Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat

Ju-Kyung Yu, Trevor M. Dake, Sukhwinder Singh, David Benscher, Wanlong Li, Bikram Gill, and Mark E. Sorrells

Abstract: Expressed sequence tags (ESTs) are a valuable source of molecular markers. To enhance the resolution of an existing linkage map and to identify putative functional polymorphic gene loci in hexaploid wheat (*Triticum aestivum* L.), over 260 000 ESTs from 5 different grass species were analyzed and 5418 SSR-containing sequences were identified. Using sequence similarity analysis, 156 cross-species superclusters and 138 singletons were used to develop primer pairs, which were then tested on the genomic DNA of barley (*Hordeum vulgare*), maize (*Zea mays*), rice (*Oryza sativa*), and wheat. Three-hundred sixty-eight primer pairs produced PCR amplicons from at least one species and 227 primer pairs amplified DNA from two or more species. EST-SSR sequences containing dinucleotide motifs were significantly more polymorphic (74%) than those containing trinucleotides (56%), and polymorphism was similar for markers in both coding and 5′ untranslated (UTR) regions. Out of 112 EST-SSR markers, 90 identified 149 loci that were integrated into a reference wheat genetic map. These loci were distributed on 19 of the 21 wheat chromosomes and were clustered in the distal chromosomal regions. Multiple-loci were detected by 39% of the primer pairs. Of the 90 mapped ESTs, putative functions for 22 were identified using BLASTX queries. In addition, 80 EST-SSR markers (104 loci) were located to chromosomes using nullisomic-tetrasomic lines. The enhanced map from this study provides a basis for comparative mapping using orthologous and PCR-based markers and for identification of expressed genes possibly affecting important traits in wheat.

Key words: wheat, EST, SSR mapping.

Résumé : Les étiquettes de séquences exprimées (EST) constituent une source intéressante de marqueurs moléculaires. Afin d'augmenter la résolution d'une carte génétique existante et d'identifier des locus géniques fonctionnels polymorphes chez le blé hexaploïde (*Triticum aestivum* L.), plus de 260 000 EST provenant de 5 espèces de graminées ont été analysés et 5418 séquences contenant un microsatellite ont été identifiées. Par analyse de séquences, 156 super-groupes interspécifiques et 138 séquences uniques ont été exploités pour développer des paires d'amorces, lesquelles ont ensuite été testées sur l'ADN génomique de l'orge (*Hordeum vulgare*), du maïs (*Zea mays*), du riz (*Oryza sativa*) et du blé. Trois cent soixante-huit paires d'amorces ont produit des amplicons chez au moins une des espèces et 227 paires d'amorces ont amplifié l'ADN chez deux espèces ou plus. Les microsatellites dinucléotidiques étaient significativement plus polymorphes (74 %) que les microsatellites à motif trinucléotidique (56 %). Par ailleurs, le polymorphisme était semblable au sein des régions codantes et 5'UTR. Parmi 112 microsatellites, 90 ont permis d'ajouter 149 locus sur la carte génétique de référence chez le blé. Ces locus étaient situés sur 19 des 21 chromosomes du blé et étaient groupés au sein des régions distales des chromosomes. Des locus multiples ont été détectés pour 39 % des paires d'amorces. Des 90 EST qui ont été cartographiés, une fonction putative a été déterminée au terme d'une interrogation BLASTX. De plus, 80 marqueurs EST-SSR (104 locus) ont été assignés à un chromosome à l'aide de lignées nullisomiques-tétrasomiques. La carte bonifiée issue de ce travail permet de faire de la cartographie comparée à l'aide de marqueurs PCR orthologues et d'identifier des gènes exprimés qui pourraient influencer des caractères importants chez le blé.

Mots clés : blé, EST, cartographie de microsatellites.

[Traduit par la Rédaction]

Received 27 October 2003. Accepted 9 December 2003. Published on the NRC Research Press Web site at http://genome.nrc.ca on 8 October 2004.

Corresponding Editor: P. Donini.

