bins that were constructed from combined FL values across the three chromosomes, starting from the bins closest to the centromeres. For the consensus chromosome short arm, the bins are C-4S1-0.37, 4S1-0.37-0.43, 4S2-0.43-0.53, 4S3-0.53-0.57, 4S4-0.57-0.66, 4S5-0.66-0.67, 4S6-0.67-0.82, and 4S7-0.82-1.00. For the long arm, they are C-4L1-0.20, 4L1-0.20-0.31, 4L2-0.31-0.56, 4L3-0.56-0.63, 4L4-0.63-0.71, 4L5-0.71-0.76, 4L6-0.76-0.86, and 4L7-

0.86-1.00. For example, wheat EST BE497446 mapped to deletion bins 4AL12-0.43-0.59, 4BS8-0.57-0.81, and 4DS3-0.67-0.82, indicating a consensus map position between fractional lengths 0.57 and 0.66 on the short arm of the consensus chromosome.

Loci that mapped to chromosomes outside of group 4 were not taken into consideration in the construction of the consensus map. The ESTs were sorted according to their bin-mapping pattern; for example, BM138075 mapped to 4AS3-0.76-1.00, C-4BL1-0.71, and 4DL9-0.31-0.56, which is different from the mapping pattern of BG274627, which mapped to 4AS3-0.76-1.00, C-4BL1-0.71, and 4DL12-0.71-1.00. All ESTs showing a particular mapping pattern were grouped together and were assumed to be located close to one another on the chromosome; however, linear order within the mapping-pattern groups could not be determined. A mapping-pattern group was included in the consensus map only if it contained three or more ESTs. The mapping-pattern groups were arranged in linear order on the basis of the known homoeology of group 4 chromosomes reported in publications: the homoeology of 4AS to 4BL to 4DL, the homoeology of 4AL to 4BS to 4DS (DEVOS *et al.* 1995; GALE *et al.* 1995; MICKELSON-YOUNG *et al.* 1995), and the structural rearrangement of chromosome 4A (NARANJO *et al.* 1987; ANDERSON *et al.* 1992; LIU *et al.* 1992; DEVOS *et al.* 1995; GALE *et al.* 1995; MICKELSON-YOUNG *et al.* 1995; NELSON *et al.* 1995; CHEN and GUSTAFSON 1997). The mapping-pattern groups were organized to maximize the integrity of these trends. Groups with patterns that deviated from the trend and could not be reconciled by reordering the groups within a bin were considered signals of undefined chromosomal rearrangement such as previously undetected inversions; these aberrant groups were not included in the consensus chromosome.

**Wheat-rice sequence comparisons:** The 119 group 4 consensus ESTs were compared to rice genomic sequences using blastN (<http://www.ncbi.nlm.nih.gov/BLAST/>) with cutoff parameters: *E*-value  $\leq 1E-15$ , identity (ID)  $\geq 80\%$  for at least 100 bp. ESTs that matched with a rice BAC were tabulated. The ESTs were then tentatively ordered on the basis of the established rice BAC orders.

**Wheat-Arabidopsis sequence comparisons:** Wheat-Arabidopsis sequence comparisons were performed in the same

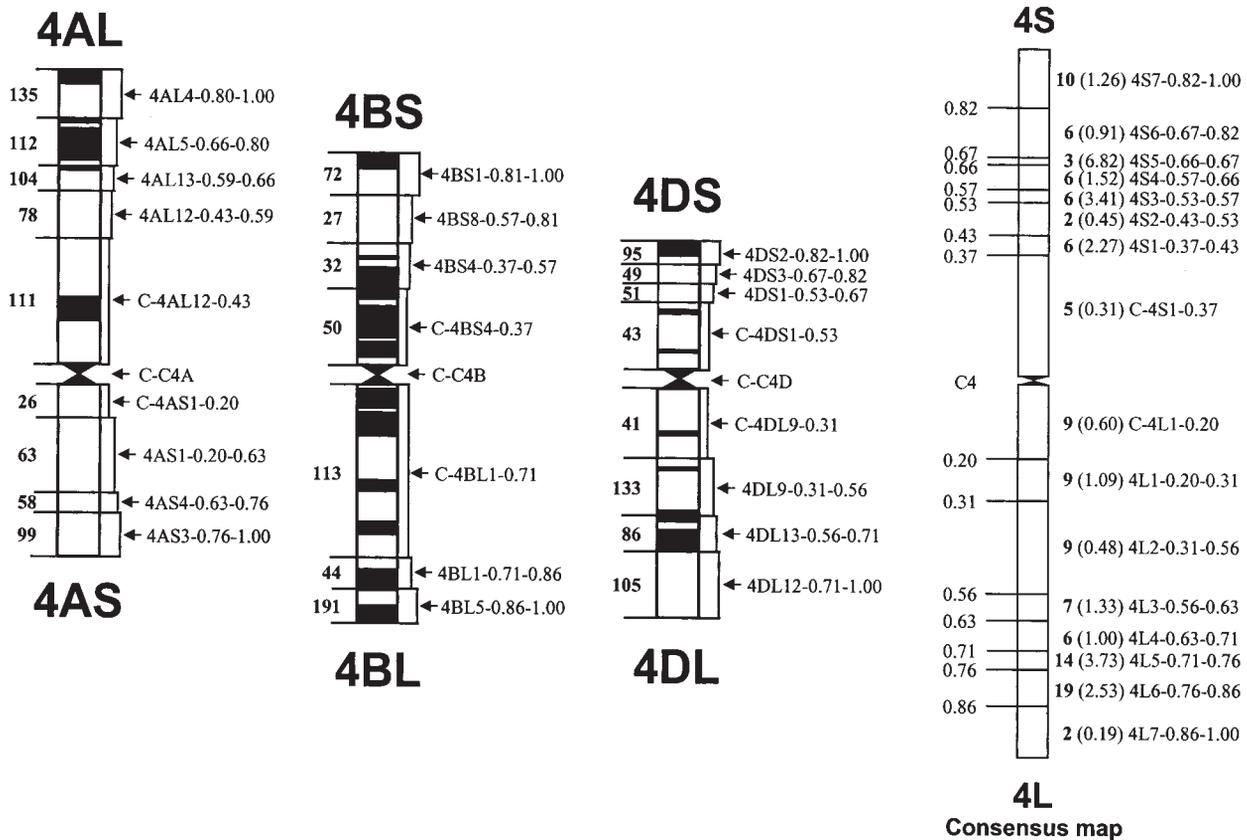


FIGURE 1.—C-banding ideograms of the three wheat homoeologous group 4 chromosomes and the group 4 consensus map. For the three ideograms, the number of EST loci mapped to the corresponding bins is given on the left and the deletion bin designation is on the right of each chromosome. For the consensus map, the consensus bin breakpoints are identified on the left and given on the right are the number of consensus ESTs in the bin (boldface), the density of that bin (in parentheses), and the bin designation. More detailed information on the identity of the consensus ESTs and their distribution on the three group 4 chromosomes can be found in the supplemental online material at <http://wheat.pw.usda.gov/pubs/2004/Genetics/>. Note that as the text and Figure 2 show, it is the long arm of 4A that is homoeologous to the short arms of 4B and 4D and the short arm of 4A that is homoeologous to the long arms of 4B and 4D. Therefore the ideogram for chromosome 4A is presented here with the long arm at the top to emphasize that homoeology.

manner as the wheat-rice comparison, using the 119 consensus ESTs. Nucleotide-nucleotide comparisons and protein-protein comparisons were analyzed. The cutoff values for protein-protein comparisons were  $E\text{-value} \leq 1E-5$  and  $ID \geq 70\%$  for at least 50 amino acid pairwise comparisons.

