Theor Appl Genet (2004) 108:1392–1400
DOI 10.1007/s00122-003-1554-z

ORIGINAL PAPER

L. F. Gao · R. L. Jing · N. X. Huo · Y. Li · X. P. Li · R. H. Zhou · X. P. Chang · J. F. Tang · Z. Y. Ma · J. Z. Jia

One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat

Received: 16 October 2003 / Accepted: 21 November 2003 / Published online: 14 February 2004

Abstract Four hundred and seventy-eight microsatellite markers derived from expressed sequence tags (EST-SSRs) were screened among three mapping populations (W-7984×Opata 85, WOpop; Lumai×Hanxuan, LHpop; Wenmai×Shanhongmai, WSpop). The number of polymorphic EST-SSR primer pairs found in WOpop, LHpop and WSpop was 92, 58 and 29 respectively. A total of 101 EST-SSR loci amplified from 88 primer sets were distributed over the 20 chromosomes of the reference maps (no markers were located on chromosome 4B). These 101 mapped EST-SSR markers add to the existing 450 microsatellite loci previously mapped in bread wheat. Seventy-four of the 101 loci showed significant similarities to known genes, including 24 genes involved in metabolism, 4 in cellular structures, 9 in stress resistance, 12 in transcription, 2 in development, 2 transporters and 21 storage proteins. Besides gliadin and glutenin, most of the 53 genes with putative functions were mapped for the first time by EST-SSR markers in bread wheat. Sequence alignment of the mapped wheat EST-SSR loci allowed tentative assignment of functionality to the other members of grasses family. Colinearity combined with homology information offers an attractive approach to comparative genomics.

Communicated by R. Hagemann

L. F. Gao · R. L. Jing · N. X. Huo · Y. Li · X. P. Li · R. H. Zhou · X. P. Chang · J. F. Tang · J. Z. Jia (✉)
Key Laboratory of Crop Germplasm and Biotechnology,
Ministry of Agriculture, Institute of Crop Germplasm Resources,
Chinese Academy of Agricultural Sciences,
100081 Beijing, China
e-mail: jzjia@mail.caas.net.cn
Tel.: +86-010-62186623
Fax: +86-010-62186623

Y. Li · Z. Y. Ma
Department of Agronomy,
Heibei Agricultural University,
171001 Bao ding, China

X. P. Li
Northwest Sci-Tech University of Agriculture and Forestry,
712100 Yang ling, China

Introduction

Bread wheat (*Triticum aestivum* L em. Thell.) is one of the most important crops in the world. It is an allohexaploid ($2n=6x=42$) containing three distinct but related genomes, A, B and D, each with seven chromosomes. It has a large genome of 16×10^9 bp (Bennett and Smith 1976), of which more than 80% is repetitive DNA. Detailed RFLP genetic maps (Nelson et al. 1995a, 1995b, 1995c; Marino et al. 1996) and physical maps (Gill et al. 1993; Kota et al. 1993; Hohmann et al. 1994; Delaney et al. 1995a, 1995b; Mickelson-Young et al. 1995) for all the seven homologous groups are now available. RFLP analysis, however, is limited by its labor-intensiveness and low polymorphism in wheat. In contrast, PCR-based molecular markers such as microsatellites or simple sequence repeats (SSRs) are easy to use and exhibit a higher degree of polymorphism. As of now, a total of 450 microsatellite markers have been added to wheat genetic maps by different research groups (Röder et al. 1998; Stephenson et al. 1998; Pestsova et al. 2000; Gupta et al. 2002). However, traditional SSR markers have some disadvantages. First, genomic SSR markers were mostly derived from the intergenic regions with no gene function. Second, procedures for developing those markers are complex, which include isolating and sequencing clones containing putative SSR motifs, and subsequently designing and testing the flanking primers. Recent large-scale sequencing projects have produced a large amount of single-pass sequences of complementary DNAs (cDNAs) from different plant species (http://www.ncbi.nlm.nih.gov; http://www.graingenes.org). The number of expressed sequence tags (ESTs) deposited in GenBank for wheat, maize, rice and soybean has mounted to 416,000, 197,000, 113,000 and 308,000 sequences respectively (released 10/1/03, http://www.ncbi.nih.nlm.gov/dbEST). Studies on the distribution of microsatellites in ESTs

(EST-SSRs) have been carried out in both eukaryotic (Cardle et al. 2000; Tóth et al. 2000; Kantety et al. 2002) and prokaryotic genomes (Gur-Arie et al. 2000). The estimated frequency of EST-SSRs is higher in the coding than that in the non-coding sequences (Temnykh et al. 2001; Morgante et al. 2002), suggesting that a significant proportion of ESTs can be used as polymorphic SSR markers. Furthermore, if a relatively stringent threshold for sequence similarity is used, EST-SSR mining across plant species could lead to the development of anchor markers with putative function in related plant species. Thus, traditional marker assisted selection (MAS) can be replaced by direct gene selection for targeted traits.

As of 4 February 2003, 6,636 ESTs has been mapped in wheat by RFLP analysis using mapping populations and deletion lines (http://wheat.pw.usda.gov/NSF/progress_mapping.html). However, genetic maps built from EST-SSRs have not been reported. We present here a genetic map containing 101 EST-SSR loci based on three mapping populations of hexaploid wheat, of which 74 loci represent gene mapping according to sequence similarity.

Materials and methods

Microsatellite marker development

As reported previously (Gao et al. 2003), a total of 71,495 wheat EST sequences were obtained from the wheat NSF project homepage (http://wheat.pw.usda.gov/NSF/progress_est.html). ESTs containing SSRs of at least 18 bp long for 1–6 repeat patterns were extracted and primers were designed using the program Primer 3 (http://www.basic.northwestern.edu/biotools/Primer3. html).