J.-K. Yu, T.M. Dake, D. Benscher, and M.E. Sorrells.[1] Department of Plant Breeding, Cornell University, 413 Bradfield Hall, Ithaca, NY 14850, USA.
S. Singh, W. Li, and B. Gill. Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA.

[1]Corresponding author (e-mail: mes12@cornell.edu).

 doi: 10.1139/G04-057

Introduction

Bread wheat (*Triticum aestivum* L. em. Thell, $2n = 6x = 42$) is one of the most important crops in the world, and its genetics and genome organization have been extensively studied by molecular markers. These studies are of value for locating genes of interest, facilitating marker-assisted breeding, map-based cloning, and for understanding the biological basis of complex traits. Despite a large genome size (16×10^9 bp), polyploidy, and a low level of polymorphism, there are more than 1800 mapped loci for wheat owing to concerted efforts to construct genetic linkage maps (Gupta et al. 1999; http://wheat.pw.usda.gov/ggpages/maps.shtml). A comprehensive hexaploid wheat map has been developed for the 'W7984' × 'Opata 85' cross, consisting of more than 1250 marker loci with over 3000 cM of genome coverage (Nelson et al. 1995*a*, 1995*b*, 1995*c*; Marino at el. 1996; Röder et al. 1998; http://wheat.pw.usda.gov/ggpages/maps.shtml). In addition, researchers have constructed maps based on cytogenetic stocks missing complete chromosomes, chromosome arms (Anderson et al. 1992), or chromosome arm segments (Endo and Gill 1996). Recently, more than 6700 expressed sequence tags (EST) have been mapped on to wheat deletion lines (http://wheat.pw.usda.gov/cgi-bin/westsql/westsql.cgi).

Simple sequence repeat (SSR) markers are microsatellite loci that can be amplified by polymerase chain reaction (PCR) using primers designed for unique flanking sequences. Polymorphism is based on variation in the number of repeats in different genotypes owing to polymerase slippage and point mutations (Kruglyak et al. 1998). SSR markers are (*i*) highly informative, (*ii*) locus-specific and frequently show co-dominant inheritance, (*iii*) adaptable to high-throughput genotyping, and (*iv*) are simple to maintain and distribute (Röder et al. 1998; Temnykh et al. 2000; Yu et al. 2002*a*). Approximately 450 SSR loci have been mapped in hexaploid wheat (Röder et al. 1998; Stephenson et al. 1998; Pestsova et al. 2000; Eujayl et al. 2002; Gupta et al. 2002; Song et al. 2002). Conventional experimental methods for developing SSRs are based on isolating and sequencing clones containing putative SSR tracts, followed by the designing and testing of flanking primers. To reduce cost and time, various cloning procedures to create libraries enriched for SSRs have been developed (Edwards et al. 1996); however, genomic SSR marker development is still labor intensive (Gupta and Varshney 2000).

cDNA sequences are also potentially a valuable source of molecular markers. Analysis of ESTs is a simple strategy to study the transcribed parts of genomes, thus rendering even complex and highly redundant genomes like wheat amenable to large-scale analysis. There are a number of advantages using expressed genes compared with anonymous sequences as genetic markers. First, if an EST marker is found to be genetically associated with a trait of interest, it may be possible that this could be the gene affecting the trait directly (Chen et al. 2001; Thiel et al. 2003). Therefore, EST-derived markers can provide opportunities for gene discovery and enhance the role of genetic markers by assaying variation in transcribed and known-function genes. Second, EST-derived markers are likely to be more highly conserved and therefore may be more transferable between species than anonymous sequence-derived markers (Cordeiro et al. 2001; Taylor et al. 2001; Decroocq et al. 2003). Third, ESTs that share homology to candidate genes can be specifically targeted for genetic mapping and can be useful for aligning genome linkage across distantly related species for comparative analysis (Holton et al. 2002). The occurrence of microsatellites in EST sequences has been reported for different species including barley (Holton et al. 2002; Thiel et al. 2003), rice (Temnykh et al. 2001), durum wheat (Eujayl et al. 2002), bread wheat (Kantety et al. 2002), rye (Hackauf and Wehling 2002), sugarcane (Cordeiro et al. 2001), and grape (Scott et al. 2000). Thus, the generation of EST-derived SSR markers has become an attractive complement to existing SSR collections.

