

J. H. Peng · Nora L. V. Lapitan

## Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers

Received: 21 July 2004 / Revised: 27 October 2004 / Accepted: 30 October 2004 / Published online: 14 January 2005  
© Springer-Verlag 2005

**Abstract** EST-derived microsatellites or simple sequence repeats (eSSR) occur in expressed sequence tags (EST). Here we report characteristics of eSSRs in the wheat genome, construction of consensus chromosome bin maps of SSR-containing ESTs ( $_{SSR}$ ESTs), and development of eSSR markers for the 21 wheat chromosomes. A Perl script known as MISA was used to identify eSSRs in wheat ESTs available in the database <http://wheat.pw.usda.gov/cgi-bin/acc/search/wEST>). Among 492,832 ESTs from the database, 36,520 (7.41%) contained 43,598 eSSRs. This is equivalent to 1 eSSR per 5.46 kb EST sequence. About 60% of the eSSRs were trinucleotides, 19.7% were mononucleotide, 16.7% were dinucleotides, and the remaining ~3% consisted of tetra-, penta-, and hexanucleotides. Among the identified eSSRs, (CCG/CGG) $_n$  is the most frequent (20.5%) followed by (A/T) $_n$  at 13.6%, (AAC/GTT) $_n$  at 11.7%, and (AG/CT) $_n$  at 8.7%. Among ESTs previously mapped to wheat chromosome bins, a total of 1,010 eSSR loci were derived from 341  $_{SSR}$ ESTs. Consensus chromosome bin maps showing the chromosome locations of  $_{SSR}$ ESTs, SSR sequence motif, and cDNA library were constructed. A  $\chi^2$  test indicated that the distribution pattern of eSSR loci was generally similar to that of the original mapped ESTs in the wheat genome. Forty-eight  $_{SSR}$ ESTs were converted into PCR-based eSSR markers, and 266 eSSR loci were mapped to specific chromosome arms using wheat cytogenetic stocks. The average polymorphism information content ( $0.45\pm 0.16$ ) of eSSR markers was lower than that reported for genomic SSRs ( $0.54\pm 0.19$ ), but higher than RFLPs ( $0.30\pm 0.27$ ). The eSSR markers were transfer-

able among related Triticeae species, *Triticum aestivum*, *T. durum*, *T. dicoccoides*, *Hordeum spontaneum*, *H. vulgare*, and *Secale cereale*. The results confirm the presence of SSRs in expressed genes of wheat and demonstrate another application of ESTs in genomics research. eSSRs will be useful for gene tagging, gene cloning, and comparative genomics studies of cereal crops.

**Keywords** Expressed sequence tags (EST) · EST-derived simple sequence repeats (eSSR) · Molecular markers · Chromosome bin maps · Triticeae

### Introduction

The genomes of all eukaryotes contain iterations of 1- to 6-bp nucleotide motifs. This class of DNA sequences is known as microsatellites (Litt and Luty 1989) or simple sequence repeats (SSRs; Tautz et al. 1986). SSRs are abundant across genomes, show high levels of polymorphism, and are usually considered as evolutionarily neutral DNA sequences. These are randomly or nearly randomly distributed across the euchromatic genome (reviewed in Li et al. 2002). SSRs are ubiquitously interspersed in eukaryotic genomes and can be frequently used as highly variable and multi-allelic PCR-based genetic markers (Brown et al. 1996). In wheat, Röder et al. (1998) developed the first molecular map consisting of 279 loci of genomic SSRs. These SSR markers have been extensively applied to molecular genetic studies and breeding research in wheat (Peng et al. 1999, 2000a,b,c, 2003; Peng 2000). Recently, many more SSRs were mapped and high density microsatellite consensus maps (Somers et al. 2004) or chromosome deletion bin maps (Sourdille et al. 2004) have become available. However, development of SSR markers is expensive, labor intensive and time consuming, especially if they are developed from genomic libraries (Röder et al. 1998; Varshney et al. 2002).

Recently, SSRs have been identified in genes and expressed sequence tags (ESTs; reviewed in Li et al. 2004). ESTs are single-pass sequence segments of expressed genes

**Electronic Supplementary Material** Supplementary material is available for this article at <http://dx.doi.org/10.1007/s10142-004-0128-8>.

J. H. Peng · N. L. V. Lapitan (✉)  
Department of Soil and Crop Sciences  
Colorado State University,  
Fort Collins, CO, 80523-1170, USA  
e-mail: nlapitan@lamar.colostate.edu  
Tel.: +1-970-4911921  
Fax: +1-970-4910564

(Adams et al. 1991). They are derived from cDNA libraries made from multiple tissues under various treatments and used to identify as many genes as possible in an organism. ESTs have been developed for numerous organisms and sequences are deposited in GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>) and in specific species databases. Among plants, wheat currently has the largest number of ESTs. There are over 500,000 wheat ESTs deposited in the database as of 15 July 2004 (<http://wheat.pw.usda.gov/wEST>). This collection includes 90,016 hexaploid wheat ESTs developed by a group of United States researchers funded by the National Science Foundation (<http://wheat.pw.usda.gov/NSF/project/>). Out of these ESTs, 8,241 were physically mapped to chromosomes or chromosome bins using wheat aneuploid stocks and Southern hybridization (Hossain et al. 2004; Lazo et al. 2004; Linkiewicz et al. 2004; Miftahudin et al. 2004; Munkvold et al. 2004; Peng et al. 2004; Qi et al. 2004; Randhawa et al. 2004; Conley et al. 2004; <http://wheat.pw.usda.gov/NSF/project/>).

SSRs have been identified in the ESTs (designated eSSR hereafter) of wheat (Eujayl et al. 2002; Kantety et al. 2002; Varshney et al. 2002; Gao et al. 2003; Nicot et al. 2004), and other cereals (Hackauf and Wehling 2002; Kantety et al. 2002; Varshney et al. 2002; Gao et al. 2003; Thiel et al. 2003). The frequency of eSSRs was 1/6.2 kb–1/17.4 kb in wheat, 1/3.9 kb–1/11.8 kb in rice, and 1/7.5 kb–1/28.3 kb in maize (Varshney et al. 2002; Gao et al. 2003). The proportion of SSR-containing ESTs (designated <sub>SSR</sub>ESTs hereafter) in cebreedreals varies from 1.5% in maize to 6.6% in rye (Hackauf and Wehling 2002; Kantety et al. 2002). However, the frequency of eSSRs was variable in different studies even for the same species. Moreover, information on the physical location of eSSRs was not available in previous studies. The objectives of this study were (1) to analyze the frequency of eSSRs in the wheat genome, (2) to construct consensus chromosome bin maps of eSSRs in wheat, and (3) to develop and evaluate eSSR markers for application in breeding and comparative genomics.

## Materials and methods

### Identification of SSRs in expressed sequence tags

The Perl script MISA (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) was used to identify eSSRs in a wheat EST database consisting of about half a million ESTs (<http://wheat.pw.usda.gov/cgi-bin/ace/search/wEST>). In the present study, eSSRs were considered to contain motifs one to six nucleotides in size. The minimum repeat unit was defined as ten for mononucleotides, six for dinucleotides, and five for all the higher order motifs including tri-, tetra-, penta-, and hexanucleotides. Frequency of eSSR refers to kilo-base pairs of EST sequences containing one SSR. Percentage of <sub>SSR</sub>EST refers to number of <sub>SSR</sub>ESTs per 100 ESTs examined. Percentage of multi-SSR EST refers to number of ESTs containing more than one SSR per 100

<sub>SSR</sub>ESTs. Percentage of compound eSSR refers to number of eSSRs present in compound formation ( $\geq 2$  SSRs interrupted by  $\leq 100$  bases in an identified <sub>SSR</sub>EST) per 100 identified eSSRs.

### Chromosome localization of SSR-containing ESTs

Eight thousand two hundred and forty-one ESTs were physically mapped to chromosome bins of wheat using cytogenetic stocks (Hossain et al. 2004; Lazo et al. 2004; Linkiewicz et al. 2004; Miftahudin et al. 2004; Munkvold et al. 2004; Peng et al. 2004; Qi et al. 2004; Randhawa et al. 2004; Conley et al. 2004; <http://wheat.pw.usda.gov/NSF/project/progress.html>). A bin is a chromosome segment flanked by two deletion breakpoints (Endo and Gill 1996). The MISA software (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) was used to identify <sub>SSR</sub>ESTs in this set of wheat ESTs. In order to construct a reliable consensus physical map of eSSRs and detect patterns of genomic distribution of eSSRs, we first verified all the mapping data on the project webpage ([http://wheat.pw.usda.gov/cgi-bin/westsq/locus\\_rev.cgi](http://wheat.pw.usda.gov/cgi-bin/westsq/locus_rev.cgi)) by careful inspection of the online mapping images and included only those that we were able to confirm. The chromosome locations of <sub>SSR</sub>ESTs are graphically presented in consensus bin maps for the seven homoeologous chromosome groups.

### Development of eSSR markers and evaluation for use in mapping and comparative genomics

Three <sub>SSR</sub>ESTs were chosen from each arm of the seven consensus maps and an additional six from 1S for primer design. The software program, Primer3 ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)), was used to design primer pairs flanking eSSRs in the selected 48 <sub>SSR</sub>ESTs. A set of 59 aneuploid stocks including 21 nulli-tetrasomic and 38 ditelosomic lines of Chinese Spring (CS) background (Sears 1969) was used to map eSSR marker loci to chromosome arms. A set of 12 Triticeae cultivars or accessions including seven common wheat cultivars/accessions, two tetraploid accessions (*T. dicoccoides*, *T. durum*), two barley accessions and one rye cultivar (Table 1) were used to test polymorphism and transferability among Triticeae species of wheat eSSR markers. PCR procedures were according to Peng et al. (1999), and PCR products were detected using 3% agarose gel with EtBr staining or 5% sequencing gels with silver staining.

### Statistical analysis

Distributions of <sub>SSR</sub>ESTs among three genomes, seven homoeologous groups, 21 chromosomes, and between the long and short arms of each chromosome were compared with that of the original mapped ESTs (Hossain et al. 2004; Lazo et al. 2004; Linkiewicz et al. 2004; Miftahudin et al.