Plant materials, DNA extraction and PCR amplification

Three mapping populations were used in this study. The main one was an ITMI mapping population, generated by single-seed descent (F_7) hybrids from the cross of a synthetic hexaploid wheat W-7984 with *T. aestivum* variety 'Opata 85' (WOpop); a RIL population generated from a cross of released Chinese variety 'Wenmai 6' and a land variety 'Shanhongmai'(WSpop); and a double haploid (DH) population generated from cross of Chinese released variety 'Lumai 14' and 'Hanxuan 10' (LHpop). Ninety to 100 lines were selected randomly from each population and genomic DNAs were extracted from the 3-week-old leaves using an SDS-phenol-chloroform method (Devos et al. 1992). Conditions for PCR reactions were as described by Röder et al. (1998) except that the annealing temperatures were adjusted depending on the different primer pairs.

Genetic mapping

Reference maps consisting of 519 anchor markers (mainly RFLPs) for WOpop, 320 (mainly SSRs and AFLPs) for LHpop (unpublished data) and 197 (mainly SSRs and AFLPs) for WSpop (unpublished data) were prepared using MAPMAKER/Exp v3.0b (Lander et al. 1987). New microsatellite markers were integrated into the skeleton maps according to the procedures described by Gupta et al. (2002) at a LOD score of 2.5. Centimorgan units were calculated using the Kosambi mapping function (Kosambi 1944). Mapped wheat microsatellite loci were designated as Cwm for "wheat microsatellites derived from cDNAs". To construct a consensus linkage map from the three individual maps, anchor markers (mainly SSR) were chosen as standard markers that were mapped in both WOpop and LHpop or WOpop and WSpop. The positions of these loci in LHpop and WSpop were then assigned approximately for the WOpop linkage groups.

Homology searching

ESTs containing mapped microsatellites were searched against GenBank nonredundant (nr) database using TBLASTX or BLASTN algorithms (http://www.ncbi.nlm.nih.gov/BLAST). First, sequences with expected value$<10^{-7}$ by TBLASTX or $<10^{-15}$ by BLASTN were assigned putative functions. For example, *Ivrv-A1a-SSR* and *Ivrv-A1b-SSR*, two alleles of gene *Ivrv*, were mapped to the A genome by EST-SSR markers. If the *E* value was greater than 10^{-7}–10^{-15}, the markers followed Cwm designation, with a number indicating the primer pairs designed originally.

Results

Marker evaluation: functionality and polymorphism

One thousand two hundred and twenty-eight ESTs containing microsatellites were mined from the 71,495 ESTs. A total of 597 primer pairs were designed from the 1,228 ESTs-SSRs, of which 478 (80%) amplified products successfully, based on DNAs from the parents of the mapping populations (WOpop, LHpop, and WSpop). The other 20% of these primer pairs either amplified products of larger sizes than expected (fragments above 500 bp), or in most cases, produced no products. Ninety-two, 58 and 29 primers showed polymorphism between parents of WOpop, LHpop and WSpop, respectively. Thirty-one polymorphic primer sets were shared by the parents of WOpop and LHpop/WSpop, and five were shared by the parents of LHpop and WSpop. Therefore, 29.9% (143/478) of the tested primers yielded unique polymorphic EST-SSRs.

Mapping and distribution of microsatellite loci

Of the 143 polymorphic EST-SSR markers based on the three mapping populations, 88 primer sets demonstrated reproducible amplification and were used for genetic mapping. A total of 101 microsatellite loci amplified by the 88 primer pairs were integrated into the three reference maps: 67 on WOpop, 25 on LHpop, and 9 on WSpop (Fig. 1). The genetic length of WOpop reached up to 4,641.2 cM, and contained 65 EST-SSR loci together with 519 anchor markers. The average distance between loci was 7.9 cM. Two loci, *Atp-D1-SSR* and *Cwm224*, were not included in the calculation because they were assigned to the intervals of chromosome1D and chromosome 6D of the RFLP map respectively (Fig. 1). Ten of the 88 primer sets amplified more than one marker, and the highest number of loci was produced by *Cwm231* with 4 loci mapped to the non-homologous groups.

Of the 101 loci, 24 were mapped on the D genome, whereas 40 and 37 were mapped to the A and B genomes