Kantety et al. (2002) estimated the frequency of SSR from the publicly available EST databases for five different cereal crops (barley, maize, rice, sorghum, and wheat). The frequency of SSRs in the ESTs varied from 1.5% for maize to 4.7% for rice with an average of 3.2% across five species. The SSR frequency in wheat ESTs was 3.2% with a representation of di-, tri-, and tetra-nucleotide repeat motifs at 10%, 83%, and 7%, respectively. The GA–CT motif is the most abundant dinucleotide repeat motif in ESTs from all the species even though the AT–TA repeat was predicted to be the most abundant motif in plant genomes (Morgante and Olivieri 1993). In wheat ESTs, the most abundant trinucleotide repeat motif is AAC–TTG (Kantety et al. 2002) and this differed from genomic DNA (Song et al. 2002). Based on a study of wheat genomic SSRs by Song et al. (2002), the TAA–ATT motif was more abundant and informative as a marker than the AAC–TTG motif. However, the GGC–CCG motif is the most common trinucleotide repeat motif in most cereals, consistent with a high proportion of G–C content in coding regions (Moore 2000; Varshney et al. 2002). As genetic markers, EST-SSRs have been evaluated in several studies and tend to be considerably less polymorphic than those from genomic DNA for rice (Cho et al. 2000), sugarcane (Cordeiro et al. 2001), barley (Thiel et al. 2003), or durum wheat (Eujayl et al. 2002). Eujayl et al. (2002) showed that EST-SSR primers produce high quality markers, but expose a lower level of polymorphism (25%) compared with the genomic SSR markers (53%) in durum wheat. In barley, EST-SSRs from the 3′ untranslated region (UTR) were more polymorphic than those from 5′ UTR (Thiel et al. 2003), but a study of sugarcane has shown the opposite result (Cordeiro et al. 2001). Scott et al. (2000) reported that the degree of polymorphism depended on the genotypes used for the screening panel. SSRs from the 3′ UTR were most polymorphic within species, while those from the 5′ UTR were most polymorphic between species. Despite fewer polymorphisms, a small number of EST-SSRs (10 markers) generated a very high discriminatory power (Eujayl et al. 2002) and showed a high frequency of cross-species transferability in several studies (Cordeiro et al. 2001; Taylor et al. 2001; Decroocq et al. 2003).

The objectives of this study were (*i*) to analyze publicly available wheat and rice EST sequences containing SSRs, (*ii*) to develop SSR markers based on these EST sequences, (*iii*) to examine the efficiency of marker transferability within cereal crop species, and (*iv*) to integrate EST-SSR marker loci into an existing framework genetic map of wheat.

Materials and methods

Plant material

Newly developed EST-SSR primer pairs were screened against five *Triticum aestivum* accessions ('W7984', 'Opata85', 'Clark's Cream', 'NY-6432-18', 'Chinese Spring'), two *Triticum monococcum* accessions ('DV92' and 'AUS18913'), one *Triticum durum* accession ('Cham1'), four *Oryza sativa* subsp. *japonica* accessions ('Nipponbare', 'IR64', and 'Azucena'), one *Oryza rufipogon* accession, two *Zea mays* accessions ('B73' and 'Mo17'), and two *Hordeum vulgare* accessions ('Steptoe' and 'Morex'). This screening panel was also used for the investigation of the cross-species amplification of EST-SSR marker primers. The wheat mapping used 46 recombinant inbred lines (RIL) from the cross between W7984 and Opata 85, which was derived by single-seed descent (Van Deynze et al. 1995). This population, also known as the International Triticeae Mapping Initiative (ITMI) population, has been used for the construction of a hexaploid wheat map including RFLP (Nelson et al. 1995*a*, 1995*b*, 1995*c*) and genomic SSR (Röder et al. 1998) maps. Genomic DNA from the above plant material was isolated from young leaves as described by Heun et al. (1991).