**Table 1** Triticeae materials used for polymorphism test of wheat EST-derived simple sequence repeat (eSSR) markers

Genotype	Species	Genome	Origin
Chinese spring	<i>Triticum aestivum</i>	AABBDD	China
PI294994	<i>Triticum aestivum</i>	AABBDD	Bulgaria
PI262660	<i>Triticum aestivum</i>	AABBDD	Azerbaijan
CI2401	<i>Triticum aestivum</i>	AABBDD	Tajikistan
CO970547	<i>Triticum aestivum</i>	AABBDD	USA
Above	<i>Triticum aestivum</i>	AABBDD	USA
Glupro	<i>Triticum aestivum</i>	AABBDD	USA
Langdon	<i>Triticum durum</i>	AABB	USA
Hermon H52	<i>Triticum dicoccoides</i>	AABB	Israel
Mona	<i>Hordeum vulgare</i>	HH	Sweden
Wadi Qilt 23–39	<i>Hordeum spontaneum</i>	HH	Israel
Turkey 67	<i>Secale cereale</i>	RR	Turkey

2004; Munkvold et al. 2004; Peng et al. 2004; Qi et al. 2004; Randhawa et al. 2004; Conley et al. 2004; <http://wheat.pw.usda.gov/NSF/project/progress.html>). Correction factors were calculated as the percentage of ESTs in the specific genome, homoeologous groups, chromosome, and chromosome arms out of the total mapped ESTs. The reference values were obtained by multiplying the corresponding correction factors with total SSR-ESTs mapped to genome, homoeologous group, chromosome, and chromosome arms. The distribution difference was tested by  $\chi^2$  analysis. A software program, STATISTICA (StatSoft 1996), was used to conduct  $\chi^2$  analysis.

Polymorphism information content (PIC) is a measure of the polymorphism-detecting capacity of markers. PIC refers to the value of a marker for detecting polymorphism within a population, and depends on the number and frequency distribution of detectable alleles. We used a simplified formula (Anderson et al. 1993) to evaluate the PIC of eSSR markers, assuming that the inbred wheat lines are homozygous. The calculation of PIC for the  $i$ th marker is:

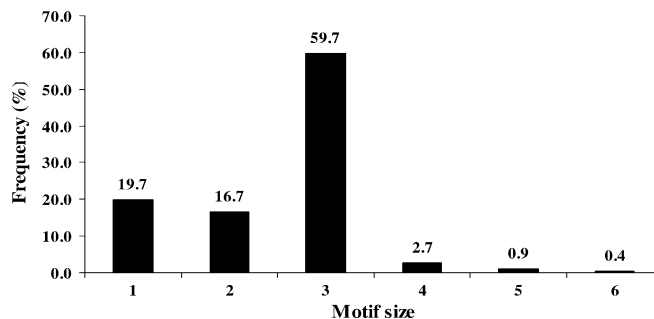
$$PIC_i = 1 - \sum p_{ij}^2 (j = 1, 2, \dots, n)$$

where  $p_{ij}$  is the frequency of the  $j$ th pattern for the  $i$ th marker and the summation extends over  $n$  patterns.

## Results

### Frequency of eSSRs in wheat

We examined nearly 500,000 wheat gene segments (ESTs) comprising 238.05 Mb DNA sequence and detected a total of 43,598 SSRs. Taking into account those ESTs which contained more than one eSSR, the total number of SSR-ESTs was 36,520 (7.41%). The frequency of occurrence for eSSRs was one eSSR in every 5.46 kb of EST sequences. Out of the total SSR-ESTs, 4,647 (12.72%) contained two or more eSSRs. Among the total eSSRs, 5,133 (11.77%) appeared to have compound formation.

**Fig. 1** Frequency distribution of EST-derived simple sequence repeats (SSRs) based on motif size

### Distribution of wheat eSSRs based on motif sizes

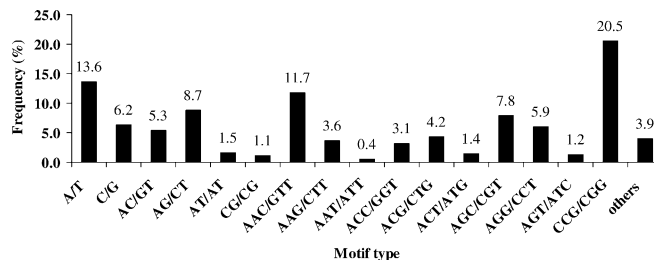
Figure 1 shows the distribution pattern of eSSRs in terms of motif size. Among the eSSRs detected, the trinucleotides are the most frequent, with a frequency of 59.7%, followed by mono- (19.7%) and dinucleotides (16.7%). The tetranucleotide eSSRs have a much lower frequency (2.7%) than those with mono-, di-, and trinucleotide motifs. The frequencies of eSSRs with motif sizes of five or six are less than 1%.

### Distribution of wheat eSSRs based on motif types

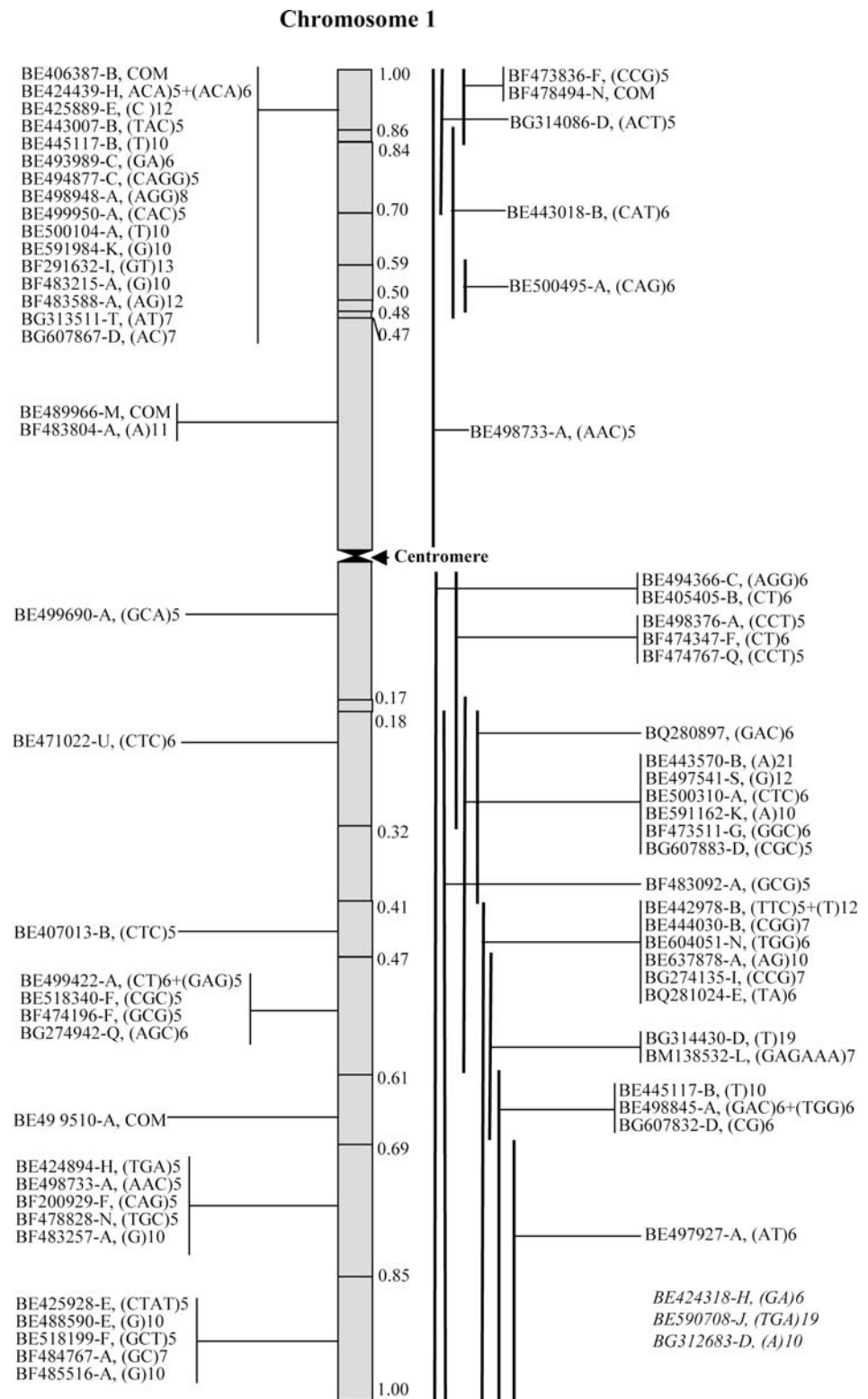
The eSSRs identified in the present study could be basically characterized by 16 types of motifs (Fig. 2). In general, the eSSRs were found to be unevenly distributed across motif types. The motif CCG/CGG had the highest frequency of 20.5% followed by the motifs A/T (13.6%), AAC/GTT (11.7%), AG/CT (8.7%), AGC/CGT (7.8%), C/G (6.2%), AGG/CCT (5.9%), and AC/GT (5.3%). The other eight types of motifs possessed a frequency of <5%.

### Physical mapping of wheat SSR-ESTs

Among 8,241 wheat ESTs physically mapped by the ten United States laboratories including ours as of February 2004 (<http://wheat.pw.usda.gov/NSF/project/>), 341 were SSR-ESTs. A total of 1,010 EST loci derived from these 341 SSR-ESTs were mapped to specific bins, arms or chromosomes of seven homoeologous groups in wheat. The consensus physical (chromosome bin) maps showing chro-

**Fig. 2** Frequency distribution of EST-derived SSRs based on motif sequence type

**Fig. 3** Consensus bin maps of SSR/ESTs for wheat homoeologous group 1, 2, 3, 5, 6, and 7. The accession numbers of SSR/ESTs mapped to specific bins are shown on the *left side*, and those mapped to combined bins are shown on the *right side* of the chromosomes. Vertical lines on the right side span the range of combined bins. Letters after EST accession numbers refer to the code for cDNA libraries: *A* wheat pre-anthesis spike, *B* wheat etiolated seedling root, *C* *Secale cereale* anther, *D* *Triticum monococcum* early reproductive apex, *E* wheat un-stressed seedling shoot, *F* wheat vernalized crown, *G* wheat 5–15 DAP spike, *H* wheat endosperm, *I* *Aegilops speltoides* anther, *J* wheat 20–45 DAP spike, *K* wheat heat-stressed flag leaf, *L* wheat *Fusarium graminearum*-infected spike, *M* wheat cold-stressed seedling, *N* wheat drought-stressed leaf, *O* wheat heat-stressed spike, *P* durum wheat whole plant, *Q* wheat salt-stressed crown, *R* wheat un-stressed root tip, *S* wheat heat-stressed seedling, *T* wheat salt-stressed sheath, *U* wheat drought-stressed seedling, and *V* wheat dormant embryo. Following the code is the eSSR motif sequence. *COM* Compound formation,  $\geq 2$  SSRs interrupted by  $\leq 100$  bases in an identified SSR/ESTs. SSR/ESTs in *italics* are those which only mapped to specific chromosomes, and not to chromosome arms or bins