**Functional analysis of mapped ESTs:** All 938 mapped ESTs were used to perform blastX analysis against all protein databases. The results were then grouped on the basis of the functional category of Arabidopsis protein (<http://mips.gsf.de/proj/thal/db>). All groupings were performed manually.

## RESULTS AND DISCUSSION

**Chromosome deletion map and distribution pattern of EST loci on wheat homoeologous group 4:** From the March 17, 2003, data set, 938 ESTs mapped unambiguously to 1918 loci in homoeologous group 4 deletion bins. The number of mapped EST loci for group 4 was significantly higher than expected ( $P < 0.001$ ) on the basis of the relative size of the chromosomes (GILL *et al.* 1991; Table 1), with a mean of 2.04 loci per EST mapped. Chi-square analysis showed that the number

of EST loci mapped to 4A (786, 41%) was significantly higher than expected whereas those mapping to 4B (529, 28%) and 4D (603, 31%) were lower than expected, on the basis of the assumption that the chromosomes of group 4 have equal numbers of loci ( $P < 0.001$ ; Figure 1, Table 1). In contrast, the total number of EST loci mapping to the entire B genome (5774; 36%) exceeded those mapping to the A (5173; 32%) and D (5146; 32%) genomes (LAZO *et al.* 2004).

The total number of EST loci mapped to the long arms on the three chromosomes of homoeologous group 4 was twice the number mapped to the short arms and was higher than expected, on the basis of relative size ( $P < 0.001$ ; Table 1). Individual chromosome analyses of the EST loci distribution between the short and long chromosome arms showed that the number of EST loci mapped to all three short arms was lower than expected, while, except for 4DL, the long arms contained a higher than expected number. In general, the distributions of EST loci along group 4 chromosome

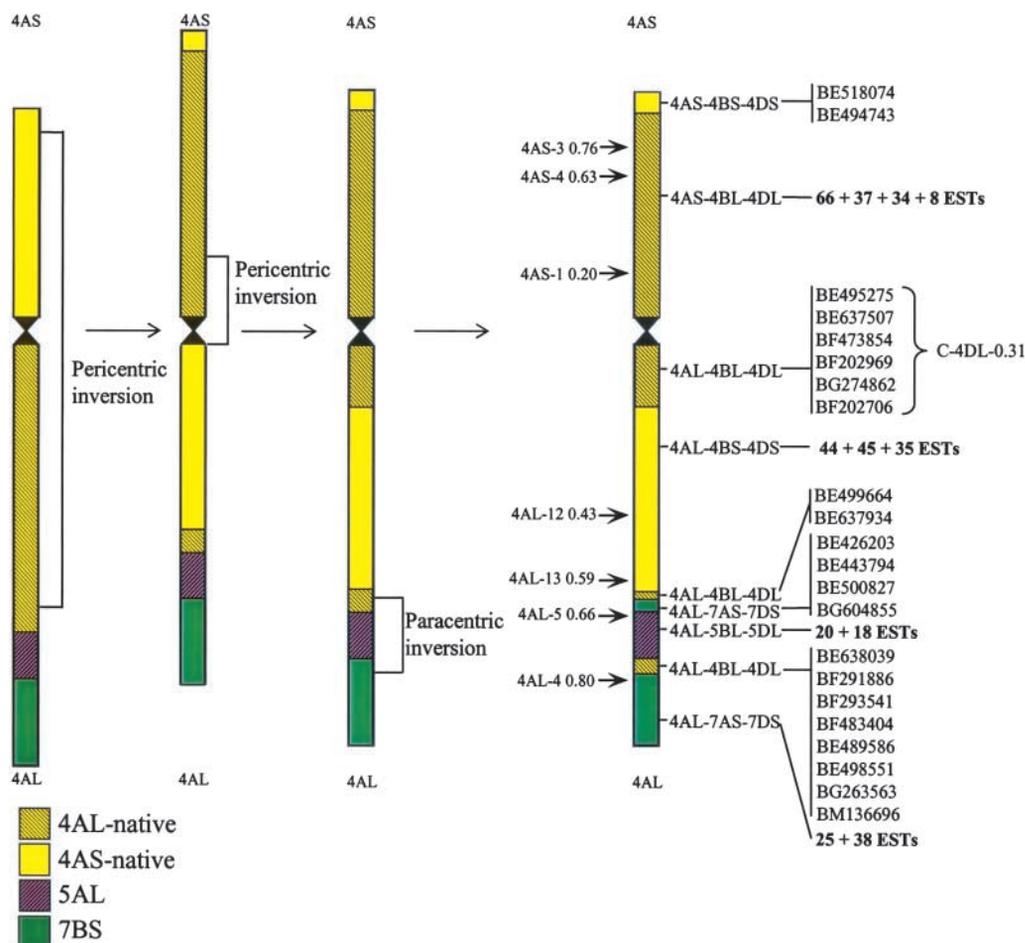


FIGURE 2.—Translocations involving chromosome arms 4AL, 5BL, and 7AS and inversions within chromosome 4A detected by EST mapping. In the rightmost ideogram, where >10 ESTs mapped in the corresponding segments, the total number of ESTs rather than their accession numbers are noted. The “+” sign between these numbers differentiates ESTs mapped to opposite sides of a deletion breakpoint when the segment covers more than one mapping bin.

arms were similar to those observed in the other homoeologous groups (AKHUNOV *et al.* 2003a; CONLEY *et al.* 2004; HOSSAIN *et al.* 2004; LINKIEWICZ *et al.* 2004; MUNKVOLD *et al.* 2004; PENG *et al.* 2004; RANDHAWA *et al.* 2004). The distal regions of the chromosome arms contained more loci than the proximal regions and more loci than expected, with the exception of 4DL. In chromosome arm 4DL, the highest number of EST loci was found in bin 4DL9-0.31-0.56, which is shorter than the terminal bin (Figure 1).

**Homeologous group 4 intra- and interchromosome gene duplication:** Gene duplication occurred within and among all three homoeologous group 4 chromosomes. A total of 375 or 40% of all mapped EST probes had duplicated (two or more) loci. This number is an approximation, because single loci with *EcoRI* sites within the region recognized by a probe would have been erroneously counted as duplications. This overestimation would have been partially balanced by the exclusion of some real duplication because of a lack of polymorphism among the duplicated copies. Among 631 loci duplicated among the three group 4 chromosomes, 6% were located on chromosomes 4A and 4B, 17% on 4A and 4D, and 23% on 4B and 4D. The remaining 54% were triplicated loci found on all three chromosomes. Of these 631 loci, 70 (11%) also mapped to chromo-

somes other than the three homoeologous group 4 chromosomes.

The intrachromosome distribution of duplicated loci on individual group 4 chromosomes was similar to the distribution of total loci. Chromosome 4B contained the fewest duplicated loci (157; 25%) when compared to 4A (268; 42%) and 4D (206; 33%) (Table 1). The long arms had twice as many duplicated loci as the short arms. The duplicated loci were mainly concentrated in the distal region of all chromosome arms with the exception of 4DL. The number of duplicated loci was positively correlated ( $R^2 = 0.78$ ;  $P < 0.001$ ) with the total number of mapped loci on the three homoeologous group 4 chromosomes.

**Chromosome-arm-specific genes:** We identified 10 ESTs that hybridized with single fragments that mapped to specific chromosome arms. However, further verification showed only two ESTs, BE490806 and BM140483, that were unique to a specific chromosome arm. Both ESTs were mapped to chromosome arm 4AL, one to bin 4AL13-0.59-0.66 and the other to bin 4AL4-0.80-1.00. In addition, we also identified one EST, BE426090, having two loci that both mapped to bin 4AL13-0.59-0.66. Those three chromosome arm-unique ESTs have considerable value as chromosome 4AL-specific markers.