Marker development

In silico data mining for SSR markers

An analysis of EST sequences for SSR identification from five different cereal crops (barley, maize, rice, sorghum, and wheat) was obtained from Kantety et al. (2002). In that study, over 260 000 EST sequences were analyzed (*i*) to identify SSRs from EST sequences of different species, (*ii*) to cluster EST sequences containing SSRs within species, and (*iii*) to construct EST-SSR clusters across five different species, supercluster, or cross-species cluster (Kantety et al. 2002). We focused on wheat and rice ESTs and these sequences were re-analyzed to determine the positions of SSRs and identify the putative functions of SSR-containing ESTs in developing markers for this study. A total of 219 superclusters containing 5418 SSR-bearing EST sequences was used to develop markers. A PERL5 script (named MISA, microsatellite; http://pgrc.ipk-gatersleben.de/misa/) was used for identification of SSRs. The minimum repeat number was five for di- and tri-nucleotides, and four for tetra- and penta-nucleotides. Superclusters that did not contain at least one EST sequence with more than 30 bp of flanking sequences to the SSR were excluded. To identify the positions of SSRs, initially BLASTX query was used to find published mRNA sequences (or genomic DNA sequences for rice) of the EST. If not, the positions of SSRs relative to the coding region were determined by BLASTX or ORF (open reading frame) finder analysis. Putative functions of SSR-containing ESTs were obtained using BLASTX queries with an E value $< 1 \times 10^{-20}$ cutoff.

Primer selection and screening

EST sequences containing SSRs of wheat and rice within a supercluster were aligned to identify conserved SSR flanking sequences using BLAST2, pairwise BLAST search, or ClustalW (http://www.ebi.ac.uk/clustalw). Conserved regions were used for the design of a common primer (species non-specific primer) where possible. Otherwise, species-specific primer pairs were designed for ESTs within a supercluster. PRIMER3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) was used to design primers with the following criteria: 57–61 °C melting temperature, 40%–60% GC content, 18–22 bp primer length, and 100–300 bp amplicon size. PCRs were carried out in a 25-μL solution containing 50 and 25 ng of genomic DNA template for wheat and other species, respectively, 5.5 pmol of each primer, 2.5 mM $MgCl_2$, 0.125 mM of each dNTP, 10× reaction buffer, 0.8 U *Taq* polymerase. After one denaturing step of 3 min at 94 °C, a touchdown amplification program was performed (Don et al. 1991) on a Peltier Thermal Cycler (MJ Research, Waltham, Mass.). This profile consisted of a denaturing step of 1 min at 94 °C and an extension step of 2 min at 72 °C. The initial annealing step was 1 min at 64 °C for one cycle and subsequently was reduced by 1 °C for every cycle until a final temperature of 55 °C was reached. The annealing temperature of 55 °C was employed for the last 32 cycles of the amplification, followed by one cycle of 72 °C for 10 min. PCR products from all species were first run on 1.5% *w*/*v* agarose gels for the confirmation of successful amplication, and then PCR products of wheat and rice were separated on denaturing polyacrylamide gels consisting of 4% *w*/*v* acrylamide solution, followed by silver staining.

A total of 263 primer pairs amplifying wheat-specific EST-SSRs was provided by Kansas State University (KSU; Manhattan, Kans.). To eliminate redundancy, a cluster analysis was conducted using stackPACK software (Miller et al. 1999) that identified 125 duplicates, leaving 138 KSU primer pairs for mapping.

Linkage mapping

The EST-SSR marker loci were integrated into a framework map composed of 442 loci that were retrieved from the GrainGenes Web site (http://wheat.pw.usda.gov/index.shtml). The framework map loci were selected using the following criteria: less than 30% of data missing for the 46 progeny used in this study, no significant segregation distortion ($p > 0.05$), and no duplicated, co-segregating loci. The linkage map was constructed using G-MENDEL (Holloway and Knapp 1993; www.css.orst.edu/g-mendel). Loci were grouped and ordered using a logarithm of odds (LOD) threshold of 3.0 and a recombination frequency threshold less than 0.25. If loci coalesced into more than one linkage group corresponding to a single known linkage group, then groups and orders were re-estimated using an LOD threshold of 4.0. Map distances (cM) were calculated using the Kosambi (1944) mapping function. The reproducibility of locus orders was checked by performing 100 replicated runs of the locus-ordering algorithm (via the MONTE function). The orientation of the chromosomes was determined using information from the GrainGenes Web site.