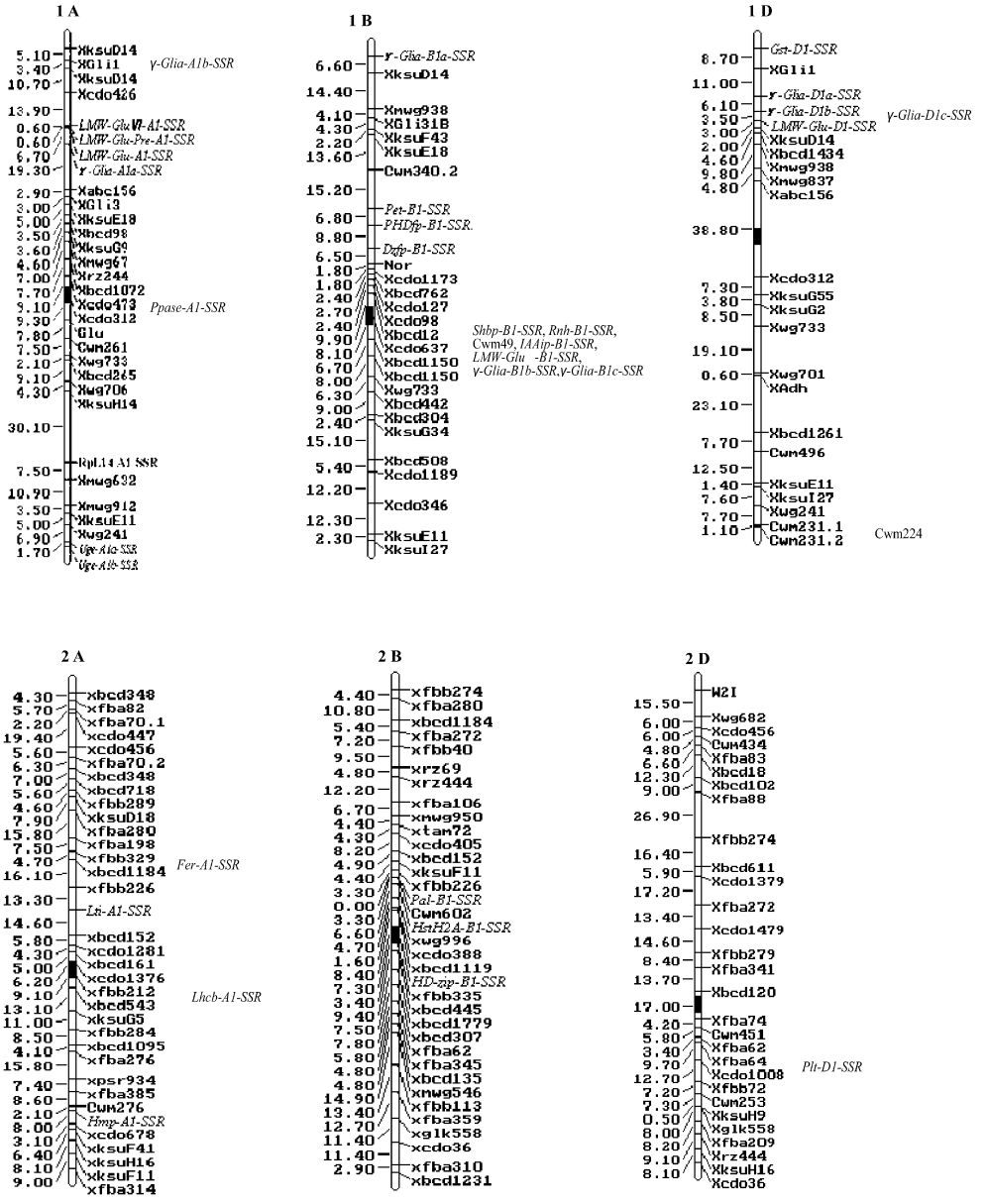


Fig. 1 Molecular linkage map of bread wheat by EST-SSR markers using three mapping populations: W7984×Opata85 (WOpop), Lumai×Hanxuan (LHpop) and Wenmai6×Shanhongmai (WSpop). For WOpop, markers with a LOD>2.5 were integrated into the RFLP framework; the other markers were placed to the most probable interval. The approximate positions of markers mapped on LHpop and WSpop were assigned to the *right* of the chromosomes on WOpop correspondingly. Estimated centromere locations were shown in *black lines*

respectively (Fig. 1). The distribution of microsatellites among seven homologous groups was not random. Thirty-one loci were mapped to the three chromosomes of homologous group1 whereas only 4 loci were mapped on chromosomes of group 4 and no locus was mapped on chromosome 4B.

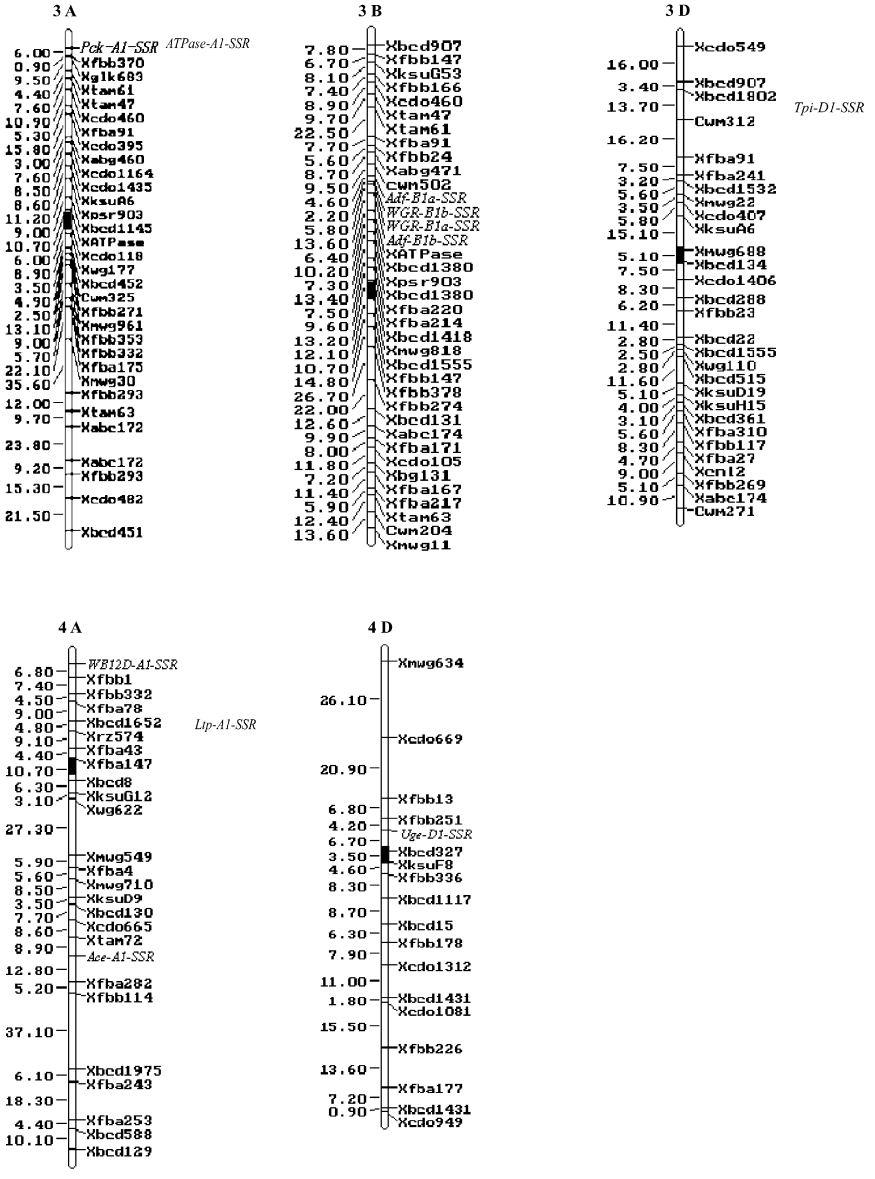
Putative functions of the mapped genes

Table 1 listed the loci for which sequence homology had been determined. The data indicated that 74 (73.2%) of 101 loci were corresponded to genes of known function. Sequence similarity searches revealed storage proteins, regulatory factors as well as structural genes and genes involved in such diverse processes as DNA synthesis, cell cycle regulation, carbon metabolism, fatty acid metabolism, membrane transport and signal transduction. Except for glutenins and gliadins that had been mapped previously as RFLP markers, all the putative genes represented by EST-SSRs were mapped for the first time in wheat (Fig. 1).