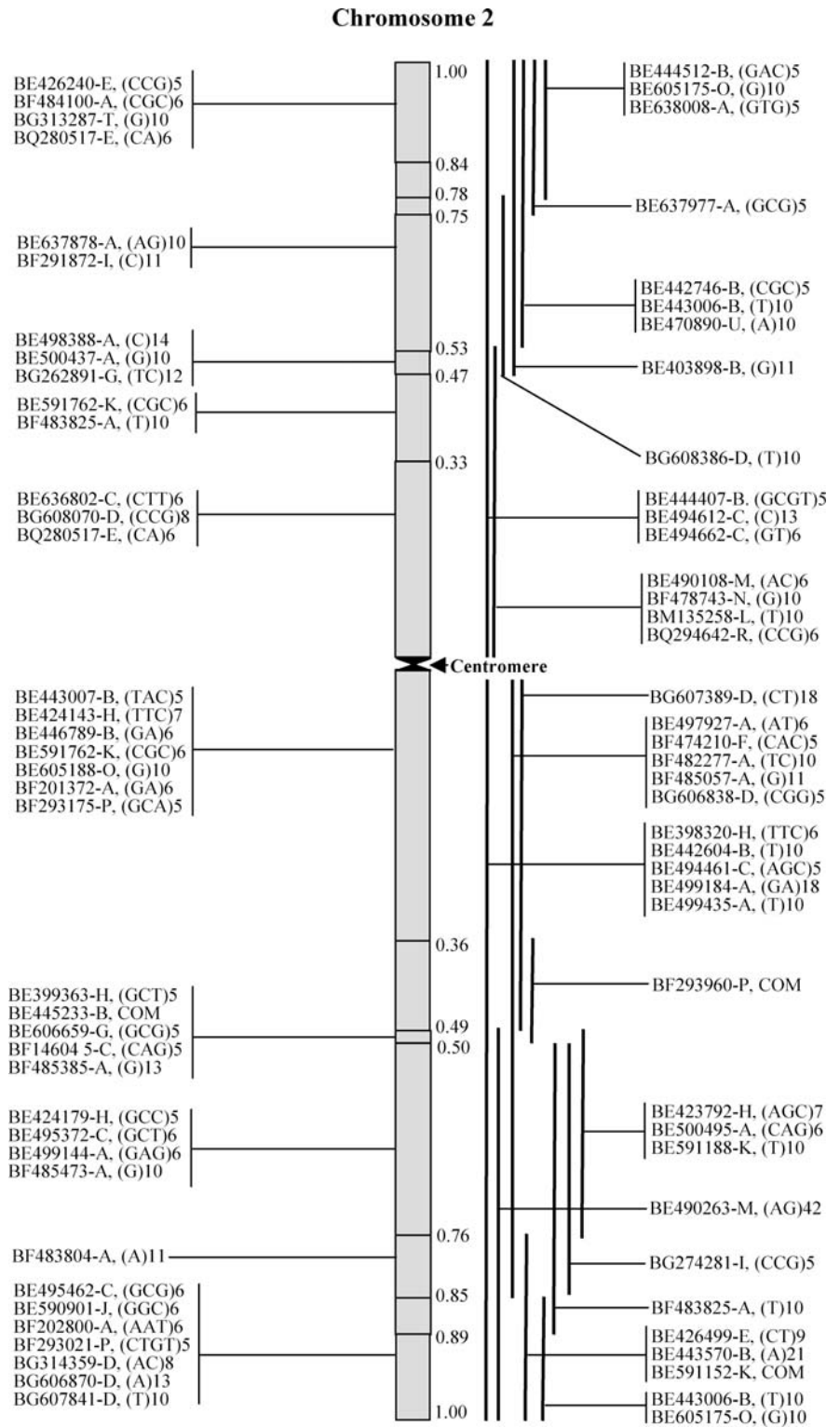


mosomal locations of SSR/ESTs, the corresponding cDNA libraries and the containing SSR sequences were constructed and are presented in Figs. 3 and 4.

Due to the significant structural changes in group 4 chromosomes, bin maps for chromosome 4A and a consensus chromosome 4 incorporating 4B and 4D are pres-



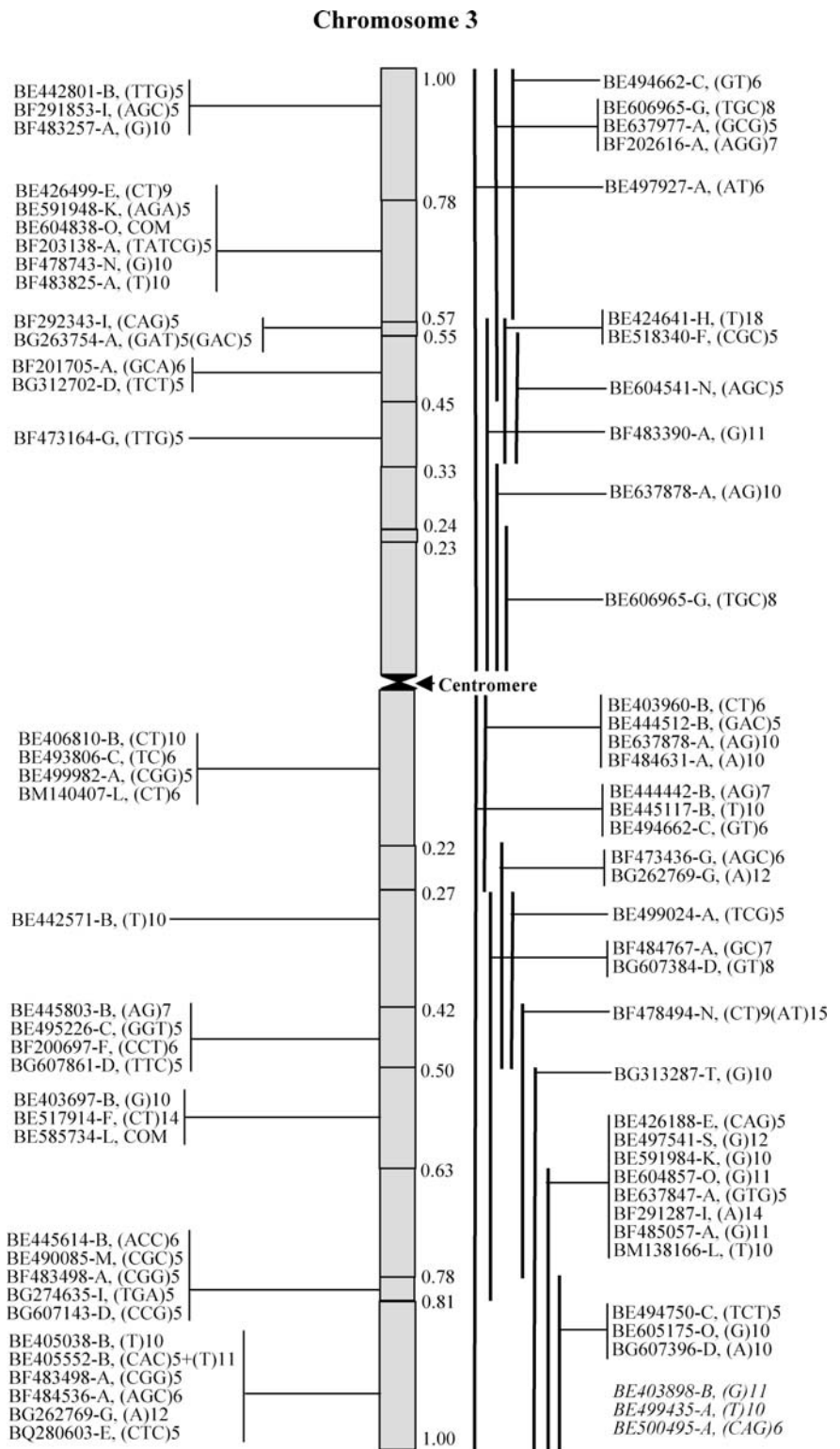
Fig. 3 (continued)



ented separately (Fig. 4). Among 15  $_{SSR}$ ESTs mapped to chromosomes 4A, 4B and/or 4D, 14 detected structural changes between 4A and 4B/4D. The pericentric inversion in 4A established arm homoeologies as 4AS =4BL =4DL, and 4AL =4BS =4DS. This result further confirms the

findings of structure changes of group 4 chromosomes based on chromosome morphology and C-banding patterns (Endo and Gill 1984; Gill et al. 1991) and molecular mapping (Mickelson-Young et al. 1995).

Fig. 3 (continued)

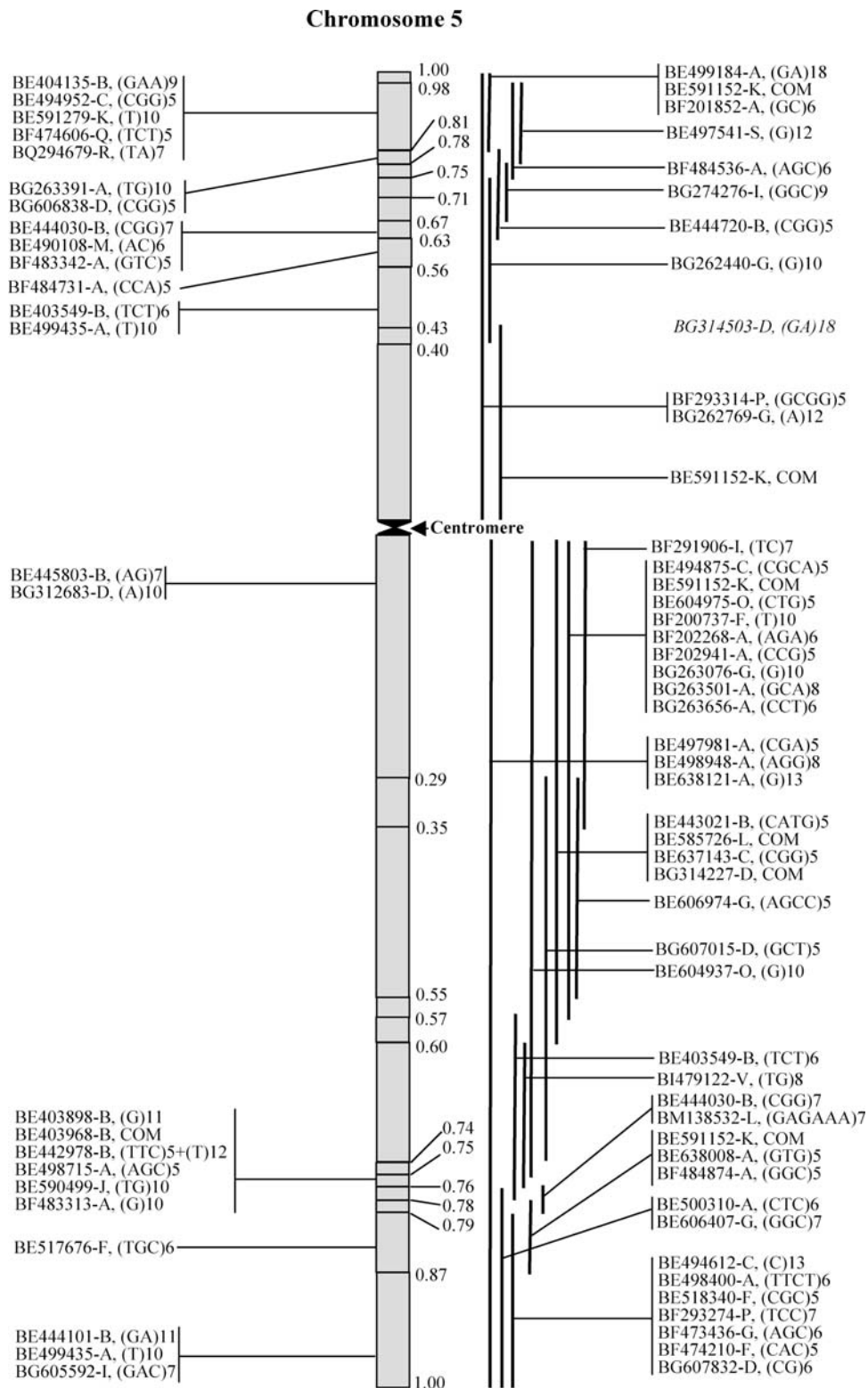


The mapped  $SSR$ -ESTs covered the whole genome of hexaploid wheat. In general, however, there were more  $SSR$ -ESTs in the distal regions than in the proximal regions. The distribution patterns of  $SSR$ -ESTs or eSSRs were variable among chromosome groups and between chromosome arms (Figs. 3, 4).