Analysis of blastN against the rice genomic sequence

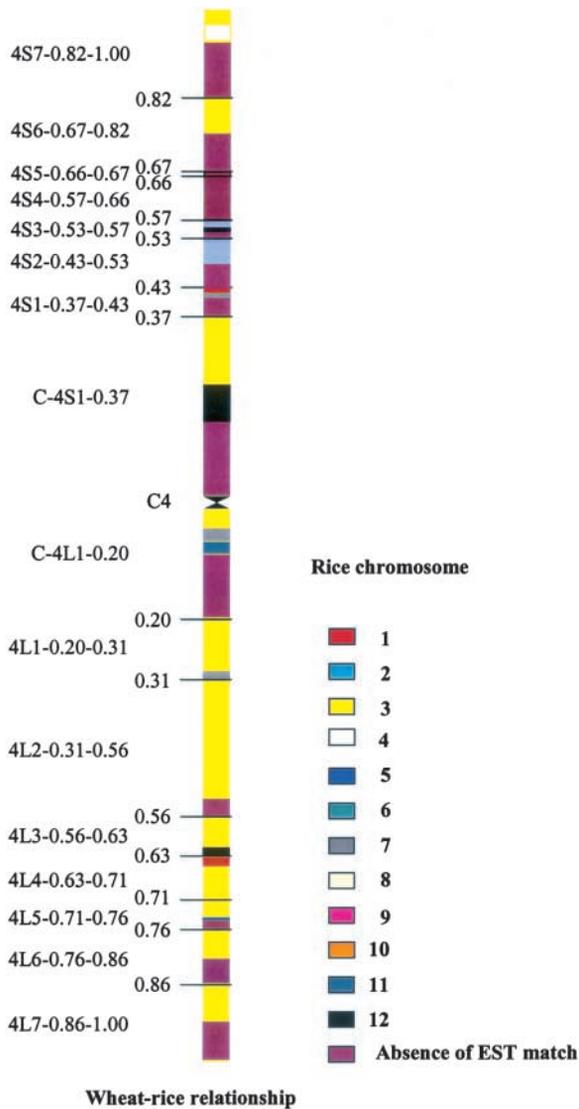


FIGURE 3.—Wheat-rice colinear relationships for the wheat ESTs composing the group 4 consensus chromosome. The consensus bin designations and bin breakpoints are given on the left of the wheat consensus chromosome. The colors on the wheat consensus chromosome represent specific rice chromosomes that had genomic sequences homologous to ESTs in those wheat consensus bins. The colors used correspond to the rice chromosome color designations used by *SORRELLS et al.* (2003). The order of the colors along the consensus wheat chromosome within a bin does not imply any order of ESTs within a bin with respect to rice homologs.

showed that only EST BE426090 had high homology with a rice BAC chromosome. A blastX analysis of the three ESTs also showed that only EST BE426090 had homology with a protein, namely a glutaredoxin-related protein.

**Chromosome 4A:** The chromosome 4A deletion map was generated by loci from 635 ESTs. Most of the 4A ESTs also mapped to 4B and 4D, but, in addition, 59 of them detected loci on 5BL and 5DL, and 72 detected loci on 7AS and 7DS. These data confirm the presence of two reciprocal translocations involving chromosome

arms 4AL, 5AL, and 7BS, and two inversions (one pericentric and one paracentric; Figure 2; *NARANJO et al.* 1987; *ANDERSON et al.* 1992; *LIU et al.* 1992; *DEVOS et al.* 1995; *MICKELSON-YOUNG et al.* 1995; *NELSON et al.* 1995; *CHEN and GUSTAFSON* 1997). In addition, the present study confirmed the chromosome arm location of the 4A translocations previously detected by *in situ* hybridization (*CHEN and GUSTAFSON* 1997).

**New pericentric inversion:** This study established one additional inversion, a small pericentric inversion in the centromeric region of 4A (Figure 2). This new inversion was found within the larger pericentric inversion previously described by *DEVOS et al.* (1995). This inversion was detected by 6 ESTs from four different mapping laboratories (Figure 2). Among those, 5 ESTs—BE49-5275, BE637507, BF473854, BG274862, and BF202-706—were mapped to C-4AL12-0.43, C-4BL1-0.71, and C-4DL9-0.31. However, 1 EST, BF202969, was mapped to C-4AL12-0.43, C-4BS4-0.37, and C-4DL9-0.31 as a result of a centromeric inversion in chromosome 4B. The fragment size involved in this inversion is unknown. Of the 88 ESTs mapped in the C-4AL12-12-0.43 bin, 44 showed a 4AL-4BS-4DS relationship. The remaining 35 ESTs were not colinear with either 4AL-4BL-4DL or 4AL-4BS-4DS. If the ESTs in this bin were evenly distributed, the expected length of the 4AL-native fragment would be  $\sim 11\%$  (6 ESTs out of 53) of the bin C-4AL12-0.43, which would correspond to  $\sim 5\%$  of the total 4AL arm. Centromeric regions typically exhibit the lowest gene density when compared to other chromosome regions, and thus the length of the inverted 4AL-native fragment might be much larger than expected. Greater detail and confirmation is required for a full interpretation of this evidence.

**Previously described 4AL paracentric inversion:** A native 4AL segment inversion, detected by RFLP analysis, near the distal end of 4AL was described by *MICKELSON-YOUNG et al.* (1995). The inversion was also reported by the observation of an inverted 5AL segment in the 4AL linkage map (*DEVOS et al.* 1995). Data from this study showed that this paracentric inversion involved the native 4AL segment, the translocated 5AL segment, and the translocated 7BS segment (Figure 2). Consequently, only a small piece of native 4AL was found in the original location within bin 4AL13-0.59-0.66. Two ESTs (BE49-9664 and BE637934) detected this fragment in chromosome bin 4AL13-0.59-0.66. The rest of this 4AL-native segment is now located between the 5AL-native and 7BS-native regions. Eight ESTs (BE638039, BF291886, BF293541, BF483404, BE489586, BE498551, BG263563, and BM136696) detected this segment in chromosome bin 4AL5-0.66-0.80, which appeared to be colinear with the distal regions of 4BL and 4DL (Figure 2). A native 7BS segment was inverted in bin 4AL13-0.59-0.66 (Figure 2). Four ESTs (BE426203, BE443794, BE500827, and BG604855) detected this 7BS inversion. These ESTs also mapped to the distal regions of the 7AS and 7DS.