Discussion

We present here the first genetic map of the bread wheat genome based on microsatellites derived from ESTs. By comparison with the traditional method for developing microsatellites, mining microsatellites from ESTs can save considerable time and cost. In this study, 80% of the primer pairs successfully amplified products, a rate much higher than that reported by Röder et al. (1998) from wheat genomic DNA (30%) and that by Stack et al.

Fig. 1 (continued)



(2000) from wheat ESTs (50%). Most of the primer pairs produced clear and strong amplification products. Under similar PCR conditions (same thermocycler and *Taq* polymerase and buffer), there were no obvious differences between SSR markers designed from genomic sequences or ESTs. However, it appears that EST-SSRs show fewer alleles than SSRs designed from genomic DNAs. Functional constraints on ORFs may account for the lower percentage of polymorphism (19.2%) between the parents of OWmap population as compared to the polymorphism rate of 33% reported by Gupta et al. (2002) using SSR designed from genomic DNA. On the other hand, we were able to map 65 markers (73.9% of all mapped primer sets) with putative functions to 20 chromosomes. Such results will be valuable for targeted traits selection in crop breeding. For instance, EST-SSRs associated with gliadin or glutenin will be helpful for evaluating bread-making quality, whereas markers related to stress responsive genes may facilitate selection for tolerance against biotic and abiotic stresses.

This study presents a starting point for the construction of a candidate gene map of bread wheat genomes using EST-SSRs. Since 478 functional EST-SSR markers were developed from 71,000 wheat ESTs in the present study, we estimate that approximately 2,600 EST-SSR markers could be derived from the 400,000 wheat ESTs publicly available. Combined with the cSNP makers generated from ESTs, construction of a high-resolution and marker-dense transcriptional map of bread wheat is feasible in the near future.

The EST-SSR markers presented in this study have many advantages over those from anonymous genomic DNAs. Most of the EST-SSR markers (74) represented genes based on a stringent sequence similarity threshold. For example, at an *E* value of 10^{-35}, a Dof protein was identified and mapped to chromosome 1BS. The Dof

Fig. 1 (continued)