Primer pairs for several unlinked polymorphic markers and markers that were monomorphic with respect to the mapping parents were assayed on DNAs of nullisomic-tetrasomic lines to determine the chromosome location of the amplicon(s).

The EST-SSR markers, developed at Cornell University (Ithaca, N.Y.) and Kansas State University were mapped at

Cornell University. The marker loci were designated for *Xcnl39* to *Xcnl158* (developed at Cornell University) and *Xksum1* to *Xksum255* (developed at KSU). Where duplicated loci were mapped within the same chromosome, consecutive numbers were used to identify individual loci (e.g., *Xcnl143.1* and *Xcnl143.2*; http://wheat.pw.usda.gov/ggpages/wgc/98/Intro.htm#Intro5). Further information regarding EST-SSRs, primer sequences, mapping data, and chromosome location has been posted on the GrainGenes Web site under the section "Triticeae EST-SSR coordination" (http://wheat.pw.usda.gov/ITMI/EST-SSR/Cornell/).

Results and discussion

EST-SSR sequence analysis

Over 260 000 EST sequences from barley, maize, rice, sorghum, and wheat were examined to identify SSRs and for clustering of EST-SSR sequences within each species to produce a consensus sequence (Kantety et al. 2002). The consensus and singleton sequences from each species were pooled and clustered to construct 172 cross-species clusters, hereafter referred to as superclusters, and 138 singletons (Kantety et al. 2002). Two superclusters (Nos. 1 and 3) contained more than one putative function by BLASTX queries. Therefore, the default similarity stringency was increased from 85% to 96%; thus, superclusters 1 and 3 were subdivided into 26 and 23 sub-superclusters, respectively. The majority of these sub-superclusters were formed between the most closely related species, wheat and barley. As a result, a total of 219 superclusters and 138 singletons was evaluated to develop SSR markers (Fig. 1; http://wheat.pw.usda.gov/ITMI/EST-SSR/Cornell/).

Out of 357 EST-SSRs examined, 70% of the motifs were trinucleotides and 74% of these were found in coding regions, the remainder were located in 5′ untranslated regions (UTRs) (20%) and 3′ UTRs (6%). In contrast to Kantety et al. (2002), but in agreement with Varshney et al. (2002), the sequence CCG–GGC was the most frequent trinucleotide motif (36%), followed by CAA–GTT (14%). The most frequent occurrence of the CCG–GGC motif was in coding regions (71%), as reported by other researchers for sugarcane (Cordeiro et al. 2001), barley (Thiel et al. 2003), rice, maize, and sorghum (Kantety et al. 2002; Varshney et al. 2002). In rice, among the more than 7000 SSRs found in predicted genes, 92% were trinucleotides, and the motif CCG was the most frequent (Temnykh et al. 2001; Goff et al. 2002; Yu et al. 2002*b*). Dinucleotides made up 16% of the SSRs and AG–TC was the predominant motif (63%). The majority of the repeats were localized in the non-translating region (81%), and in contrast to trinucleotides, there was no significant difference between the frequency of 5′ and 3′ UTR localization (42% and 39%).