Distribution of  $SSR$ -ESTs among genomes, chromosomes and between chromosome arms

The number of  $SSR$ -EST loci mapped to the B genome was 17.7% and 16.6% higher than that mapped to the A and D genomes, respectively (Table 2). In comparison to the

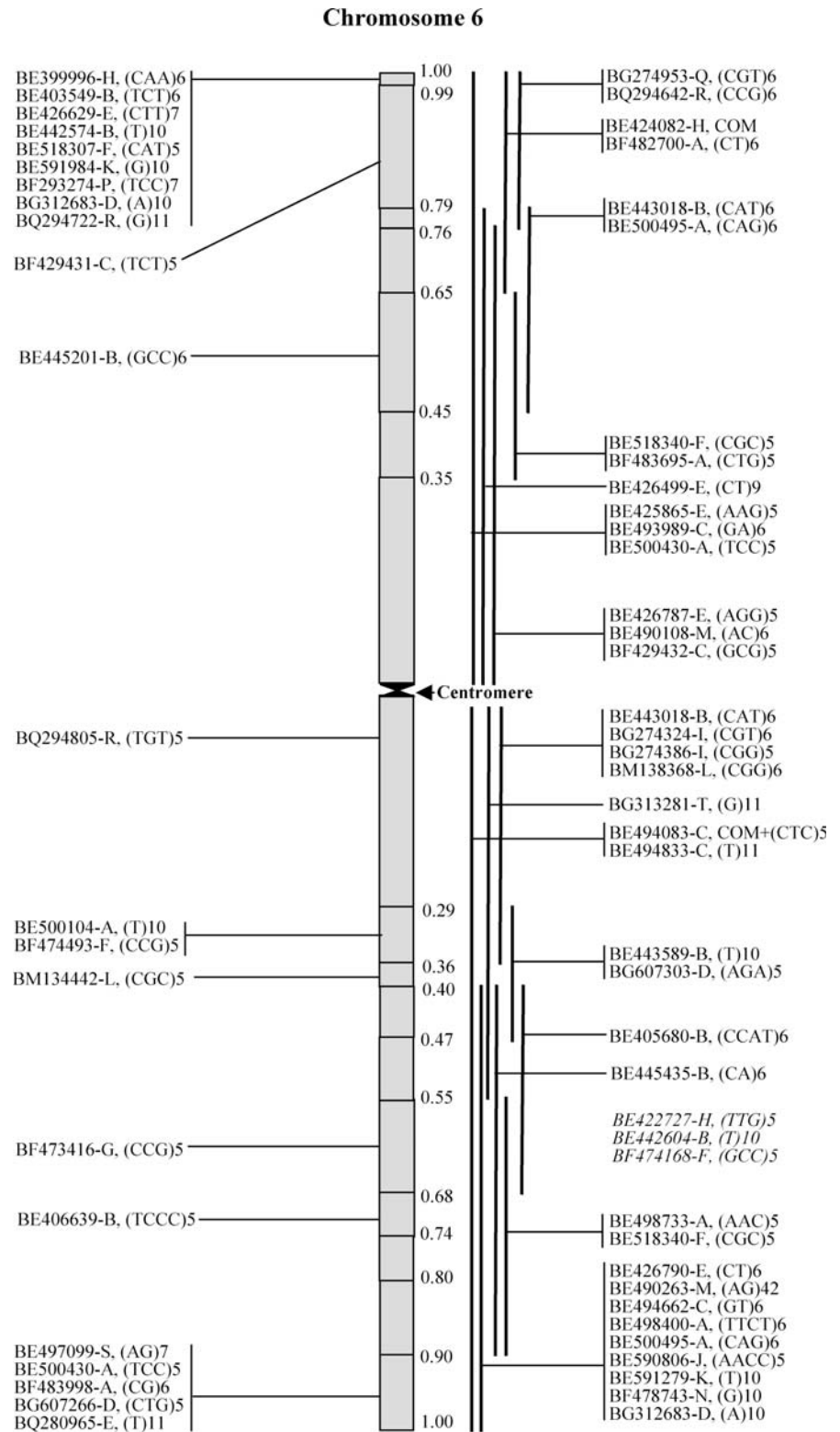
Fig. 3 (continued)



reference number of  $_{SSR}$ ESTs based on the original mapped ESTs, the number of  $_{SSR}$ ESTs in the B genome was in slight excess, while the number of  $_{SSR}$ ESTs in the A and D genomes was slightly less. However, the deviation was not significant ( $P = 0.823$ ) as indicated by the  $\chi^2$  test (Table 2).

As shown in Table 2, the number of  $_{SSR}$ EST loci mapped to homoeologous groups 1, 2, 3, 5 and 7 exceeded the reference values based on the original mapped ESTs, while the reverse was true for groups 4 and 6. The  $\chi^2$  test indicated that the deviation was nearly significant ( $P = 0.053$ ).

Fig. 3 (continued)



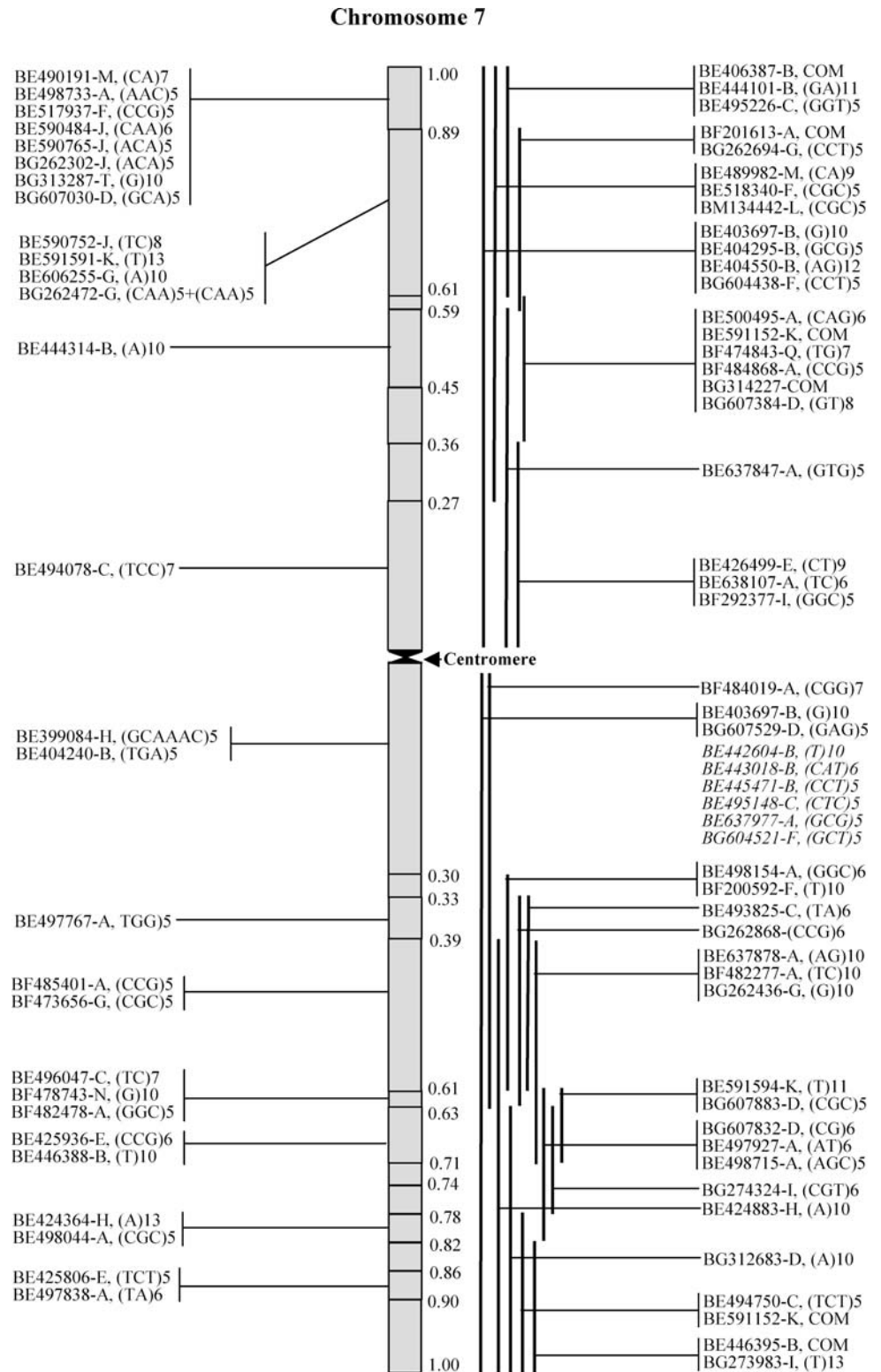
The number of  $_{SSR}EST$  loci mapped to chromosomes 1B, 3B, 4A, 6B and 7D exceeded the reference values based on the original mapped ESTs. The number of  $_{SSR}EST$  loci was similar to or lower than the reference value for the other 16 chromosomes. However, the devia-

tion was not significant ( $P=0.963$ ) as indicated by the  $\chi^2$  test (Table 2).