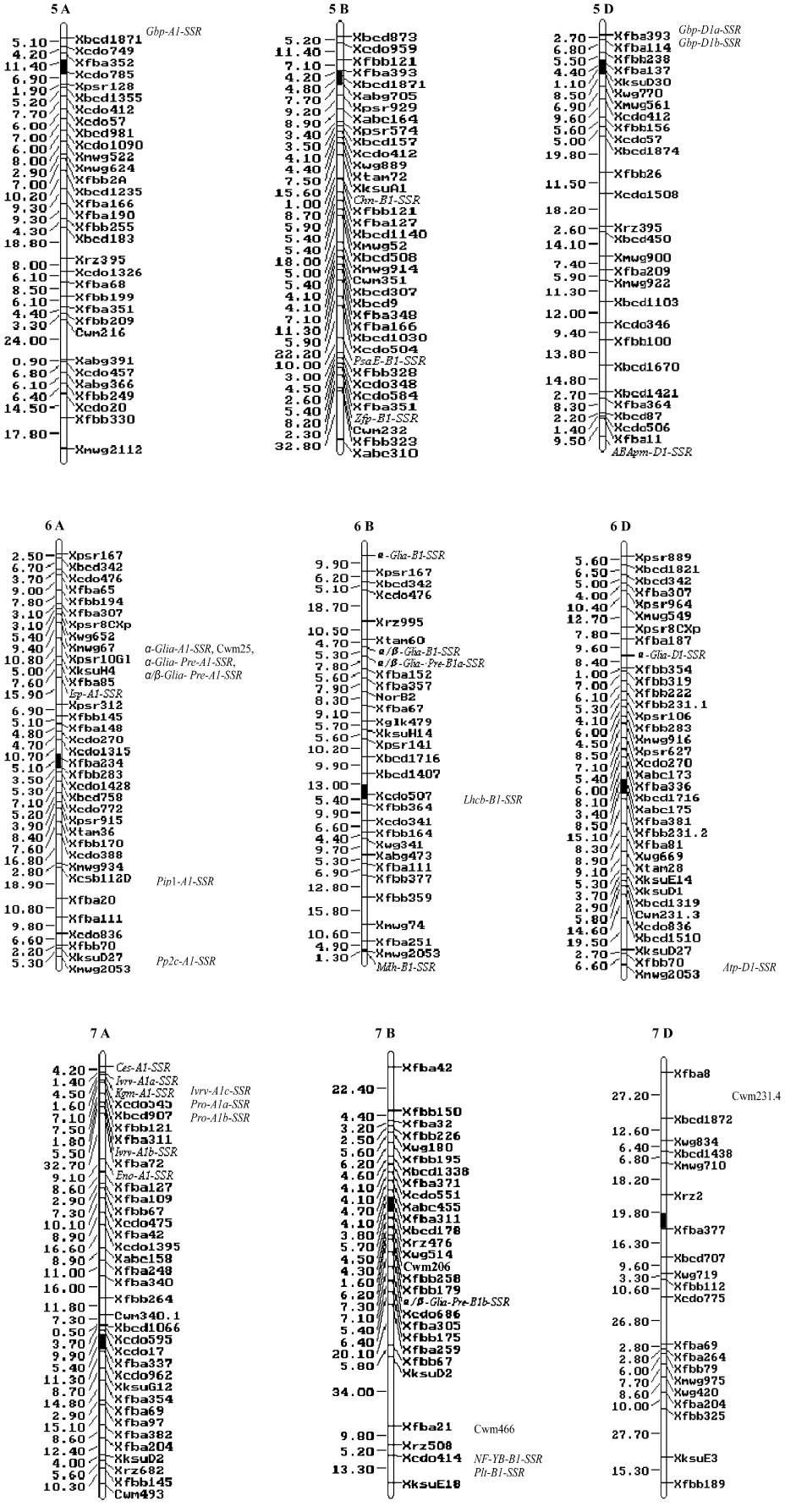


Table 1 Description of EST-derived wheat microsatellites mapped on the three populations: W-7984×Opata85 (WOpop), Lumai×Hanxuan (LHpop), and Wenmai6×Shanhongmai (WSpop). Loci were named after their GenBank homologues if the alignment on tBlastX or BlastN searching gave an *E* value $\leq 10^{-7}$ (tBlastX) or $\leq 10^{-15}$ (BlastN) respectively; otherwise, the locus was named Cwm followed by the primer pair numbers. Loci on WOpop were assigned to chromosome arms, while the loci on LHpop and WSpop were allocated to possible positions on WOpop, and then assigned to the appropriate chromosome

Locus name	Chr/arm	Repeat	*E* value	Accession no.	Function	Organism
Mdh-B1-SSR	6BL	$(a)_{46}$	3e-32	Q08062	Cytoplasmic malate dehydrogenase	*Zea mays*
Shbp-B1-SSR	1B	$(a)_{37}$	5e-24	P46285	Sedoheptulose-bisphosphatase precursor	*Triticum aestivum*
Ivrv-A1a-SSR	3AS	$(atg)_7$	8e-78[a]	AF069309	Vacuolar invertase (WIVRV)	*Triticum aestivum*
Ivrv-A1b-SSR	7AS	$(atg)_7$	8e-78[a]	AF069309	Vacular invertase (WIVRV)	*Triticum aestivum*
Ivrv-A1c-SSR	7AS	$(atg)_7$	e-09	T06226	Vacuolar invertase (WIVRV)	*Triticum aestivum*
Eno-A1-SSR	7AS	$(acg)_8$	4e-74	P26301	Enolase	*Zea mays*
Ltp-A1-SSR	4A	$(a)_{18}$	3e-47	S45370	Nonspecific lipid transfer protein Cw-18 precursor	*Hordeum vulgare*
Atp-D1-SSR	6DL	$(atggcg)_3$	7e-54	P26360	H^+-transporting ATP synthase gamma chain precursor	*Ipomoea batatas*
Ppase-A1-SSR	1A	$(aggct)_4$	8e-21	NP-196527	Inorganic pyrophosphatase-like protein	*Arabidopsis thaliana*
Pp2c-A1-SSR	6A	$(acgcgg)_4$	2e-16	AAC36699	Protein phosphatase 2C	*Mesembryanthemum crystallinum*
Plt-B1-SSR	7B	$(ac)_{10}$	e-112[a]	U18127	Phospholipid transfer protein precursor gene	*Hordeum vulgare*
Plt-D1-SSR	2D	$(ac)_{10}$	e-112[a]	U18127	Phospholipid transfer protein precursor gene	*Hordeum vulgare*
Pro-A1a-SSR	7A	$(ccg)_6$	3e-08	CAC80649	Protease P27	*Hypocrea Lixii*
Pro-A1b-SSR	7A	$(ccg)_6$	3e-08	CAC80649	Protease P27	*Hypocrea Lixii*
Isp-A1-SSR	6AS	$(aag)_6$	3e-25	T09557	Rieske iron-sulfur protein	*Arabidopsis thaliana*
Uge-A1a-SSR	1AL	$(at)_{18}$	3e-22	AY303682	UDP-glucose 4-epimerase OsUGE-1	*Oryza sativa*
Uge-A1b-SSR	1AL	$(at)_{20}$	3e-09	BAC02925	UDP-galactose 4-epimerase-like protein	*Arabidopsis thaliana*
Uge-D1-SSR	4DS	$(at)_{18}$	3e-22	AY303682	UDP-glucose 4-epimerase OsUGE-1	*Oryza sativa*
Tpi-D1-SSR	3D	$(tggcgg)_4$	5e-43[a]	Z32521	Triose-phosphate isomerase precursor	*Secale cereale*
Lhcb-B1-SSR	6B	$(a)_{50}$	2e-26	AAT23819	Chlorophyll a/b binding protein precursor	*Hordeum vulgare*
Lhcb-A1-SSR	2A	$(a)_{50}$	2e-26	AAT23819	Chlorophyll a/b binding protein precursor	*Hordeum vulgare*
PsaE-B1-SSR	5BL	$(a)_{35}$	e-157[a]	Y00966	PsaE	*Hordeum vulgare*
ATPase-A1-SSR	3A	$(at)_{23}$	4e-29	BAA90510	Ca^{2+}-ATPase	*Oryza sativa*
Pck-A1-SSR	3AS	$(aaat)_5$	4e-28[a]	AJ250829	Phosphoenolpyruvate carboxy-kinase	*Penaeus vannamei*
*Hst*H2A-B1-SSR	2BS	$(a)_{50}$	3e-50	P02277	Histone H2A	*Triticum aestivum*
*Pip*1-A1-SSR	6A	$(tgg)_7$	e-142	AAM00368	Aquaporin PIP1	*Triticum aestivum*
Ces-A1-SSR	7AS	$(agc)_6$	3e-61	T51546	Cellulose synthase catalytic subunit-like protein	*Arabidopsis thaliana*
Fer-A1-SSR	2A	$(ccg)_6$	1e-44	AAM74942	Ferritin	*Oryza sativa*
Rnh-B1-SSR	1B	$(agccgc)_3$	4e-28	NM-130144	RNA helicase	*Arabidopsis thaliana*
Kgm-A1-SSR	7AS	$(agg)_6$	2e-89[a]	AY167561	GAMYB-binding protein mRNA	*Hordeum vulgare*
PHDfp-B1-SSR	1BS	$(agctgc)_3$	6e-39	AA072583	PHD-finger protein	*Arabidopsis thaliana*
NF-YB-B1-SSR	7B	$(ggc)_6$	1e-63	BAC76332	NF-YB	*Oryza sativa*
Zbp-B1-SSR	5BL	$(tccg)_6$	2e-28	AA038447	Zinc binding protein	Oryza sativa
Dzfp-B1-SSR	1BS	$(tttc)_6$	2e-11	BAA78572	Dof zinc finger protein	*Oryza sativa*
HD-zip-B1-SSR	2BL	$(atggcg)_3$	1e-39	AA072559	Homeodomain leucine zipper protein	*Oryza sativa*
Adf-B1a-SSR	3BS	$(acccg)_4$	2e-23	AAK09235	Actin-depolymerizing factor	*Oryza sativa*
Adf-B1b-SSR	3BS	$(acccg)_4$	3e-67	AAK09235	Actin-depolymerizing factor	*Oryza sativa*
Gbp-A1-SSR	5A	$(ccggcg)_3$	1e-27	AF112964	Small GTP-binding protein	*Triticum aestivum*
Gbp-D1a-SSR	5D	$(ccggcg)_3$	1e-27	AF112964	Small GTP-binding protein	*Triticum aestivum*
Gbp-D1b-SSR	5D	$(ccggcg)_3$	1e-25	AF112964	Small GTP-binding protein	*Triticum aestivum*
Pet-B1-SSR	1BS	$(ggc)_6$	4e-41	NP-173670	Oligopeptide transporter	*Arabidopsis thaliana*
RpL14-A1-SSR	1AL	$(ccg)_8$	1e-37	NP-179635	60S ribosomal protein L14	*Arabidopsis thaliana*
WB12D-A1-SSR	4AS	$(a)_{18}$	7e-28	S60284	HvB12D	*Hordeum vulgare*
α-Glia-B1-SSR	6BS	$(aac)_{20}$	4e-60	CAA10257	Alpha-gliadin	*Triticum aestivum*
α-Glia-A1-SSR	6A	$(aac)_{13}$	2e-63	CAB76961	Alpha-gliadin	*Triticum aestivum*
α-Glia- Pre-A1-SSR	6A	$(aac)_{11}$	3e-11	T06282	Alpha-gliadin precursor	*Triticum aestivum*
α-Glia-D1-SSR	6DS	$(aac)_{14}$	3e-27	AAB48476	Alpha-gliadin	*Triticum aestivum*
α/β-Glia-B1-SSR	6BS	$(aac)_{11}$	9e-64[a]	AJ133605	Alpha/beta gliadin	*Triticum aestivum*