SSR marker development

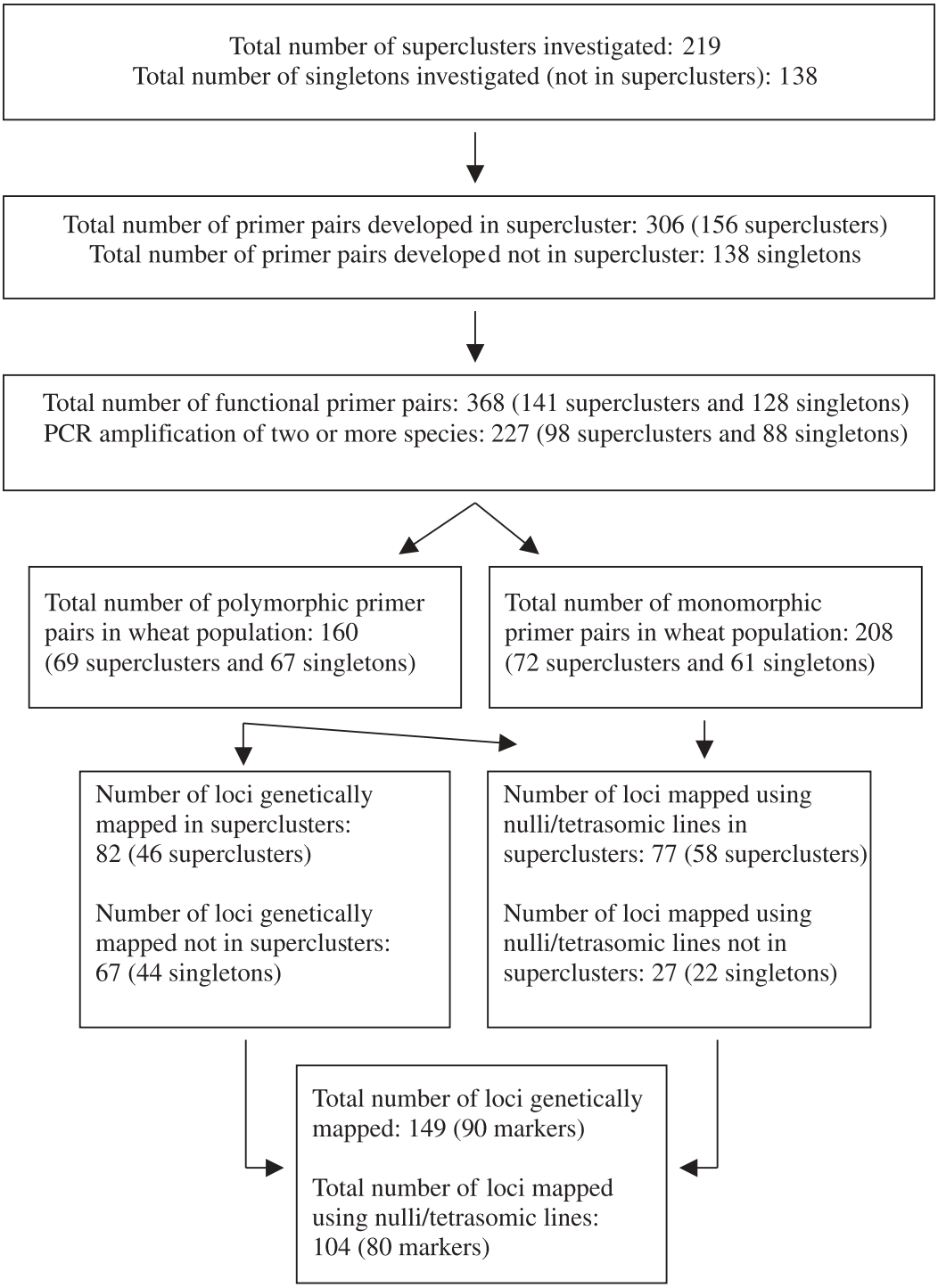
Cluster analysis was previously used to study the EST-SSR sequence relationships among five different species (Kantety et al. 2002); however for this study, priority was given to the EST sequences of wheat and rice to facilitate the assessment of interspecific transferability and comparative relationship beyond marker development and mapping. Out of 219 superclusters, it was possible to design flanking oligonucleotide primer pairs for 156 superclusters (71% of all superclusters) (Fig. 1). Primers could not be designed for 63 superclusters because of insufficient flanking sequence. From 156 superclusters, 79 were used to design more than one primer pair resulting in 306 primer pairs. A total of 57 out of the 79 superclusters was used to design species-specific primers for wheat and rice within the same supercluster. In contrast, the consensus sequences of five superclusters (number 25, 101, 166, 223, and 1007) showed high similarity between wheat and rice, thus allowing the design of common primer pairs (species non-specific primers) for wheat and rice. The EST-SSR sequences developed by Kansas State University (KSU) were integrated into this cluster analysis later resulting in several multiple primer pairs for the same cluster. Fifteen KSU primer pairs turned out to be in supercluster number one, and 14 were in supercluster 3p3946. However, each primer pair in the same supercluster was developed from slightly different EST sequences and only one was chosen from each supercluster for mapping. In addition, 138 EST-SSR markers were designed for the singleton ESTs resulting in a total of 444 primer pairs (306 from superclusters and 138 from singletons) (Fig. 1). These were evaluated using 16 genotypes from four species for marker functionality and transferability on agarose gels (http://wheat.pw.usda.gov/ITMI/EST-SSR/Cornell/).