TABLE 2  
Distribution of homoeologous group 4 wheat consensus ESTs on rice chromosomes

Wheat bin	Rice chromosomes												No match	Total matches	P first	P second	
	1	2	3	4	5	6	7	8	9	10	11	12					
4S7-0.82-1.00			2	2										6	4	0.035	0.007
4S6-0.67-0.82			3											3	3	0.001	—
4S5-0.66-0.67														3	0	—	—
4S4-0.57-0.66														6	0	—	—
4S3-0.53-0.57											2	2		2	4	0.035	0.007
4S2-0.43-0.53											1			1	1	0.083	—
4S1-0.37-0.43	1						1							4	2	0.152	0.152
C-4S1-0.37			2											2	3	0.019	0.083
C-4L1-0.20			2				1				1			5	4	0.035	0.152
4L1-0.20-0.31			8				1							0	9	<0.001	0.083
4L2-0.31-0.56			8											1	8	<0.001	—
4L3-0.56-0.63			6									1		0	7	<0.001	0.083
4L4-0.63-0.71	1		5											0	6	<0.001	0.083
4L5-0.71-0.76			10		1									3	11	<0.001	0.083
4L6-0.76-0.86			11											8	11	<0.001	—
4L7-0.86-1.00			1											1	1	0.083	—
Total ESTs	2	0	58	2	1	0	3	0	0	0	4	4		45	119		

Bin 4AL13-0.59-0.66 contains adjacent segments of 4AS, 4AL, 7BS, and 5AL. In total, 80 ESTs were mapped in this bin, among which 35 mapped on the 4AS-native segment, 2 mapped on the 4AL-native segment, 4 mapped on the 7BS segment, and 20 mapped on the 5AL segment. The 19 remaining ESTs within this bin did not show a colinear relationship with the other homoeologous group 4 chromosomes, on the basis of the available data.

**Evolutionary order of the 4A structural changes:** Altogether, the current mapping data indicated two reciprocal translocations (4AL/5AL and 4AL/7BS), two pericentric inversions, and one paracentric inversion. The major remaining questions involve the evolutionary order of the five chromosome rearrangement events. DEVOS *et al.* (1995) found that the same 4AL/5AL translocation also appeared in *T. monococcum* and concluded that the 4AL/5AL translocation occurred at the diploid level. On the basis of their linkage data and the cytological study by RILEY *et al.* (1967), DEVOS *et al.* (1995) also concluded that the 4AL/7BS translocation and the paracentric and the larger pericentric inversions arose at the tetraploid level. However, the temporal order of those events could not be fully resolved, although the 4AL/7BS translocation must have occurred before the paracentric inversion due to the observation of an inverted 5AL segment in the bread wheat chromosome arm 4AL (DEVOS *et al.* 1995). The larger pericentric inversion could have taken place either before or after the 4AL/7BS translocation, or even after the paracentric inversion on 4AL (DEVOS *et al.* 1995). The present results provide new evidence that the 4AL/7BS translo-

cation occurred before the paracentric inversion on the 4AL arm because the two 4AL-native segments within the region are interstitial and separated by the 5AL segments, as are the two 7BS segments. However, the temporal order of the larger pericentric inversion, the smaller pericentric inversion, and the paracentric inversion on 4AL remains unclear (Figure 2).

**Chromosomes 4B and 4D:** Chromosome 4B had the fewest mapped ESTs (458) and loci of the three group 4 chromosomes. In general, the EST mapping data support previously reported data showing that chromosome arm 4BL is homoeologous to 4DL and 4AS, and 4BS is homoeologous to 4DS and 4AL (DEVOS *et al.* 1995; MICKELSON-YOUNG *et al.* 1995). The ESTs located in the centromeric region and again in the distal region of 4BL show a 4BL-4DL-4AL pattern reflecting the 4AL inversions explained above.

Four earlier studies (ENDO and GILL 1984; DVOŘÁK *et al.* 1984; GILL *et al.* 1991; MICKELSON-YOUNG *et al.* 1995) found evidence of a pericentric inversion in 4B and four ESTs in this study support this. Three of these, BF202211, BE406512, and BF202969, mapped to C-4BS4-0.37 and chromosome arm 4DL. BF202211 did not map to chromosome 4A. BE406512 mapped to 4AS1-0.20-0.63. B202969 mapped to C-4BS, C-4DL, and C-4AL. BE497309 mapped to C-4BL1-0.71, C-4DS1-0.53, and C-4AL12-0.43.

A total of 549 informative ESTs mapped to chromosome 4D. One probe accession, BE442750, was unique to chromosome 4D and mapped to bin 4DL12-0.71-1.00. In the set of data used for this study, 108 ESTs, which mapped to 4DL, also mapped to 4AS and 4BL.

**TABLE 3**  
**cDNA libraries containing ESTs mapped to wheat homoeologous group 4 chromosomes**

Name <sup>a</sup>	Tissue and condition <sup>b</sup>	ESTs generated	ESTs mapped to group 4 <sup>c</sup>
TA016E1X	Crown, vernalized	2,286	47
TA038E1X	Crown, salt stressed	943	22
TA001E1X	Endosperm (Cheyenne)	2,728	40
TA036E1X	Leaf, drought stressed	641	14
TA027E1X	Leaf, drought stressed (TAM W101)	905	8
TA031E1X	Leaf, flag, heat stressed	973	36
TA037E1X	Leaf, sheath, salt stressed	964	15
TA007E1X	Seedling, cold stressed	938	25
TA005E1X	Seedling, drought stressed	795	4
TA015E1X	Seedling, heat stressed	821	22
TA008E1X	Root, seedling, etiolated	4,017	99
TA008E3N	Root, seedling, etiolated, normalized	4,308	102
TA006E1X	Shoot, seedling, unstressed	2,261	28
TA006E2N	Shoot, seedling, unstressed, normalized	1,686	13
TA018E1X	Spike, 5–15 DAP	2,860	43
TA032E1X	Spike, 5–20 DAP, heat stressed	1,012	27
TA017E1X	Spike, 20–45 DAP	1,076	35
TA009XXX	Spike, <i>Fusarium graminearum</i> infected (Sumai3)	10,287	47
TA006G1X	Spike, <i>F. graminearum</i> -infected (Sumai3)	727	15
TA019E1X	Spike, preanthesis	11,194	206
TM043E1X	Apex, shoot, early reproductive, 7-wk vernalized (acc. Dv92)	2,647	42
TT039E1X	Whole plant (Langdon-16)	1,194	17
AS040E1X	Anther	2,466	42
SC024E1X	Anther (Blanco)	4,631	66

<sup>a</sup> The species source of the library is indicated by the first two letters of the name: TA, *Triticum aestivum*; TM, *T. monococcum*; TT, *T. turgidum*; SC, *Secale cereale*; and AS, *Aegilops speltoides*.

<sup>b</sup> All of the TA libraries are from the hexaploid wheat Chinese Spring except where indicated otherwise in parentheses. DAP, days after pollination.

<sup>c</sup> This column totals 1015 ESTs; however, the analyses reported in this article involve the subset of 938 of them that remained after ambiguous ESTs were discarded.

Sixteen ESTs, which mapped to 4DL, mapped to 4AL and 4BL, of which 6, noted above, were involved in the new pericentric inversion on 4A. A total of 44 ESTs showed a colinear mapping relationship with 4BL and 5AL23-0.87-1.00. These data confirm the 5AL/4AL reciprocal translocation in *T. aestivum* proposed by NARANJO *et al.* (1987).

**Consensus EST map of homoeologous group 4 chromosomes:** Developing the group 4 consensus map from the three homoeologous group 4 chromosomes was complicated by the several translocations and inversions of chromosome 4A. A total of 141 ESTs were detected as unique for group 4 chromosomes, 4 of which were excluded because they did not map to a specific bin or the data were unclear. We detected 18 ESTs, which mapped to various bins among the three chromosomes, making it difficult to assign them to a consensus bin. These anomalies were also excluded from the consensus map. Most of the anomalies involved the long arm of 4A and the short arms of 4B and 4D, which was not surprising owing to the significant rearrangements that have occurred in the long arm of 4A. Thus, we used

119 ESTs that could be clearly assigned to 16 consensus bins. The number of ESTs assigned to each bin varied from 2 to 19. The numbers of ESTs in long and short arms were not equal, with the long arm containing 70% more ESTs than the short arm. Unlike the distribution of ESTs mapping to individual chromosomes, there was no clear distribution pattern along the consensus chromosome. The EST density ranged from 0.19 to 6.82 ESTs per consensus bin (Figure 1). However, it should be pointed out that the relative error in the estimation of a bin size is larger in the smaller bins, and therefore these values have a large variance. As observed in other homoeologous groups in this project, the smaller the bin the higher the EST density.