Table 1 (continued)

Locus name	Chr/arm	Repeat	*E* value	Accession no.	Function	Organism
α/β-Glia- Pre-A1-SSR	6A	$(aac)_{11}$	8e-28	P04721	Alpha/beta-gliadin A-I precursor	*Triticum aestivum*
α/β-Glia- Pre-B1a-SSR	6BS	$(aac)_{15}$	4e-10	P04723	Pre-alpha-/beta-gliadin A-III	*Triticum aestivum*
α/β- Glia- Pre-B1b-SSR	7BL	$(aac)_{16}$	3e-11	E22364	Alpha/beta-gliadin precursor	*Triticum aestivum*
γ-Glia-A1a-SSR	1AS	$(aac)_6$	3e-18	AAK84773	Gamma-gliadin	*Triticum aestivum*
γ-Glia-A1b-SSR	1AS	$(ttg)_{13}$	2e-29	AAK84773	Gamma-gliadin	*Triticum aestivum*
γ-Glia-B1a-SSR	1BS	$(aac)_{10}$	4e-09	CAC10616	Gamma-gliadin	*Triticum aestivum*
γ-Glia-B1b-SSR	1B	$(aac)_6$	3e-18	AAK84773	Gamma-gliadin	*Triticum aestivum*
γ-Glia-B1c-SSR	1B	$(aac)_{10}$	8e-49	AAK84779	Gamma-gliadin	*Triticum aestivum*
γ-Glia-D1a-SSR	1DS	$(aac)_{10}$	4e-09	CAC10616	Gamma-gliadin	*Triticum aestivum*
γ-Glia-D1b-SSR	1DS	$(aac)_{10}$	4e-09	CAC10616	Gamma-gliadin	*Triticum aestivum*
γ-Glia-D1c-SSR	1D	$(aac)_6$	7e-48	AAK84779	Gamma-gliadin	*Triticum aestivum*
LMW-*Glu*-A1-SSR	1AS	$(aac)_6$	8e-08	P10385	LMW glutenin	*Triticum aestivum*
LMW-*Glu*-D1-SSR	1DS	$(aac)_6$	7e-10	CAB41921	LMW glutenin	*Triticum aestivum*
LMW-*Glu*-Pre-A1-SSR	1AS	$(aac)_9$	8e-54	AA017157	LMW glutenin precursor	*Triticum aestivum*
LMW-*Glu*II-B1-SSR	1B	$(aac)_6$	7e-10	BAB78749	LMW glutenin subunit group 4 typeII	*Triticum aestivum*
LMW-*Glu*VI-A1-SSR	1AS	$(agc)_5(aac)_6$	5e-13	T06982	LMW glutenin subunit group 11 type VI	*Triticum aestivum*
Ace-A1-SSR	4AL	$(aac)_6$	7e-15	AAB32025	Alcohol-soluble avenin	*Avena sativa*
Gst-D1-SSR	1DS	$(actccc)_3$	8e-48[a]	AF387085	Glutathione-s-transferase	*Triticum aestivum*
Chn-B1-SSR	5BL	$(agctg)_4$	7e-33	NP-172076	Class I chitinase	*Arabidopsis thaliana*
Pal-B1-SSR	2BS	$(actccg)_3$	5e-07	Z49147	Phenylalanine ammonia-lyase	*Hordeum vulgare*
ABApm-D1-SSR	5DL	$(ac)_{10}$	1e-18[a]	U80037	ABA induced plasma membrane protein	*Triticum aestivum*
WGR-B1a-SSR	3BS	$(aag)_9$	2e-18	BAA74805	ZmGR2a	*Zea mays*
WGR-B1b-SSR	3BS	$(aag)_9$	6e-15	BAA74805	ZmGR2a	*Zea mays*
Lti-A1-SSR	2AS	$(ag)_{18}$	2e-15	BAC16385	Low temperature and salt responsive protein LTI6B	*Oryza sativa*
Hmp-A1-SSR	2AL	$(ccg)_8$	4e-08	NP-171656	Heavy-metal-associated domain-containing protein	*Arabidopsis thaliana*
IAAip-B1-SSR	1B	$(accagc)_3$	1e-18	AAD32147	IAA induced protein	*Nicotiana tabacum*
Cwm25	6A	$(aac)_6$	–	–	–	–
Cwm49	1B	$(agc)_6$	–	–	–	–
Cwm204	3BL	$(atgt)_6$	–	–	–	–
Cwm206	7BL	$(ag)_{34}$	–	–	–	–
Cwm216	5AL	$(agtg)_5$	2e-10	AAP21432	Unknown protein	*Oryza sativa*
Cwm224	1DL	$(cggg)_5$	–	–	–	–
Cwm231.1	1DL	$(aacgcc)_4$	–	–	–	–
Cwm231.2	1DL	$(aacgcc)_4$	–	–	–	–
Cwm231.3	6DL	$(aacgcc)_4$	–	–	–	–
Cwm231.4	7D	$(aacgcc)_4$	–	–	–	–
Cwm232	5BL	$(aaccct)_3$	–	–	–	–
Cwm253	2DL	$(acg)_7$	–	–	–	–
Cwm261	1AL	$(ttc)_7$	–	–	–	–
Cwm271	3DL	$(tgc)_7$	–	–	–	–
Cwm276	2AL	$(ccg)_8$	–	–	–	–
Cwm312	3DS	$(cgg)_6$	7e-10	AAL76183	Hypothetical protein	*Oryza sativa*
Cwm325	3AL	$(ac)_{12}$	7e-38	AAM65538	Hypothetical protein	*Oryza sativa*
Cwm340.1	7AS	$(acg)_7$	–	–	–	–
Cwm340.2	1BS	$(acg)_7$	–	–	–	–
Cwm351	5BL	$(agg)_7$	–	–	–	–
Cwm434	2DS	$(ttc)_6$	2e-60	CAE02123	Hypothetical protein	*Arabidopsis thaliana*
Cwm451	2DL	$(agc)_6$	7e-21	CAE03770	Putative protein	*Arabidopsis thaliana*
Cwm466	7B	$(ccg)_6$	–	–	–	–
Cwm493	7AL	$(acgc)_6$	–	–	–	–
Cwm496	1DL	$(agcc)_6$	–	–	–	–
Cwm502	3BS	$(atgg)_6$	–	–	–	–
Cwm602	2BS	$(acg)_5$at-g$(acg)_6$	–	–	–	–