Out of 444 primer pairs, 368 (141 superclusters and 128 singletons) produced amplicons for at least one species, and the wheat and rice amplicons generated were screened for length polymorphisms (Fig. 1). Of the EST-SSR primer pairs that produced amplicons, 62% were from the template of coding regions and 38% from UTRs; however, this difference was not statistically significant suggesting that primer functionality and their origin are unrelated. The 76 primer pairs that were rejected either produced no amplification products or complex non-specific banding patterns. Out of the 181 primer pairs developed at Cornell University, 129 primer pairs (71%) produced amplicons. Because the functionality of KSU primer pairs was tested at KSU and only functional primer sequences were provided for mapping at Cornell, statistics were not available for those primers. This functionality rate was higher than EST-SSRs reported for rice (54%) (Cho et al. 2000), sugarcane (60%) (Cordeiro et al. 2001), and barley (64%) (Thiel et al. 2003), but not rye (74%) (Hackauf et al. 2002). A total of 52 primer pairs failed to produce amplicons, possibly because either an intron occurred within the primer sequences interrupting amplification, or a large intron disrupted PCR extension. Approximately 46% of the primer pairs amplified at least one PCR fragment size larger than expected. For example, the expected product size for cnl97 was 355 bp, but one of the PCR amplicons was about 2 kb (http://wheat.pw.usda.gov/ITMI/EST-SSR/Cornell/). The occurrence of such variant alleles is not likely due to repeat length variability within the SSRs, but rather is the result of insertion–deletion (indel) polymorphism within the amplicon.

SSR marker polymorphism

Out of 368 functional primer pairs (141 superclusters and 128 singletons), 160 identified polymorphism between hexaploid wheat mapping parents. These represented 69 superclusters and 67 singletons (51% of 269 markers)

Fig. 1. Schematic overview summarizing the strategy for EST-SSR marker development and mapping in wheat.



(Fig. 1). A previous study of durum wheat (Eujayl et al. 2002) reported approximately 25% polymorphism from EST-derived SSRs and concluded that EST-SSR primers produced high-quality markers, but had low polymorphism compared with genomic SSR markers. However, in this study, marker polymorphism from ESTs was comparable to genomic SSRs (53%) in hexaploid wheat (Röder et al. 1998). Of the four primer pairs (cnl52, cnl63, cnl77 and cnl78) based on consensus sequences of wheat and rice (supercluster numbers 25, 101, and 223), only one was polymorphic (cnl77, supercluster number 223). The high cDNA sequence similarity across two different species implies that the genes in these superclusters are highly conserved. The common primer pairs (species non-specific) from superclusters 166 and 1007 did not produce amplification products.

Dinucleotides were significantly more polymorphic (74%) than trinucleotides (56%). The majority of dinucleotides

Fig. 2. Linkage map of the hexaploid wheat cross between W7984 and Opata 85 showing the position of 149 loci from 90 EST-SSR markers. The distances (cM) between marker loci are shown on the right and the marker loci on the left of each map. The new EST-SSR loci (*Xcnl* and *Xksum*) that were integrated into a framework map of 442 loci are shown in bold italics. Maps of chromosome 1D, 6D and 7D were not presented because of the lack of newly mapped EST-SSR loci. Mapped loci prefixed with asterisk (*) showed segregation distortion, $p < 0.05$.