The consensus chromosome map revealed a clearer picture of the EST density along the homoeologous group 4 chromosomes. The consensus map showed that the highest EST density was not located in the most distal bins of the consensus map, but closer to the middle of both consensus arms (Figure 1). Since the consensus map took into account the vagaries of bin size and utilized more bins per arm, it gave a considerably higher-

**TABLE 4**  
**The highest-copy-number genes/gene families present in homoeologous group 4**

Putative function	EST accession no.	cDNA library	Copy no. <sup>a</sup>	Chromosome location <sup>b</sup>
Oxalate oxidase	BE406676	TA008E1X	11 (13)	C4AL12 (2); 4AL13; 4BS1 (2); 4DS2 (6); C2AS5; 3BL2
Cytochrome P450s	BF145360	SC024E1X	3	4AL12; 4BS4; 4DS1
	BG314543	TM043E1X	2 (12)	4A (2)
	BM136696	TA009XXX	3	4AL5; 4BL5; 4DL12
	BM138474	TA009XXX	3	4AL12; 4BS8; 4DS3
2-Oxoglutarate/malate translocator Glutathione S-transferases	BE403439	TA008E1X	10 (13)	C4AL12; 4BS1 (3); 4DS2 (6); 3BL2
	BE490186	TA007E1X	2 (7)	4AL4 (2); 7AS1; 7DS4 (2)
	BE405625	TA008E1X	5 (11)	4BS1 (3); 4DS1 (2)
	BM138536	TA009XXX	1 (4)	4AL4; 7AS1; 7DS4 (2)
Tubulin $\alpha$ -2 chain	BE490658	TA007E1X	5 (21)	4AS1; 4AS3; 4AL13; 4BL5; 4DL13; 3AL3; 3BL2; 3DL2
	BE496999	TA015E1X	4 (5)	4AL12; 4BS1; 4DS2 (2)
ATP-dependent Clp proteases	BE490177	TA007E1X	2 (3)	4DL9 (2)
	BE490526	TA007E1X	2 (5)	4AL5 (2); 7AS5 (2); 7DS5
	BE591521	TA031E1X	3	4AS1; C4BL1; 4DL9
	BE446076	TA008E3N	7	4AS4 (3); C4BL1 (2); 4DL9 (2)
MAP kinase 1	BE497108	TA015E1X	3	4AS3; C4BL1; 4DL9;
Ascorbate peroxidases	BE497618	TA015E1X	3	4AS4; C4BL1; 4DL9;
Cell division cycle protein 48	BE444757	TA008E3N	6 (8)	4AS3; 4BL5 (2); 4DL13 (3); 1AL1; 1BL6
Elongin	BE403789	TA008E1X	6	C4AL12 (2); 4BS4 (2); C4DS1 (2)
Heat-shock protein (class I)	BE605229	TA032E1X	6 (11)	4AS4; C4BL1; 4DL9 (4); 3BS8; 3DS6
Mitochondrial processing peptidase	BE442995	TA008E3N	6	4AS3 (2); 4BL5 (2); 4DL9 (2)
S-Malonyltransferase	BE404963	TA008E1X	6	C4AL12; 4DL9 (2); C4BL1 (3)
Other kinases			53	
Other peroxidases			8	

<sup>a</sup>The number of EST loci mapped to homoeologous group 4 chromosomes for the given EST accession; the number in parentheses indicates the total number of loci generated by this EST, whether mapped or not.

<sup>b</sup>The chromosome, chromosome arm, or bin designation is given; if more than one locus maps to a location, the number of loci is given in parentheses following the location. Not all loci were mappable; thus the total number of loci in this column will not always equal the total from the previous column.

resolution view of actual EST location than the individual chromosome maps did (Figure 1). However, consensus maps will be of limited value until a much larger number of bins per chromosome arm are available to create high-resolution maps.

**EST-based wheat-rice colinear relationship:** On the basis of the homoeologous group 4 consensus map, the relationship between ESTs on wheat group 4 chromosomes and the rice genome showed that among 119 ESTs, 74 ESTs (62%) significantly matched to rice BACs (Figure 3). Of those 74 ESTs, 58 (78%) were present in wheat deletion bins that were colinear with the corresponding groups of genes from rice chromosome 3 to which they matched (Table 2). No claims can be made for wheat EST linear order within bins, but only for the order of wheat deletion bins. The greatest similarity in order of the wheat deletion bins and the corresponding groups of rice genes occurred primarily between the long arm of the consensus group 4 map (88%) and the short arm of rice chromosome 3 (Figure 3), as observed by SORRELLS *et al.* (2003). This conserved order spanned

the short arm of rice chromosome 3 from 14.4 to 55.8 cM with several interruptions primarily for the bins at the distal end of the group 4 consensus long arm. A gap in the linear order occurred in the short arm of rice chromosome 3 corresponding to the group 4 consensus centromeric bins, which suggested a large rearrangement in these regions during the evolutionary history of one or both species. This interruption of order between the wheat group 4 consensus chromosome bins and corresponding groups of rice chromosome 3 genes will complicate the use of rice-wheat synteny to facilitate gene discovery in wheat and other Triticeae species. Rice chromosome 3 also showed long segments of conserved order with wheat homoeologous group 5 (LIN-KIEWICZ *et al.* 2004).

**Wheat-Arabidopsis relationship:** The relationship between wheat ESTs and the Arabidopsis genome was analyzed using a blastN search against the Arabidopsis genome database. Among the 119 consensus ESTs, only 26 of them showed high homology ( $E$ -value  $\leq E^{-5}$ ; ID  $\geq 80\%$  for  $>50$  bp) to Arabidopsis genomic se-