[a] Searched against the nonredundant nucleotide database

proteins are a large family of transcription factors, recently discovered and present only in plants (Papi et al. 2002). A number of Dof proteins are being characterized in maize (Vicente-Carbajosa et al. 1997; Yanagisawa and Sheen 1998; Yanagisawa 2000), barley (Mena et al. 1998), pumpkin (Kisu et al. 1998; Shimofurutani et al. 1998), tobacco (Baumann et al. 1999) and *Arabidopsis* (Gualberti et al. 2002) and these genes appear to confer

distinct functions in different plant taxa. So far, only the Dof gene *DAG1* was convincingly demonstrated to have effects on seed germination in Arabdopsis (Papi et al. 2000). There has been no description of this gene family in wheat until now. With the EST-SSR approach, we were able to map a Dof homolog on chromosome 1B, enabling further research to determine the functions of this gene in wheat. Homeodomain-leucine zipper (HD-zip) proteins, another large family of transcription factors, are apparently unique to plants (Johannesson 2000). HD-zip protein Oshox1 showed repressor function in rice but conferred transcriptional activation in yeast (Meijer et al. 2000). We have mapped one member of this protein family to chromosome 2BL based on sequence similarity to rice HD-zip protein (*E* value 10^{-28}). Based on high sequence similarity to that of rice, an actin-depolymerizing factor was located on bread wheat chromosome 3BS. Studies confirmed that this actin-depolymerizing factor was involved in pollen actin reorganization (Lopez et al. 1996) and affected pollen tube elongation (Chen et al. 2002). These examples illustrate the power of using homology search from rice to place genes with putative function onto a wheat map that will facilitate functional assignment of wheat ESTs.

Consensus gene maps are important tools for comparative genetics. With the availability of the whole genome sequences of rice and *Arabidopsis* as well as the abundant EST data from many plant taxa, it is possible to reveal more information between monocots and dicots by sequence alignment since gene contents and gene orders among different plant species are highly conserved (Bennetzen and Freeling 1993; Gale and Devos 1998). It implied that alignment of common markers on maps of one member of the grass family allows tentative assignment of functionality to the other genes (Gale and Devos 1998). Alignment of sequence data across organisms will become an increasingly important aspect of future gene discovery and development strategy. Colinearity combined with homology information offers an attractive approach to comparative genomics.

Acknowledgements The authors gratefully acknowledge two anonymous referees for their constructive comments and Dr. Leung He (IRRI) for critically reading the manuscript.