For 16 of the 21 chromosomes, the number of  $_{SSR}EST$  loci mapped to the long arm exceeded that mapped to the short arm. For the remaining five chromosomes, the short



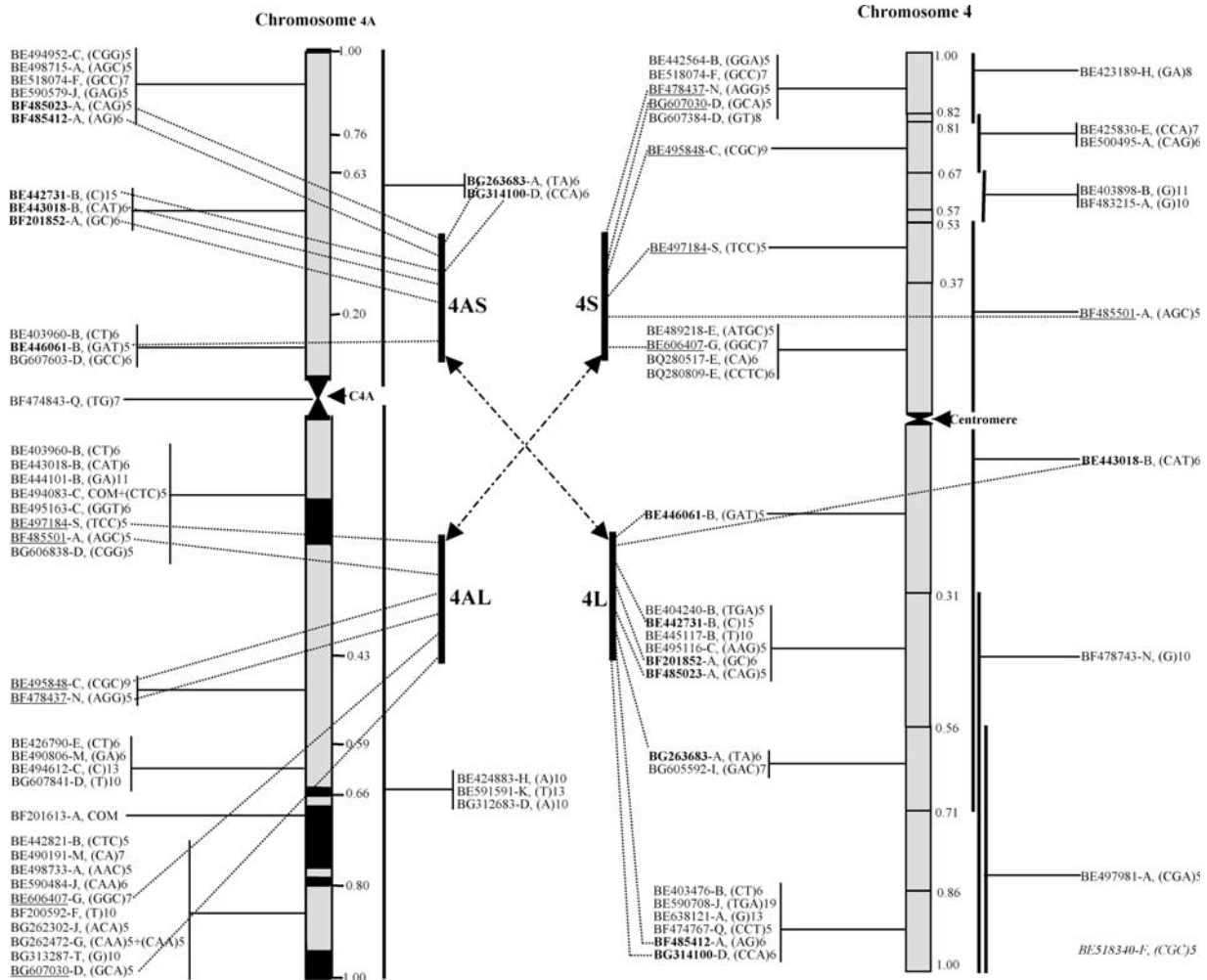
Fig. 3 (continued)



arm had the same or a larger number of  $_{SSR}$ ESTs than the long arm. Furthermore,  $\chi^2$  test indicated that the deviation from reference values based on the original mapped ESTs was significant ( $P < 0.05$ ) only for chromosomes 5B, 5D, and 7B (Table 2).

#### Experimental evaluation of eSSRs

To evaluate the usefulness of the identified eSSRs, primers were designed for 48 selected  $_{SSR}$ ESTs, and the amplified loci were mapped to chromosome arms using wheat cytogenetic stocks and a PCR approach. The Electronic Sup-



**Fig. 4** Chromosome bin map of  $SSR/ESTs$  for wheat homoeologous group 4. The consensus chromosome 4 integrated 4B and 4D.  $SSR/ESTs$  in **bold** and underlined letters detected structure changes between 4A and 4B/4D (**bold-fonts**  $SSR/ESTs$  were mapped to 4AS and

4L arms, the underlined-font  $SSR/ESTs$  were mapped to 4AL and 4S arms). BE518074 was the only  $SSR/EST$  mapped to similar arms, 4AS and 4S. All other notations are the same as in Fig. 3

plementary Material lists 266 eSSR loci derived from the 48 eSSR primer pairs and mapped to specific chromosome arms of wheat. The number of mapped loci ranged from 1 to 28 and an average of 5.5 loci were mapped for each primer pair (Table 3; Electronic Supplementary Material). Out of the 48 eSSR markers, 43 (~90%) amplified at least one band from the expected chromosome arm, based on mapping information of the original  $SSR/ESTs$  (Figs. 3, 4, Table 3). Five did not amplify loci from the expected chromosomes, possibly due to the imperfect primer design or other unknown reasons. The high coincidence in chromosome location between  $SSR/ESTs$  and eSSR markers is encouraging for the application of EST-derived markers in genetic studies.

As shown in Table 3, the number (266) of eSSR loci was obviously larger than loci (165) detected via the RFLP approach (designated eRFLP hereafter). For 25 of the 48  $SSR/ESTs$ , the number of eSSR loci exceeded that of the corresponding eRFLP loci. For 17 of the 48  $SSR/ESTs$ , the number of eSSR loci was smaller than that of the corresponding eRFLP loci. The number of eSSR loci was

equal to that of eRFLP for only 6 of the 48  $SSR/ESTs$ . However, most (79.2%) of the excessive eSSR loci relative to eRFLP loci were derived from 6 (12.5%)  $SSR/ESTs$ , i.e., BE494877, BE495462, BE495116, BE497184, BF478437 and BG262868. The other 42 (87.5%)  $SSR/ESTs$  only contributed 20.8% to the loci difference between eSSR and eRFLP (Table 3). Thus, only a small number of  $SSR/ESTs$  significantly contributed to the number difference between mapped eSSR and eRFLP loci.

The eSSRs were also evaluated for polymorphism and transferability to other cultivars or other Triticeae species. Figure 5 is an example of a PCR profile of wheat eSSR marker *Xcwm20* on 12 accessions/genotypes of Triticeae species listed in Table 1. As a measure of the informativeness of eSSR markers, PICs were obtained for 15 eSSR markers with loci covering all the seven homoeologous chromosome groups, by testing a set of seven hexaploid wheat lines used in our wheat genomics and breeding programs (Table 1). Each band in each accession was considered as an allele of a genetic locus for each eSSR marker. As shown in Table 4, the PIC of eSSR markers

**Table 2** Distribution of SSR-containing ESTs among genomes, chromosomes and between chromosome arms [ $T = A + B + D$ ;  $O$  observed,  $R$  reference value of  $ssrEST$  number which was based on the original number of mapped ESTs (Peng et al. 2004; Qi et al. 2004; [http://wheat.pw.usda.gov/cgi-bin/westsql/map\\_locus\\_rev.cgi](http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus_rev.cgi))]

Homoeologous group	Genome	Chromosome	Chromosome arm					
			Short			Long		
			O	R	O	R	O	R
1	A	35	42.60	10	11.20	22	20.80	0.20
	B	61	54.54	26	25.01	35	35.99	0.07
	D	46	46.86	13	13.44	29	28.56	0.02
	T	142	136.35					
2	A	50	49.30	15	17.50	35	32.50	0.55
	B	61	62.90	26	26.84	35	34.16	0.05
	D	59	57.80	24	23.60	35	35.40	0.01
	T	170	157.56					
3	A	49	48.43	12	16.80	30	25.20	2.29
	B	69	63.46	22	28.56	46	39.44	2.60
	D	49	55.11	13	18.33	34	28.67	2.54
	T	167	159.58					
4	A	53	45.20	15	15.50	35	34.50	0.02
	B	27	32.77	10	9.00	15	16.00	0.17
	D	33	36.16	17	12.40	14	18.60	2.84
	T	113	136.35					
5	A	50	46.08	14	12.48	34	35.52	0.25
	B	56	56.16	21	13.50	33	40.50	5.56*
	D	38	41.76	14	8.64	22	27.36	4.38*
	T	144	139.38					
6	A	35	38.28	19	13.94	15	20.06	3.11
	B	50	45.24	18	18.72	30	29.28	0.05
	D	31	32.48	17	12.47	12	16.53	2.89
	T	116	140.39					
7	A	45	48.98	19	17.64	17	18.36	0.21
	B	49	53.72	10	18.40	36	27.60	6.39*
	D	64	55.30	31	29.76	31	32.24	0.10
	T	158	139.38					
Sum	A	317	323.20					
	B	373	363.60					
	D	320	323.20					
	T	1010						
		$\chi^2_{group}=12.44$ ( $p=0.053$ )		$\chi^2_{genome}=0.39$ ( $p=0.823$ )		$\chi^2_{chromosome}=10.26$ ( $p=0.963$ )		

\*Significant at the level of 0.05

**Table 3** Chromosome location coincidence between EST-derived simple sequence repeats (*eSSR*) and eRFLP loci (detected via the RLFP approach) of SSR-containing ESTs in wheat

EST accession	Number of loci		Chromosome location <sup>a</sup>		Number of unmapped eRFLP bands <sup>b</sup>
	eSSR	eRFLP	eSSR	eRFLP	
BF200929	7	3	<b>1AL, 1BL, 1DL</b>	<b>1AL, 1BL, 1DL</b>	2
BE489966	4	4	1AL, 2AS, 3BL, 3DL	1AS, 1BS, 1DS	N/A
BG607867	9	3	<b>1AS, 1DL, 3AL, 3BL, 3BS, 5DL, 6BL, 6DS, 6DL</b>	<b>1AS, 1BS, 1DS</b>	1
BE499690	6	3	<b>1AL, 2AL, 3BS</b>	<b>1AL, 1BL, 1DL</b>	N/A
BG274942	9	3	<b>1DL, 2BS, 2AL, 2BL, 2DL, 3AL, 4AL, 6AS</b>	1AL, 1BL, <b>1DL</b>	N/A
BF483588	3	3	<b>1AS, 1BS, 1DS</b>	<b>1AS, 1BS, 1DS</b>	0
BE500104	7	6	<b>1BS, 6DS, 6AL, 6BL, 6DL</b>	1AS, 1BS, 1DS, <b>6AL, 6BL, 6DL</b>	0