**TABLE 5**  
**Group 4-specific genes and location on homoeologous chromosome group 4**

Putative function	BlastX <sup>a</sup>	Accession no.	Chromosome location <sup>b</sup>
40S ribosomal protein S3A	2E-45	BE425301	4AS4; 4BL5; 4DL9
50S ribosomal protein L15 (chloroplast)	2E-53	BE517664	4AS3; 4BL5; 4DL13 (2)
60S ribosomal protein L10A	2E-24	BG263898	4AL13; 4BS1; 4DS1
Actin 1	3E-95	BE490281	4AL12; 4BS1; 4DS2
α-Amylase inhibitor, chain CM17	2E-70	BE424208	4BL5; 4DL12
Arabinoxylan arabinofuranohydrolase AXAH-II	3E-71	BF473320	4AL; 4BS4; 4DS1
Ascorbate peroxidase	2E-34	BE497618	4AS4; C4BL1; 4DL9
Ascorbate peroxidase (stromal)	4E-40	BE497108	4AS3; C4BL1; 4DL9
Basic protein 1B, WBP1B	7E-14	BE517725	4AL12; 4BS4; 4DS1
Benzothiadiazole-induced protein	9E-21	BE403491	4BS1 (2); 4DS2; 4DS3
β-3-Glucanase	1E-4	BF428777	4AL13; 4BS8; 4DS3
β-Galactosidase	4E-64	BE426767	4AS4; 4BL1; C4D
β-Galactosidase BG1	7E-18	BF201449	4AS3; 4BL1; 4DL12
β-Glucosidase aggregating factor	2E-17	BE445831	4BS1; 4DS2
C-4 sterol methyl oxidase	1E-59	BE497184	C4AL12; 4BS4; C4DS1
Catechol O-methyltransferase	7E-31	BE405350	4AL12; 4BS8; 4DS2
CAXIP1 protein	2E-15	BE426090	4AL13 (2)
Chlorophyll a/b-binding protein CP29	2E-56	BE497446	4AL12; 4BS8; 4DS3
Cytochrome P450-1	1E-26	BM138474	4AL12; 4BS8; 4DS3
Cytochrome P450-2	3E-29	BE404334	4BL5
Cytochrome P450-3	6E-38	BM136696	4AL5; 4BL5; 4DL12
Elongin	1E-27	BE403789	C4AL12 (2); 4BS4 (2); C4DS1 (2)
Endo-1,4-β-glucanase Cell	4E-89	BE498699	4AL12; 4BS1; 4DS2
F-actin capping protein β-subunit	3E-59	BF145268	C4AS1; C4BL1; C4DL9
Flocculin	1E-4	BF201717	4A; C4BL5; C4DL9
Fructose-6-phosphate 2-kinase	2E-15	BF483393	C4BL5; 4DL9
GDP dissociation inhibitor	6E-77	BF485396	4AS4 (2); C4BL1; 4DL9
Glucosyltransferase-like protein	5E-45	BE443973	4AS1; C4BL1; 4DL9
Glutathione reductase	3E-91	BE498405	4AS3; 4BL5 (2); 4DL12
H-β-58 homologue	1E-85	BE444811	C4A; C4BS4; C4DS1
Lipoxygenase	1E-112	BE442666	4AL13; 4BS8; 4DS3
Low-temperature/salt stress-induced protein	7E-19	BE518424	4AS4 (2); C4BS1; 4DL9
MAP kinase 1	2E-64	BE446076	4AS4 (3); C4BL1 (2); 4DL9 (2)
Mitochondrial processing peptidase	3E-85	BE442995	4AS3 (2); 4BL5 (2); 4DL9 (2)
Myosin II heavy chain	3E-46	BF483405	4AS4; 4BL1; 4DL13
NADP-dependent oxidoreductase	8E-24	BE490685	4AL12 (3); 4BS8
O-Succinylhomoserine (thiol)-lyase	6E-54	BF202969	C4AL12; C4BS4; C4DL9
Oxygen-evolving complex 25.6-kD protein	6E-37	BE426317	4AS4; C4BL1; 4DL9 (2)
Peroxidase ATP13a	3E-11	BE497785	4AS3; 4BL5; 4DL12
Phosphoinositide kinase	7E-51	BE500180	C4AL12; C4BL1; C4DL9
Phosphomevalonate kinase	4E-50	BE494575	4AS3; C4BL1; 4DL9
Prenyltransferase	4E-54	BE591142	4AS4; C4BL1; 4DL9
Pseudo-response regulator 9	1E-15	BE442811	4AS4; C4BL1; 4DL9
Putative neutral invertase	9E-96	BE443291	4AS1; C4BL1; 4DL9
Putative purple acid phosphatase	1E-91	BE403881	C4AL12; 4BL1; 4DL12
Ribonuclease	2E-65	BE426266	4AS4; 4BL1; 4DL9
Ribosomal protein S5	3E-9	BE497635	4AL12; 4BS4; 4DS1
Ribulose-phosphate 3-epimerase (chloroplast)	1E-62	BE497160	4AS3; 4BL5; 4DL12
Ring-H2 finger protein	5E-32	BE443211	C4AL12; 4BL5 (2); C4DL9
RNA-binding protein	3E-15	BE406512	4AS1 (2); C4BS4; 4DL9
RNA-binding protein	2E-19	BG313203	4AS3; 4BL5; 4DL12
Serine palmitoyltransferase	3E-53	BM138075	4AS4; C4BL1; 4DL9
Serine threonine kinase	1E-29	BF428648	4AL12; 4BS4; 4DS3
SH3 domain-containing protein	4E-22	BE423455	4AS4; C4D
Small nuclear ribonucleoprotein U1	7E-8	BF429264	4BS; 4DS2
S-Malonyltransferase	8E-86	BE404963	C4AL12; C4BL1 (3); 4DL9 (2)
Spliceosome-associated protein	1E-28	BM137730	4AS3; C4BL1; C4DL13
Sugar transporter	6E-11	BF201181	4AS3; 4BL5; 4DL13

(continued)

TABLE 5  
(Continued)

Putative function	BlastX <sup>a</sup>	Accession no.	Chromosome location <sup>b</sup>
Sulfolipid synthase	2E-76	BG313162	4AS4; C4BL1; 4DL9
Transcription factor X1	2E-12	BE494478	4AS3; 4BL1; 4DL9
Transport inhibitor response 1 (TIR1)	4E-45	BG606739	4AS4; 4BL1; 4DL12
Ubiquitin-conjugating enzyme E2	2E-24	BE443190	4AS1; C4BL1; 4DL9
Vacuolar targeting receptor bp-80	5E-83	BE426016	4AS4; C4BL1; 4DL9
Zinc-finger protein C60910	4E-75	BF478884	4AS1; C4BL1; 4DL9

<sup>a</sup> BlastX *E*-value of the hit with highest similarity to the EST.

<sup>b</sup> The chromosome, chromosome arm, or bin designation is given; if more than one locus maps to a location, the number of loci is given in parentheses following the location.

quences from five different chromosomes. This result indicated that there was no structural relationship between the Arabidopsis genome and wheat homoeologous group 4 chromosomes.

Analysis of blastX for protein sequence comparisons between wheat chromosome 4 consensus ESTs and the Arabidopsis protein database showed that only 38 ESTs had significant similarity (*E*-value  $\leq E-5$ ; ID  $\geq 70\%$  for a minimum of 50 amino acids) to Arabidopsis proteins. The proteins were diverse, including those involved in metabolic processes, cellular processes, transport proteins, ribosomal proteins, and unknown proteins, and were encoded by genes located on five different Arabidopsis chromosomes with no apparent clustering pattern. Unlike the wheat-rice comparison, the results of both nucleotide and protein comparisons between wheat homoeologous group 4 consensus chromosome ESTs and the Arabidopsis genome did not show a colinear relationship between the two species.

**Functional characterization of group 4 ESTs:** The unigenes corresponding to the EST loci mapped to homoeologous group 4 chromosomes were analyzed for their functional characteristics. Information about their GenBank accession numbers, putative function, and chromosomal location can be obtained at [http://wheat.pw.usda.gov/cgi-bin/westsql/map\\_locus.cgi](http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi). Twenty-four of the 26 libraries that generated group 4 ESTs were used in this analysis (Table 3). The cDNA libraries were designed to maximize gene discovery, since virtually all wheat tissues were sampled, including several developmental stages and environmental conditions. In general, the number of homoeologous group 4 ESTs from each library correlated positively with the total number of ESTs mapped to the seven homoeologous groups and the number of ESTs generated from each library, with only a few exceptions. Homoeologous group 4 chromosomes contain genes involved in a wide range of cellular mechanisms (Tables 4 and 5, Figure 4). To refine the homoeologous group 4 chromosome transcriptome, the unigenes were grouped into functional categories on the basis of a blastX search. Only 42% of the genes could be functionally classified in this

manner, whereas the rest of the genes were similar to novel proteins for which no putative functions have been assigned (27%) or did not show significant similarity to any protein (31%). The majority of known cellular functions were represented in the 440 unigenes, although genes involved in metabolism (14%), signal transduction (12%), and defense (11%) were the most abundant (Figure 4). These results are similar to those observed in rye (*Secale cereale* L.) (RODRIGUEZ MILLA *et al.* 2002), in the rice gene prediction classification (GOFF *et al.* 2002), and also in the assignment of functional categories to the Arabidopsis proteome (ARABIDOPSIS GENOME INITIATIVE 2000). Therefore, wheat homoeologous group 4 chromosomes (and probably the entire wheat genome) follow general trends similar to those observed in other species.