References

Baumann K, De Paolis A, Costantino P, Gualberti G (1999) The DNA binding site of the Dof protein NtBBF1 is essential for tissue-specific and auxin-regulated expression of the *rolB* oncogene in plants. Plant Cell 11:323–334

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc Lond B 274:227–274

Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: genome composition, colinearity and compatibility. Trends Genet 9:259–261

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Chen CY, Wong EI, Vidali L, Estavillo A, Hepler PK, Wu HM, Cheung AY (2002) The regulation of actin organization by actin-depolymerizing factor in elongating pollen tubes. Plant Cell 14:2175–2190

Delaney DE, Nasuda S, Endo TR, Gill BS, Hulbert SH (1995a) Cytologically based physical maps of the group-2 chromosomes of wheat. Theor Appl Genet 91:568–573

Delaney DE, Nasuda S, Endo TR, Gill BS, Hulbert SH (1995b) Cytologically based physical maps of the group-3 chromosomes of wheat. Theor Appl Genet 91:780–782

Devos KM, Atkinson MD, Chinoy CN, Liu C, Gale MD (1992) RFLP-based genetic map of the homoeologous group-3 chromosomes of wheat and rye. Theor Appl Genet 83:931–939

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Gao LF, Tang JF, Li HW, Jia JZ (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 12:245–261

Gill KS, Gill BS, Endo TR (1993) A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. Chromosoma 102:374–381

Gualberti G, Papi M, Bellucci L, Ricci I, Bouchez D, Camilleri C, Costantino P, Vittorioso P (2002) Mutations in the Dof zinc finger genes *DAG2* and *DAG1* influence with opposite effects the germination of Arabidopsis seeds. Plant Cell 14:1253–1263

Gupta PK, Balyan HS, Edwards KJ, Isaac P, Korzun V, Röder MS, Gautier MF, Joudrier P. Schlatter AR, Dubcovsky J, De la Pena RC, Khairallah M, Penner G, Hayden MJ, Sharp P, Keller B, Wang RCC, Hardouin JP, Jack P, Leroy P (2002) Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor Appl Genet 105:413–422

Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM, Kashi Y (2000) Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. Genome Res 10:62–71

Hohmann U, Enso TR, Gill KS, Gill BS (1994) Comparison of genetic and physical maps of group 7 chromosomes from *Triticum aestivum* L. Mol Gen Genet 245:644–653

Johannesson H (2000) Functional analysis of homeodomain-leucine zipper transcription factors in *Arabidopsis thaliana*. Dissertation, University of Uppsala

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

Kisu Y, Esaka M, Suzuki M (1995) Putative zinc binding domain of plant transcription factor AOBP is related to DNA binding domains of steroid hormone receptors and GATA1. Proc Jpn Acad 71:288–292

Kosambi DD (1944) The estimation of map distances from recombination values. Annu Eugen 12:172–175

Kota RS, Gill KS, Gill BS, Endo TR (1993) A cytogenetically based physical map of chromosome 1B in common wheat. Genome 36:548–554

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lopez I, Anthony RG, Maciver SK, Jiang CZ, Khan S, Weeds AG, Hussey PJ (1996) Pollen specific expression of maize genes encoding actin depolymerizing factor-like proteins. Proc Natl Acad Sci USA 93:7415–7420

Marino CL, Nelson JC, Lu YH, Sorrells ME, Leroy P, Tuleen NA, Lopes CR, Hart GE (1996) Molecular genetic maps of the group 6 chromosomes of hexaploid wheat (*Triticum aestivum* L.em.Tell.). Genome 39:359–366

Meijer AH, de Kam RJ, d' Erfurth I, Shen W, Hoge JH (2000) HD-Zip proteins of families Iand II from rice: interactions and functional properties. Mol Gen Genet 263:12–21

Mena M, Vicente-Carbajosa J, Schmidt R, Carbonero P (1998) An endosperm-specific Dof protein from barley, highly conserved in wheat, binds to and activates transcription from the prola-

min-box of a native B-hordein promoter in barley endosperm. Plant J 16:53–62
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
Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200
Nelson JC, Sorrells ME, van Deynze AE, Lu YH, Atkinson M, Bernard M, Leroy P, Faris JD, Anderson JA (1995a) Molecular mapping of wheat: major genes and rearrangements in homologous groups 4, 5, and 7. Genetics 141:721–731
Nelson JC, van Deynze AE, Autrique E, Sorrells ME, Lu YH, Merlino M, Atkinson M, Leroy P (1995b) Molecular mapping of wheat, homologous group 2. Genome 38:516–524
Nelson JC, van Deynze AE, Autrique E, Sorrells ME, Lu YH, Nègre S, Bernard M, Leroy P (1995c) Molecular mapping of wheat, Homologous group 3. Genome 38:525–533
Papi M, Sabatini S, Bouchez D, Camilleri C, Costantino P, Vittorioso P (2000) Identification and disruption of an Arabidopsis zinc finger gene controlling seed germination. Genes Dev 14:28–33
Papi M, Sabatini S, Altamura MM, Hennig L, Schäfer E, Costantino P, Vittorioso P (2002) Inactivation of the phloem-specific Dof zinc finger gene *DAG1* affects response to light and integrity of the testa of Arabidopsis seeds. Plant Physiol 128:411–417
Pestsova E, Ganal MW, Röder MS (2000) Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 43:689–697
Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023
Shimofurutani N, Kisu Y, Suzuki M, Esaka M (1998) Functional analyses of the Dof domain, a zinc finger DNA-binding domain, in a pumpkin DNA-binding protein AOBP. FEBS Lett 430:251–256
Stack S, Campbell L, Henderson K, Eujayl I, Hanafey M, Powell W, Wolters P (2000) Development of EST-derived microsatellite markers for mapping and germplasm analysis in wheat. Plant and Animal Genome Conference, San Diego, California, 2000
Stephenson P, Bryan G, Kirby J, Collins A, Devos K, Busso C, Gale M (1998) Fifty new microsatellite loci for the wheat genetic map. Theor Appl Genet 97:946–949
Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon association, and genetic marker potential. Genome Res 11:1441–1452
Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10:967–981
Vicente-Carbajosa J, Moose S, Parsons RL, Schmidt R (1997) A maize zinc finger protein binds the prolamin box in zein gene promoters and interacts with basic leucine zipper transcriptional activator Opaque2. Proc Natl Acad Sci USA 94:7685–7690
Yanagisawa S (2000) Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. Plant J 21:281–288
Yanagisawa S, Sheen J (1998) Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. Plant Cell 10:75–99