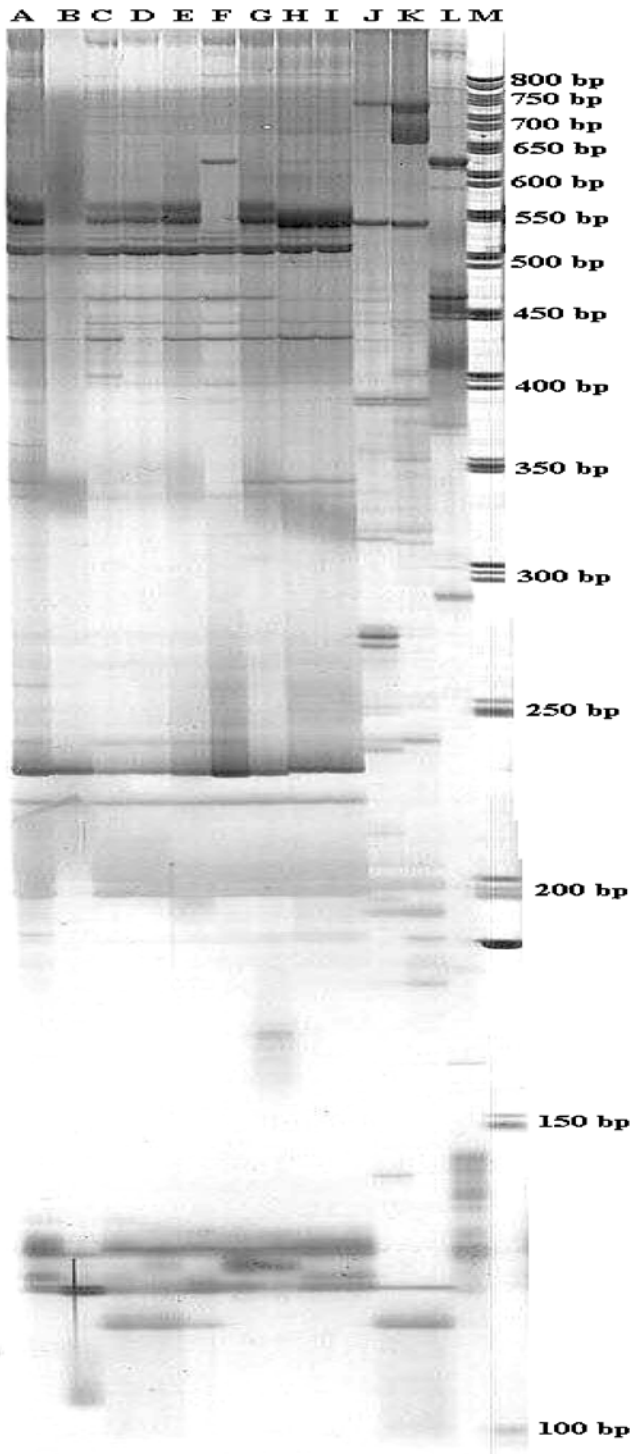
**Table 3** (continued)

EST accession	Number of loci		Chromosome location <sup>a</sup>		Number of unmapped eRFLP bands <sup>b</sup>
	eSSR	eRFLP	eSSR	eRFLP	
BE443007	5	5	<b>1DS, 2AL, 3BS, 7AL, 7BL</b>	<i>1BS, 1DS, 2AL, 2BL</i>	4
BE494877	19	3	<b>1AS, 1BS, 1AL, 1DL, 2AL, 3AL, 3BL, 4DL, 5AL, 5DL, 6BS, 6DS</b>	<b>1AS, 1BS, 1DS</b>	>3
BF483804	1	6	<b>1DS</b>	<i>1AS, 1BS, 1DS, 2AL, 2BL, 2DL</i>	1
BG314086	8	1	<i>1BS, 1DS, 2AS, 2DS, 2DL, 4BS, 4AL, 7BL</i>	<b>1DS</b>	>2
BE424439	3	4	<i>1AS, 1BS, 1DS</i>	<b>1BS, 1DS</b>	3
BE399363	2	3	<i>5AL, 5DL</i>	<i>2AL, 2BL, 2DL</i>	0
BE636802	5	3	<b>2AS, 2BS, 2DS</b>	<b>2AS, 2BS, 2DS</b>	3
BF146045	4	5	<i>1DL, 2AL, 2BL, 2DL</i>	<b>2AL, 2BL, 2DL</b>	>1
BE442746	4	3	<b>2AS, 2BS, 2DS, 3DS</b>	<b>2AS, 2BS, 2DS</b>	>1
BE495462	11	3	<b>2BL, 2DL, 3BS, 3DS, 4AL, 5AL, 5DL, 6BS, 6DS</b>	<i>2AL, 2BL, 2DL</i>	>1
BF484100	4	3	<i>1BL, 2AS, 4DL, 7AL</i>	<b>2AS, 2BS, 2DS</b>	N/A
BE585734	3	3	<b>3AL, 3BL, 3DL</b>	<b>3AL, 3BL, 3DL</b>	1
BE442801	6	3	<i>1AS, 3BS, 3DS, 5AL, 6DL</i>	<b>3BS, 3DS</b>	2
BF484536	3	11	<b>3AL, 3BL, 3DL</b>	<b>3AL, 3BL, 3DL</b>	>1
BG607143	4	5	<i>1BS, 3AL, 3BL</i>	<b>3AL, 3BL, 3DL</b>	1
BG263754	1	3	<b>3AS</b>	<b>3AS, 3BS, 3DS</b>	2
BE606965	6	3	<i>1AL, 1BL, 1DL</i>	<b>3AS, 3BS, 3DS</b>	>2
BE495116	11	2	<i>3DL, 4AS, 4BS, 4AL, 4BL, 4DL, 6BL, 6DL</i>	<b>4BL, 4DL</b>	1
BE497184	10	3	<b>4AL, 4BS, 4DS</b>	<b>4AL, 4BS, 4DS</b>	N/A
BE495848	2	3	<b>4AL, 4BS</b>	<b>4AL, 4BS, 4DS</b>	1
BE446061	3	3	<i>2AL, 5DS, 7DL</i>	<i>4AS, 4BL, 4DL</i>	>2
BF485412	4	3	<b>4AS, 4BL, 4DL, 7BL</b>	<b>4AS, 4BL, 4DL</b>	2
BF478437	20	5	<i>1BL, 1DL, 2AL, 2BL, 3BL, 3DL, 4AS, 4AL, 4BS, 4DS, 5AS</i>	<b>4AL, 4BS, 4DS</b>	>5
BE403549	3	4	<i>5AS, 5BS, 5DS</i>	<i>5AL, 5BS, 5DS, 6BS</i>	>2
BG263501	3	2	<i>4BL, 5DL, 6BL</i>	<i>5AL, 5DL</i>	>1
BE444720	6	3	<b>5AS, 5BS, 5DS</b>	<b>5AS, 5BS, 5DS</b>	>1
BE443021	2	3	<b>5BL, 5DL</b>	<i>5AL, 5BL, 5DL</i>	1
BE404135	3	6	<i>5AS, 5BS, 5DS</i>	<b>5AS, 5BS, 5DS</b>	2
BE517676	1	4	<b>5BL</b>	<i>5AL, 5BL, 5DL</i>	3
BE500430	4	5	<i>1DS, 6AS, 6DS, 7AS</i>	<b>6AS, 6BS, 6AL, 6BL, 6DL</b>	2
BE426787	4	3	<i>1BS, 3AL, 4AL, 6DS</i>	<b>6AS, 6BS, 6DS</b>	>3
BQ294805	1	2	<b>6DL</b>	<i>6BL, 6DL</i>	>2
BE445201	6	3	<b>6AS, 6BS, 6DS</b>	<b>6AS, 6BS, 6DS</b>	N/A
BF483695	3	1	<b>6AS, 6BS, 7DL</b>	<b>6AS</b>	>1
BE405680	1	3	<b>6DL</b>	<i>6AL, 6BL, 6DL</i>	N/A
BE399084	1	2	<b>7BL</b>	<b>7BL, 7DL</b>	0
BF484868	5	2	<i>4BS, 6BL, 7AS, 7BS</i>	<b>7BS, 7DS</b>	>1
BE517937	6	2	<i>1BS, 2BS, 5AL, 6DS, 7AS, 7DL</i>	<b>7AS, 7DS</b>	>1
BG262868	28	3	<i>1AL, 2AL, 2BL, 2DL, 3BS, 3DS, 3AL, 4AS, 4AL, 5A, 5AS, 5B, 5BS, 6AS, 6BS, 6BL, 7AS, 7DS, 7BL</i>	<b>7AL, 7BL, 7DL</b>	>1
BE425936	3	3	<i>2AL, 7BL</i>	<i>7AL, 7BL, 7DL</i>	N/A
BE590752	2	3	<b>7AS, 7BS</b>	<b>7AS, 7BS, 7DS</b>	N/A
Sum	266	165			>32

<sup>a</sup>*Bold fonts* denote the coincident locations of eSSR and eRFLP loci of a SSR-containing EST. *Italic fonts* denote chromosome arm with duplicated loci

<sup>b</sup>N/A Not available due to the incomplete mapping images on the web





**Fig. 5** eSSR fingerprint of 12 Triticeae genotypes generated with *Xcwem20*. Lanes A, B, C, D, E, F, G, H, I, J, K and L contain amplified bands from Chinese Spring, PI294994, PI262660, CI2401, CO970547, Above, Glupro, Langdon, Hermon H52, Mona, Wadi Qilt 23–39 and Turkey 67, respectively. The details of these genotypes are shown in Table 1. Lane M is 50-bp DNA ladder (Invitrogen cat. no. 10416–014)

ranges from 0.29 to 0.76 and has a mean value of  $0.45 \pm 0.16$  which is lower than  $0.54 \pm 0.19$  for genomic SSR markers (Röder et al. 1995) and higher than  $0.30 \pm 0.27$  for RFLP

markers (Anderson et al. 1993). The average number of alleles detected by genomic SSR markers was 4.6 (Röder et al. 1995). In the present study, the average number of alleles detected by eSSR markers reached 20.1 (Table 4) which is significantly larger than that of genomic SSR markers.

In other studies, less stringent PCR conditions had to be used to amplify bands from related species using wheat genomic SSRs (Röder et al. 1995). We were able to amplify products for all the 15 wheat eSSR markers from tetraploid wheat, barley and rye DNA samples using the same PCR conditions used for hexaploid wheat DNA (Fig. 5, Table 4).