As expected, an analysis of different genomes has revealed that a large number of genes belong to multi-copy gene families. For instance, in rice, as in Arabidopsis, extensive gene redundancy exists in all metabolic pathways, which may facilitate the tightly regulated expression of specific isozymes (GOFF *et al.* 2002). In

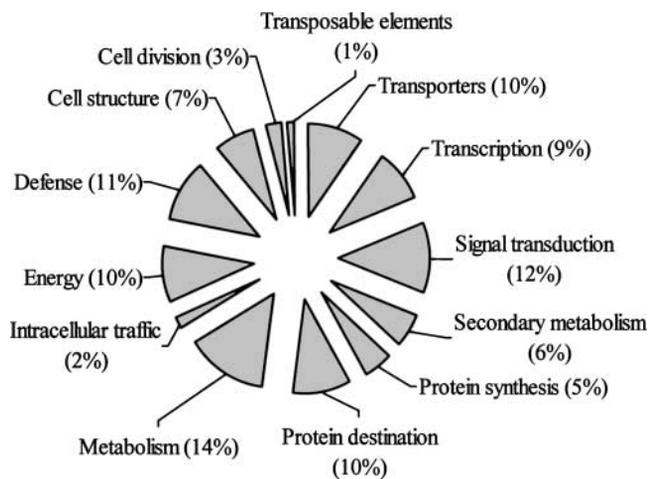


FIGURE 4.—Functional classification of wheat ESTs mapped on group 4 chromosomes based on the Arabidopsis gene classification.

addition, a significant correlation between the copy number of Arabidopsis and tomato (*Lycopersicon esculentum* L.) multigene families was reported by VAN DER HOEVEN *et al.* (2002). We used only EST loci mapping to wheat homoeologous group 4 chromosomes to estimate the number of gene families present in this transcriptome. The copy numbers in Table 4 are probably underestimated, because they are based on RFLP data. Lower stringency conditions would probably reveal a higher number of loci (*i.e.*, additional family members); however, this issue would be best addressed by analysis of DNA sequences rather than by Southern hybridization. Nevertheless, protein kinases represented the largest family in wheat homoeologous group 4, and peroxidases, cytochrome P450s, and glutathione S-transferases represented large families as well (Table 4). These four families were also among the most abundant in Arabidopsis and tomato (VAN DER HOEVEN *et al.* 2002). We noted a high copy number of genes encoding a putative oxalate oxidase (11) and a 2-oxoglutarate/malate translocator (10) present in homoeologous group 4 chromosomes.

Finally, to analyze further the nature of the genes present in homoeologous group 4, we examined more closely the putative functional role of 64 group 4-specific genes (Table 5). Of the 440 known-function unigenes studied, 56 (13%) were assigned to all three homoeologous group 4 chromosomes and 8 were assigned to only one or two of them (Table 5). The presence of one copy in each of the three homoeologous chromosomes suggests that they were also present as single-copy genes in the common ancestor of the A, B, and D genomes. It is possible that other genes are group 4 specific, because they showed three bands (one in each homoeologous chromosome), but they also showed one or two or more additional faint, nonmapped bands. Under higher-stringency conditions, the additional bands would likely disappear, especially in those cases where they represent different genes with a high degree of similarity. Two genes, encoding a mitochondrial processing peptidase and an elongin, were present in two copies in each of the homoeologous chromosomes, which indicated that they could have been duplicated in the original wheat genome ancestor. However, because in the current study RFLP deletion mapping was performed using a single restriction enzyme (*EcoRI*), additional enzymes would be necessary to prove this hypothesis. Other genes showed different copy numbers in each of the homoeologous group 4 chromosomes, suggesting that duplication events have commonly occurred following the divergence of the A, B, and D genomes. Interestingly, a gene encoding a MAP kinase 1 protein homolog was present in seven copies, three in 4AS, two in 4BL, and two in 4DL. No additional loci appear to be located anywhere else in the hexaploid wheat genome. The assignment of functional categories for the 56 group 4-specific genes was very similar to the results obtained

for the total group of 440 unigenes, as genes covering most of the cellular functions were detected.

In spite of the limitations, the present data support the idea that the overall gene composition of the wheat transcriptome follows the general trend observed in other species. Although proteins of all functional classes were present, a slight bias (more significant for group 4-specific genes) was found for genes encoding metabolic functions, which is consistent with the results obtained in other plant species. In addition, the results suggested that the hexaploid wheat genome is well "balanced," because each of its homoeologous groups appears to contain genes involved in diverse functions. However, much higher-resolution mapping is necessary to determine whether or not some chromosome regions show specific characteristics.

This material is based upon work supported by the National Science Foundation under cooperative agreement no. DBI-9975989; United States Department of Agriculture (USDA), Agricultural Research Service Current Research Information System (project no. 3622-21000-023-00D); and the University of Missouri, Agricultural Experimental Station. This article reports the results of research only. Mention of a proprietary product does not constitute an endorsement or a recommendation for its use by the USDA or the University of Missouri.

#### LITERATURE CITED

- AKHUNOV, E. D., A. W. GOODYEAR, S. GENG, L. L. QI, B. ECHALIER *et al.*, 2003a The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res.* **13**: 753–763.
- AKHUNOV, E. D., A. R. AKHUNOVA, A. M. LINKIEWICZ, J. DUBCOVSKY, D. HUMMEL *et al.*, 2003b Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. *Proc. Natl. Acad. Sci. USA* **100**: 10836–10841.
- ANDERSON, J. A., Y. OGIHARA, M. E. SORRELLS and S. D. TANKSLEY, 1992 Development of a chromosomal arm map for wheat based on RFLP markers. *Theor. Appl. Genet.* **83**: 1035–1043.
- ARABIDOPSIS GENOME INITIATIVE, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- CHAO, S., P. J. SHARP, A. J. WORLAND, E. J. WARHAM, R. M. D. KOEBNER *et al.*, 1989 RFLP-based genetic maps of wheat homologous group 7 chromosomes. *Theor. Appl. Genet.* **78**: 495–504.
- CHEN, J., and J. P. GUSTAFSON, 1994 Physical mapping of genetically mapped molecular markers in homoeologous group 7 chromosomes of wheat by *in situ* hybridization. *J. Jiangsu Agric. Coll.* **15**: 1–9.
- CHEN, J., and J. P. GUSTAFSON, 1997 Chromosomal rearrangement of wheat (*T. aestivum*) chromosome 4A by *in situ* hybridization. *Chin. J. Genet.* **24**: 39–47.
- CONLEY, E. J., V. NDUATI, J. L. GONZALEZ-HERNANDEZ, A. MESFIN, M. TRUDEAU-SPANJERS *et al.*, 2004 A 2600-locus chromosome bin map of wheat homoeologous group 2 reveals interstitial gene-rich islands and colinearity with rice. *Genetics* **168**: 625–637.
- DEVOS, K. M., J. DUBCOVSKY, J. DVOŘÁK, C. N. CHINYOY and M. D. GALE, 1995 Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theor. Appl. Genet.* **91**: 282–288.
- DVOŘÁK, J., P. E. MCGUIRE and S. MENDLINGER, 1984 Inferred chromosome morphology of the ancestral genome of *Triticum*. *Plant Syst. Evol.* **144**: 209–220.
- ENDO, T. R., 1988 Induction of chromosome structural changes by a chromosome of *Aegilops cylindrica* L. in common wheat. *J. Hered.* **79**: 366–370.

- ENDO, T. R., 1990 Gametocidal chromosomes and their induction of chromosome mutations in wheat. *Jpn. J. Genet.* **65**: 135–152.
- ENDO, T. R., and B. S. GILL, 1984 Somatic karyotype, heterochromatin distribution, and nature of chromosome differentiation in common wheat, *Triticum aestivum* L. em. Thell. *Chromosoma* **89**: 361–369.
- ENDO, T. R., and B. S. GILL, 1996 The deletion stocks of common wheat. *J. Hered.* **87**: 295–307.
- FRIEBE, B., J. JIANG, D. R. KNOTT and B. S. GILL, 1994 Compensation indices of radiation-induced wheat-*Agropyron elongatum* translocations conferring resistance to leaf rust and stem rust. *Crop Sci.* **34**: 400–404.
- FRIEBE, B., R. G. KYNAST, P. ZHANG, L. QI, M. DHAR *et al.*, 2001 Chromosome healing by addition of telomeric repeats in wheat occurs during the first mitotic divisions of the sporophyte and is a gradual process. *Chromosome Res.* **9**: 137–146.
- GALE, M. D., 1990 RFLP mapping in wheat—progress and problems, pp. 353–36 in *Gene Manipulation in Plant Improvement II: 19th Stadler Genetics Symposium*, edited by J. P. GUSTAFSON. Plenum Press, New York.
- GALE, M. D., M. D. ATKINSON, C. N. CHINYOY, R. L. HARCOURT, J. JIA *et al.*, 1995 Genetic maps of hexaploid wheat, pp. 29–40 in *Proceedings of the 8th International Wheat Genetics Symposium*, edited by Z. S. LI and Z. Y. XIN. China Agricultural Press, Beijing, China.
- GILL, B. S., B. FRIEBE and T. R. ENDO, 1991 Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*). *Genome* **34**: 830–839.
- GOFF, S. A., D. RICKE, T. H. LAN, G. PRESTING and R. WANG, 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* **296**: 92–100.
- HOSSAIN, K. G., V. KALAVACHARLA, G. R. LAZO, J. HEGSTAD, M. J. WENTZ *et al.*, 2004 A chromosome bin map of 2148 expressed sequence tag loci of wheat homoeologous group 7. *Genetics* **168**: 687–699.
- LAZO, G. R., S. CHAO, D. D. HUMMEL, H. EDWARDS, C. C. CROSSMAN *et al.*, 2004 Development of an expressed sequence tag (EST) resource for wheat (*Triticum aestivum* L.): unigene analysis, probe selection and bioinformatics for a 16,000-locus bin map. *Genetics* **168**: 585–593.
- LINKIEWICZ, A. M., L. L. QI, B. S. GILL, B. ECHALIER, S. CHAO *et al.*, 2004 A 2500-locus bin map of wheat homoeologous group 5 provides new insights on gene distribution and colinearity with rice. *Genetics* **168**: 665–676.
- LIU, C. J., K. M. DEVOS, C. N. CHINYOY, M. D. ATKINSON and M. D. GALE, 1992 Homoeologous translocations between group 4, 5 and 7 chromosomes in wheat and rye. *Theor. Appl. Genet.* **83**: 305–312.
- MICKELSON-YOUNG, L., T. R. ENDO and B. S. GILL, 1995 A cytogenetic ladder-map of the wheat homoeologous group 4 chromosomes. *Theor. Appl. Genet.* **90**: 1007–1011.
- MUNKVOLD, J. D., R. A. GREENE, C. E. BERMUDEZ-KANDIANIS, C. M. LA ROTA, H. EDWARDS *et al.*, 2004 Group 3 chromosome bin maps of wheat and their relationship to rice chromosome 1. *Genetics* **168**: 639–650.
- NARANJO, T., 1990 Chromosome structure of durum wheat. *Theor. Appl. Genet.* **79**: 397–400.
- NARANJO, T., A. ROCA, P. G. GOICOECHEA and R. GIRALDEZ, 1987 Arm homoeology of wheat and rye chromosomes. *Genome* **29**: 873–882.
- NELSON, J. C., M. E. SORRELLS, A. E. VAN DEYNZE, Y. H. LU and M. ATKINSON, 1995 Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. *Genetics* **141**: 721–731.
- PENG, J. H., H. ZADEH, G. R. LAZO, J. P. GUSTAFSON, S. CHAO *et al.*, 2004 Chromosome bin map of expressed sequence tags in homoeologous group 1 of hexaploid wheat and homoeology with rice and Arabidopsis. *Genetics* **168**: 609–623.
- QI, L. L., B. ECHALIER, B. FRIEBE and B. S. GILL, 2003 Molecular characterization of a set of wheat deletion stocks for using in chromosome bin mapping of ESTs. *Funct. Integr. Genomics* **3**: 39–55.
- RANDHAWA, H. S., M. DILBIRLI, D. SIDHU, M. ERAYMAN, D. SANDHU *et al.*, 2004 Deletion mapping of homoeologous group 6-specific wheat expressed sequence tags. *Genetics* **168**: 677–686.
- RILEY, R., H. COUCOLI and V. CHAPMAN, 1967 Chromosomal interchanges and the phylogeny of wheat. *Heredity* **22**: 233–247.
- RODRIGUEZ MILLA, M. A., E. BUTLER, A. R. HUETE, C. F. WILSON, O. D. ANDERSON *et al.*, 2002 Expressed sequence tag-based gene expression analysis under aluminum stress in rye. *Plant Physiol.* **130**: 1706–1716.
- SEARS, E. R., 1954 *The Aneuploids of Common Wheat*. Bull. 572, University of Missouri Agricultural Experiment Station, Columbia, MO.
- SEARS, E. R., 1966 Nullisomic-tetrasomic combinations in hexaploid wheat, pp. 29–45 in *Chromosome Manipulation and Plant Genetics*, edited by R. RILEY and K. R. LEWIS. Oliver & Boyd, Edinburgh.
- SORRELLS, M. E., M. LA ROTA, C. E. BERMUDEZ-KANDIANIS, R. A. GREENE, R. KANTETY *et al.*, 2003 Comparative DNA sequence analysis of wheat and rice genomes. *Genome Res.* **13**: 1818–1827.
- TSUJIMOTO, H., 1993 Molecular cytological evidence for gradual telomere synthesis at the broken chromosome ends in wheat. *J. Plant Res.* **106**: 239–244.
- VAN DER HOEVEN, R., C. RONNING, J. GIOVANNONI, G. MARTIN and S. TANKSLEY, 2002 Deduction about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *Plant Cell* **14**: 1441–1456.
- VAN DEYNZE, A. E., J. C. NELSON, E. S. YGLESIAS, S. E. HARRINGTON and D. P. BRAGA, 1995 Comparative mapping in grasses. Wheat relationships. *Mol. Gen. Genet.* **248**: 744–754.
- WERNER, J. E., T. R. ENDO and B. S. GILL, 1992 Toward a cytogenetically based physical map of the wheat genome. *Proc. Natl. Acad. Sci. USA* **89**: 11307–11311.
- ZHANG, D., D. W. CHOI, S. WANAMAKER, R. D. FENTON, A. CHIN *et al.*, 2004 Construction and evaluation of cDNA libraries for large-scale expressed sequence tag sequencing in wheat (*Triticum aestivum* L.). *Genetics* **168**: 595–608.

Communicating editor: J. P. GUSTAFSON