## Discussion

### Microsatellites in genes

Microsatellites (SSRs) are usually considered as evolutionarily neutral DNA markers (Tachida and Iizuka 1992; Awadalla and Ritland 1997; Schlötterer and Wiehe 1999) and selectively neutral sequences randomly or nearly randomly distributed over the euchromatic genome (Schlötterer and Wiehe 1999; Schlötterer 2000). However, reports have repeatedly shown that SSRs also exist within genes or functional sequences (reviewed in Li et al. 2004). Expansion of a trinucleotide repeat in coding regions of a human novel gene is sometimes associated with neuropathological diseases, e.g., Huntington's disease (Huntington's Disease Collaborative Research Group 1993). The *waxy* gene in rice is found to contain a (CT)<sub>n</sub> SSR in the 5'-untranslated region (UTR) whose length polymorphism is associated with amylose content (Ayres et al. 1997). Bao et al. (2002) also found that (CT)<sub>n</sub> SSRs in starch-synthesizing genes are related to starch physicochemical properties in waxy rice. Recently, SSRs were identified in expressed genes or ESTs of several crop plants including wheat (Eujayl et al. 2002; Hackauf and Wehling 2002; Kantety et al. 2002; Morgante et al. 2002; Schloss et al. 2002; Varshney et al. 2002; Gao et al. 2003; Thiel et al. 2003; Nicot et al. 2004). In the present study, we identified 43,598 SSRs within nearly half a million wheat ESTs or the expressed gene segments. This result further corroborates that SSRs do exist in the expressed portions of wheat genomes or genes, and must have been involved in the expression and/or function of genes. We gave the designation eSSRs for SSRs found in ESTs.

### Frequency of eSSRs in wheat

We obtained a frequency of 8.85% eSSRs or the occurrence of one SSR in every 5.46-kb sequence for the 238.05-Mb wheat EST database. Through analysis of small size genomic and EST sequences in wheat, *Arabidopsis*, rice, soybean and maize, Morgante et al. (2002) found that the frequency of SSR was significantly higher in ESTs than in genomic DNA across all species. Therefore the frequency

**Table 4** Polymorphism information content (PIC) and interspecies transferability of wheat eSSR markers

Marker	For seven hexaploid wheat lines <sup>a</sup>		PCR amplification in other species <sup>b</sup>			Polymorphism <sup>c</sup>	
	PIC	Number of alleles	Tetraploid wheat	Barley	Rye	Tetraploid wheat	Barley
Xcwem6	0.76	25	+	+	+	+	+
Xcwem8	0.38	9	+	+	+	+	–
Xcwem9	0.30	25	+	+	+	+	+
Xcwem10	0.30	37	+	+	+	+	+
Xcwem12	0.64	12	+	+	+	+	+
Xcwem14	0.41	16	+	+	+	+	+
Xcwem20	0.34	30	+	+	–	–	+
Xcwem34	0.29	16	+	+	+	+	+
Xcwem38	0.52	11	+	+	+	+	–
Xcwem39	0.29	18	+	+	+	+	+
Xcwem40	0.70	16	+	+	+	+	+
Xcwem41	0.59	30	+	+	+	+	+
Xcwem51	0.43	5	+	+	+	+	+
Xcwem55	0.35	42	+	+	+	+	+
Xcwem57	0.38	9	+	+	+	+	+
Average	0.45±0.16	20.1					

<sup>a</sup>For details of the seven lines, see Table 1

<sup>b</sup>+ indicates presence of amplified bands

<sup>c</sup>Polymorphism between Langdon and Hermon H52 in tetraploid wheat, and between Mona and Wadi Qilt 23–39 in barley (+ and – indicate polymorphic and monomorphic, respectively)

of genomic SSRs in wheat may be expected to be lower than 1/5.46 kb if a similar searching stringency is applied.

Kantety et al. (2002) and Gao et al. (2003) reported a lower frequency of 3.2% and one SSR every 17.42 kb ESTs in wheat, respectively. The lower frequency may be partially due to a more stringent criterion for defining microsatellites. For example, Kantety et al. (2002) used a criterion of a minimum length of 18 (di- and tri-) bases. Gao et al. (2003) used a criterion of a minimum length of 18 (tri-) and 20 (di-) bases. Varshney et al. (2002) adopted a similar stringency as in our study and obtained an average SSR frequency of 1/6.0 kb for five cereal crops and 1/6.2 kb for wheat, which is close to our observed frequency (1/5.46 kb).

The frequency variations of eSSRs among different studies are partly due to the sampling error, exclusion of some EST libraries, and stringency of limits set for SSR extraction. A relatively low stringency applied in the *in silico* analyses may partially explain the relatively high SSR frequencies in the present study and Varshney et al. (2002). It is necessary to note that the database containing nearly half a million ESTs or 238.05 Mb EST sequence length used in the present study is the largest compared to the previous reports of 38,238 ESTs used by Kantety et al. (2002), 37.5 Mb explored by Varshney et al. (2002), 21.4 Mb searched by Gao et al. (2003), and 170,746 ESTs examined by Nicot et al. (2004). The frequency of eSSRs identified via the whole database of wheat ESTs in the present study may be close to the actual frequency of eSSRs in the wheat genome.

#### Importance of tri-nucleotide eSSRs in wheat

The majority of genomic SSRs (48–67%) found in many species are di-nucleotides (Wang et al. 1994; Schug et al.

1998). In contrast to the tri-nucleotide SSRs, di- and tetra-nucleotide SSRs are much less frequent in coding regions than in non-coding regions (reviewed in Li et al. 2002). The tri-nucleotide SSRs are reported as the most abundant class of SSRs in the EST sequences of all the crop species including wheat (Kantety et al. 2002; Morgante et al. 2002; Varshney et al. 2002; Gao et al. 2003), which was verified in this study. Suppression of non-trinucleotide SSRs in coding regions could reduce frameshift mutations, and this may be a reason for the predominance of trinucleotide SSR (Metzgar et al. 2000). Morgante et al. (2002) suggested that mutation pressure and positive selection for specific single amino acid stretches are possible reasons for the doubled frequency of tri-nucleotide repeats relative to mono- and dinucleotide repeats in the genes of plant species.

In our study, the eSSRs with CCG/CGG motif are the most frequently observed among all types of motifs (Fig. 2) which agrees with the findings of Nicot et al. (2004). Among the trimeric repeats, the motif CCG is the most common in several cereals including wheat with a frequency of 32% (Varshney et al. 2002). Gao et al. (2003) showed that CCG/CGG is the most frequent motif followed by AAC/TTG. The CCG motif was also one of the most frequent motif sequences in Kantety et al. (2002). The abundance of CCG repeats is a specific feature of monocot genomes, which may be due to the increased G + C content (Morgante et al. 2002). In contrast, AAT/ATT motifs had the lowest frequency (<1%) in the present study (Fig. 2). This may be due to the fact that TAA-based variants code for stop codons that have a direct effect on protein synthesis (Chin et al. 1996).

## Genomic distribution patterns of eSSRs

Genetically mapped molecular markers including RFLPs (Liu and Tsunewaki 1991), genomic SSRs (Röder et al. 1998), and AFLPs (Peng et al. 2000c) were shown to be non-randomly distributed among genomes of wheat with a larger number in the B genome, and this non-randomness was attributed to higher polymorphism in the B genome. Qi et al. (2004) found that 11% more EST loci were mapped in the B genome than in the D and A genomes. A larger number of  $_{SSR}ESTs$  was also observed in the B genome (Table 2).

Wheat ESTs were not always evenly distributed among genome, homoeologous groups, chromosomes, and between arms of each chromosome (Hossain et al. 2004; Lazo et al. 2004; Linkiewicz et al. 2004; Miftahudin et al. 2004; Munkvold et al. 2004; Peng et al. 2004; Qi et al. 2004; Randhawa et al. 2004; Conley et al. 2004; <http://wheat.pw.usda.gov/NSF/project/>). Deviations of  $_{SSR}EST$  distribution patterns from that of the original mapped ESTs for genomes, homoeologous groups, chromosomes, and most of the chromosome arm pairs (long vs short arm) were not significant (Table 2). Therefore, eSSRs as represented by  $_{SSR}ESTs$  basically follow distribution patterns of the original mapped ESTs in the wheat genome (Hossain et al. 2004; Lazo et al. 2004; Linkiewicz et al. 2004; Miftahudin et al. 2004; Munkvold et al. 2004; Peng et al. 2004; Qi et al. 2004; Randhawa et al. 2004; Conley et al. 2004; <http://wheat.pw.usda.gov/NSF/project/>).

## Potential of eSSR markers in cereal breeding and comparative genomics

EST-derived SSR or eSSR markers constitute a novel source of markers that are physically associated with coding regions of the genome and thus enhance the role of genetic markers in germplasm evaluation by enabling the variation in expressed genes to be assayed. However, to date a limited number of eSSR markers has been mapped in wheat. This includes 11 in Holton et al. (2002), and 112 in Nicot et al. (2004) of which 5 are publicly available. Here we present 341 wheat eSSRs present in 1,010 loci physically mapped on specific chromosomes or bins (Figs. 3, 4). Forty-eight of these were converted into eSSR markers which amplified 266 loci physically mapped to chromosome arms (Table 3; Electronic Supplementary Material). The observed high coincidence of chromosome location between  $_{SSR}ESTs$  and the derived eSSR markers (Table 3) suggests that the mapping information presented for  $_{SSR}ESTs$  (Figs. 3, 4) can be used to derive additional eSSR markers for various applications.

In previous studies, an average of <2 loci per SSR was genetically mapped for both genomic SSR and eSSR markers (Röder et al. 1998; Nicot et al. 2004). In comparison, we physically mapped an average of 5.5 loci per eSSR marker (Electronic Supplementary Material). This higher number of loci per eSSR may be explained by the difference in mapping approaches used, i.e., genetic map-

ping in the previous studies (Röder et al. 1998; Nicot et al. 2004) and physical deletion mapping in the current study (Table 3; Electronic Supplementary Material). The high frequency of inter-chromosome duplication observed for wheat ESTs (Peng et al. 2004) also may be one of the reasons for the high number of mapped eSSR loci per marker.

In general, eSSR markers detected more loci than eRFLP did. Although a large portion (25/48) of the eSSR markers amplified more loci than the corresponding eRFLP markers detected, a significant part (17/48) of the eSSR markers amplified less loci than the corresponding eRFLP markers detected (Table 3). Furthermore, the unmapped eRFLP bands/loci due to the technical weakness of the RFLP method also seemed to play an important role (>31.7%) in the number differences between eSSR and eRFLP loci as shown in Table 3. Therefore, a combination of eSSR and eRFLP approaches is necessary in genomic analyses of wheat  $_{SSR}ESTs$ .

We observed an interesting phenomenon. For seven eSSR markers, *Xcewm1*, 5, 14, 35, 39, 42 and 49, a single band was observed to disappear in two nullisomic-tetrasomic and the corresponding ditelosomic lines. For example, the *Xcewm1c* band (134 bp) disappeared in both nulli-1A/Dt1AS and nulli-1D/Dt1DS. Hence, the same band was physically mapped to two different chromosome arms, 1AL and 1DL (Electronic Supplementary Material). *Xcewm35* and *Xcewm39* detected single bands residing in 4AL and 4DS (Electronic Supplementary Material) because of the arm homoeologies, 4AS=4BL=4DL and 4AL=4BS=4DS as shown in Fig. 4 and previous studies (Endo and Gill 1984; Gill et al. 1991; Mickelson-Young et al. 1995). The possibility that more than one chromosome or chromosome arm is missing in the nulli-tetrasomic and ditelosomic lines used in this study can be ruled out, since these lines were cytogenetically characterized before use in this study. Further studies are underway to elucidate the nature of this phenomenon.

The PIC value of markers indicates the usefulness of DNA markers for gene mapping, molecular breeding and germplasm evaluation. We obtained an average PIC of  $0.45 \pm 0.16$  for 15 eSSR markers on seven hexaploid wheat lines. This value is lower than that reported for genomic SSR markers ( $0.54 \pm 0.19$ , Röder et al. 1995), but higher than RFLP markers ( $0.30 \pm 0.26$ , Anderson et al. 1993). This agrees with the observations of Eujayl et al. (2002) and Nicot et al. (2004). The eSSRs are located in genes or coding regions of genome. Thus the lower PIC of eSSRs relative to genomic SSRs is perhaps due to the high conservation of genes/coding regions among cultivars within a species. Artificial selection pressure in the breeding process also has played a role in reduction of polymorphism or PIC values among cultivars. Nevertheless, compared to genomic SSRs, eSSR markers possess additional advantages in that (1) they are cheaper to develop using huge publicly available EST databases, (2) they exist within genes and are therefore perfect genetic markers for target genes; and (3) they appear to be more readily transferable among species (Holton et al. 2002; Table 4).



Genomic SSRs are frequently not transferable to closely related species (Röder et al. 1995; Sourdille et al. 2001), and thus not suitable for comparative genomics study. Holton et al. (2002) reported a relatively high level of transferability (26–55%) for eSSRs between wheat and barley. Tall fescue eSSR markers were also transferable across several grass species including rice, ryegrass and wheat (Saha et al. 2004). All the 15 eSSR markers we tested amplified PCR products in tetraploid wheat (*T. dicoccoides*, *T. durum*), barley (*H. spontaneum*, *H. vulgare*), and rye (*S. cereale*). The majority of the 15 eSSR markers could also detect polymorphism between the two tetraploid wheat accessions and between the two barley accessions (Table 4). The transferability of eSSR markers across related Triticeae species makes them a valuable source for comparative genomics studies.

**Acknowledgements** The EST data is based upon work supported by a grant from the National Science Foundation, USA under Cooperative Agreement No. DBI-9975989. This study was also partially supported by the US Department of Agriculture under Cooperative Agreements USDA Contract No. 2001-52100-11293, USDA Contract No. 2003-34205-13636, and Hatch Funds. We sincerely thank Dr. M. Tahir for his help in SSR analysis, H. Zadeh-Gardel, H. Wang and K. J. Morey for their technical support, and all the group members of US wheat EST project (<http://wheat.pw.usda.gov/NSF/project/participants.html>) for their contributions to the mapping data.

## References

- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH et al (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252:1651–1656
- Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome* 36:181–186
- Awadalla P, Ritland K (1997) Microsatellite variation and evolution in the *Mimulus guttatus* species complex with contrasting mating systems. *Mol Biol Evol* 14:1023–1034
- Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA et al (1997) Microsatellites and single-nucleotide polymorphism differentiate apparent amylase classes in an extended pedigree of US rice germplasm. *Theor Appl Genet* 94:773–781
- Bao JS, Corke H, Sun M (2002) Microsatellites in starch-synthesizing genes in relation to starch physiochemical properties in waxy rice (*Oryza sativa* L.). *Theor Appl Genet* 105:898–905
- Brown SW, AK Szewc-McFadden, S Kresovich (1996) Development and application of simple sequence repeat (SSR) loci for plant genome analysis. In: Jauhar PP (ed) *Methods of genome analysis in plants*. CRC, Boca Raton, Fla., pp 147–159
- Chin ECL, Senior ML, Shu H, Smith JSC (1996) Maize simple repetitive DNA sequences: abundance and allele variation. *Genome* 39:866–873
- Conley E, Nduati V, Gonzalez-Hernandez JL, Mesfin A, Trudeau-Spanjers M et al (2004) A 2,600-locus chromosome bin map of wheat group 2 reveals interstitial gene-rich islands and colinearity with rice. *Genetics* 168:625–637
- Endo TR, Gill BS (1984) Somatic karyotype, heterochromatin distribution, and nature of chromosome differentiation in common wheat, *Triticum aestivum* L. em Thell. *Chromosoma* 89:361–369
- Endo TR, Gill BS (1996) The deletion stocks of common wheat. *J Hered* 87:295–307
- Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theor Appl Genet* 104:399–407
- Gao L, Tang J, Li H, Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol Breed* 12:245–261
- Gill BS, Friebe B, Endo TR (1991) Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*). *Genome* 34:830–839
- Hackauf B, Wehling P (2002) Identification of microsatellite polymorphisms in an expressed portion of the rye genome. *Plant Breed* 121:17–25
- Holton TA, Christopher JT, McClure L, Harker N, Henry RJ (2002) Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. *Mol Breed* 9:63–71
- Hossain KG, Kalavacharla V, Lazo GR, Hegstad J, Wentz MJ et al (2004) A chromosome bin map of 2,148 EST loci of wheat homoeologous group 7. *Genetics* 168:687–699
- Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72:971–983
- Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* 48:501–510
- Lazo GR, Chao S, Hummel DH, Edwards H, Crossman CC et al (2004) Development of an expressed sequence tag (EST) resource for wheat (*Triticum aestivum* L.): EST generation, unigene analysis, probe selection, and bioinformatics for a 16,000-locus bin-delineated map. *Genetics* 168:585–593
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* 11:2453–2465
- Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function and evolution. *Mol Biol Evol* 21:991–1007
- Linkiewicz AM, Qi LL, Gill BS, Echalié B, Chao S et al (2004) A 2,500-locus bin map of wheat homoeologous group 5 provides new insights on gene distribution and colinearity with rice. *Genetics* 168:665–676
- Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44:397–401
- Liu YG, Tsunewaki K (1991) Restriction fragment length polymorphism analysis of wheat. II. Linkage maps of the RFLP sites in common wheat. *Jpn J Genet* 66:617–633
- Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res* 10:72–80
- Mickelson-Young L, Endo TR, Gill BS (1995) A cytogenetic ladder-map of the wheat homoeologous group-4 chromosomes. *Theor Appl Genet* 90:1007–1011
- Miftahudin, Ross K, Ma XF, Mahmoud A, Layton J et al (2004) Analysis of EST loci on wheat chromosome group 4. *Genetics* 168:651–663
- Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet* 30:194–200
- Munkvold JD, Greene RA, Bermudez-Kandianis CE, La Rota CM, Edwards H et al (2004) Group 3 chromosome bin maps of wheat and their relationship to rice chromosome 1. *Genetics* 168:639–650
- Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F et al (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). *Theor Appl Genet* 109:800–805



- Peng JH (2000) Genomics of wild emmer wheat, *Triticum dicoccoides*: genetic maps, mapping of stripe rust resistance genes and QTLs for agronomic traits. PhD thesis, University of Haifa, Israel
- Peng JH, Fahima T, Röder MS, Li YC, Dahan A et al (1999) Microsatellite tagging of stripe-rust resistance gene *YrH52* derived from wild emmer wheat, *Triticum dicoccoides*, and suggestive negative crossover interference on chromosome 1B. *Theor Appl Genet* 98:862–872
- Peng JH, Fahima T, Röder MS, Huang QY, Dahan A et al (2000a) High-density molecular map of chromosome region harboring stripe-rust resistance gene *YrH52* and *Yr15* derived from wild emmer wheat, *Triticum dicoccoides*. *Genetica* 109:199–210
- Peng JH, Fahima T, Röder MS, Li YC, Grama A et al (2000b) Microsatellite high-density mapping of stripe-rust resistance gene *YrH52* region on chromosome 1B and evaluation of its marker-assisted selection in F<sub>2</sub> generation in wild emmer wheat. *New Phytol* 146:141–154
- Peng JH, Korol AB, Fahima T, Röder MS, Ronin YI et al (2000c) Molecular genetic maps in wild emmer wheat, *Triticum dicoccoides*: genome-wide coverage, massive negative interference, and putative quasi-linkage. *Genome Res* 10:509–1531
- Peng JH, Ronin YI, Fahima T, Röder MS, Li YC et al (2003) Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. *Proc Natl Acad Sci USA* 100:2489–2494
- Peng JH, Zadeh H, Lazo GR, Gustafson JP, Chao S 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 LL, B Echalié, Chao S, Lazo GR, Butler GE et al (2004) A chromosome bin map of 16,000 EST loci and distribution of genes among the three genomes of polyploid wheat. *Genetics* 168:701–712
- Randhawa HS, Dilbirligi M, Sidhu D, Erayman M, Sandhu D et al (2004) Deletion mapping of homoeologous group 6-specific wheat ESTs. *Genetics* 168:677–686
- Röder MS, Plaschke J, König SU, Börner A, Sorrells ME et al (1995) Abundance, variability and chromosomal location of microsatellites in wheat. *Mol Gen Genet* 246:327–333
- Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH et al (1998) A microsatellite map of wheat. *Genetics* 149:2007–2023
- Saha MC, Rouf Mian MA, Eujayl I, Zwonitzer JC, Wang L et al (2004) Tall fescue EST-SSR markers with transferability across several grass species. *Theor Appl Genet* 109:783–791
- Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109:365–371
- Schlötterer C, Wiehe T (1999) Microsatellite, a neutral marker to infer selective sweeps. In: Goldstein DB, Schlötterer C (eds) *Microsatellites: evolution and applications*. Oxford University Press, Oxford, pp 238–247
- Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers JE et al (2002) Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench. *Theor Appl Genet* 105:912–920
- Schug MD, Wetterstrand KA, Gaudette MS, Lim RH, Hutter CM et al (1998) The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. *Mol Ecol* 7:57–70
- Sears ER (1969) Wheat cytogenetics. *Annu Rev Genet* 3:451–468
- Somers DJ, Isaac P, Edwards K (2004) A high density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 109:1105–1114
- Sourdille P, Tavaud M, Charmet G, Bernard M (2001) Transferability of wheat microsatellites to diploid Triticeae species carrying the A, B, and D genome. *Theor Appl Genet* 103:346–352
- Sourdille P, Singh S, Cadalen T, Brown-Guedira GL, Gay G et al (2004) Microsatellite-based deletion bin system for the establishment of genetic–physical map relationships in wheat (*Triticum aestivum* L.). *Funct Integr Genomics* 4:12–25
- StatSoft (1996) STATISTICA for Windows [computer program manual]. Tulsa, OK 74104
- Tachida H, Iizuka M (1992) Persistence of repeated sequences that evolve by replication slippage. *Genetics* 131:471–478
- Tautz D, Trick M, Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322:652–656
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422
- Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett* 7:537–546
- Wang Z, Weber JL, Zhong G, Tanksley (1994) Survey of plant short tandem DNA repeats. *Theor Appl Genet* 88:1–6