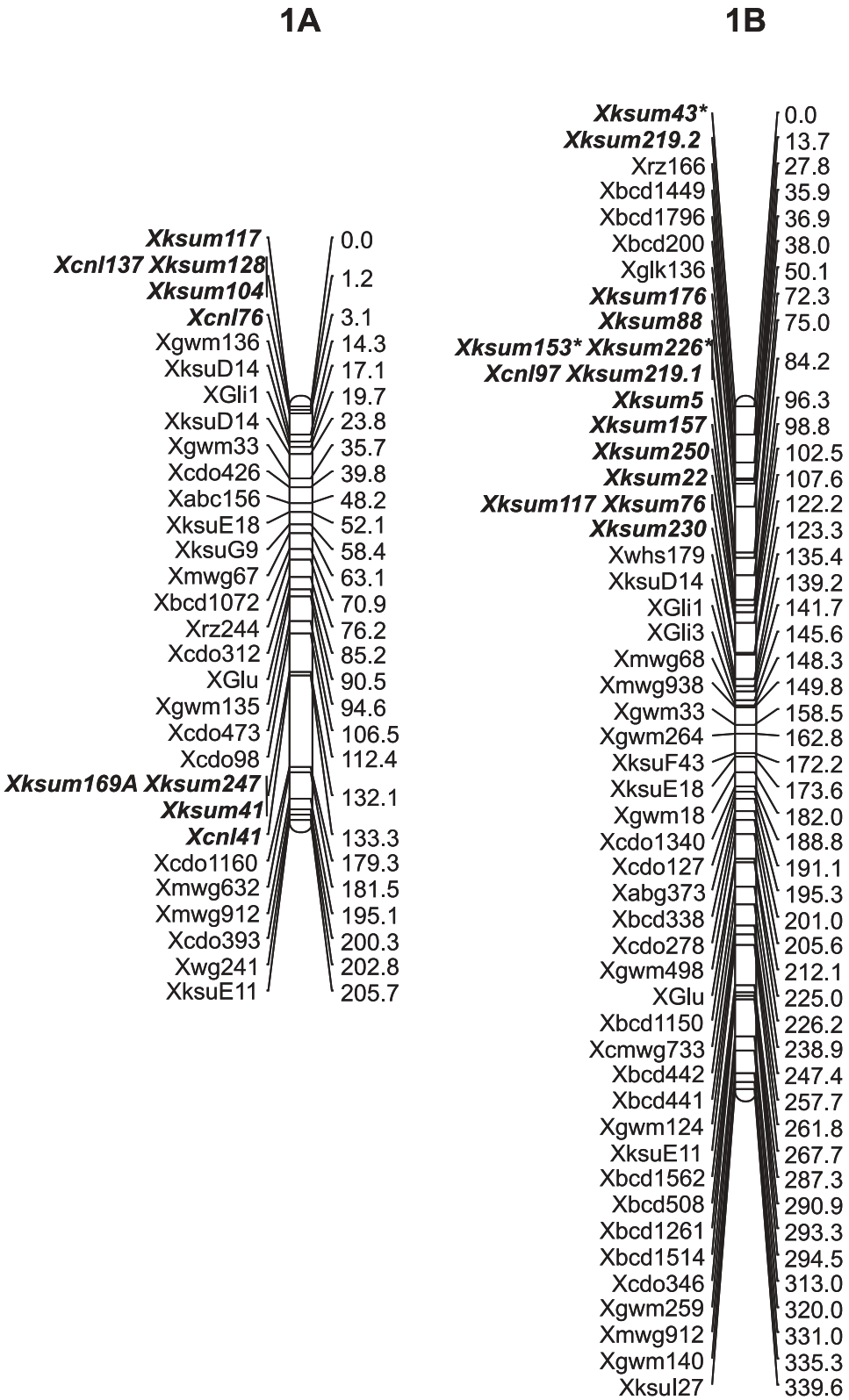
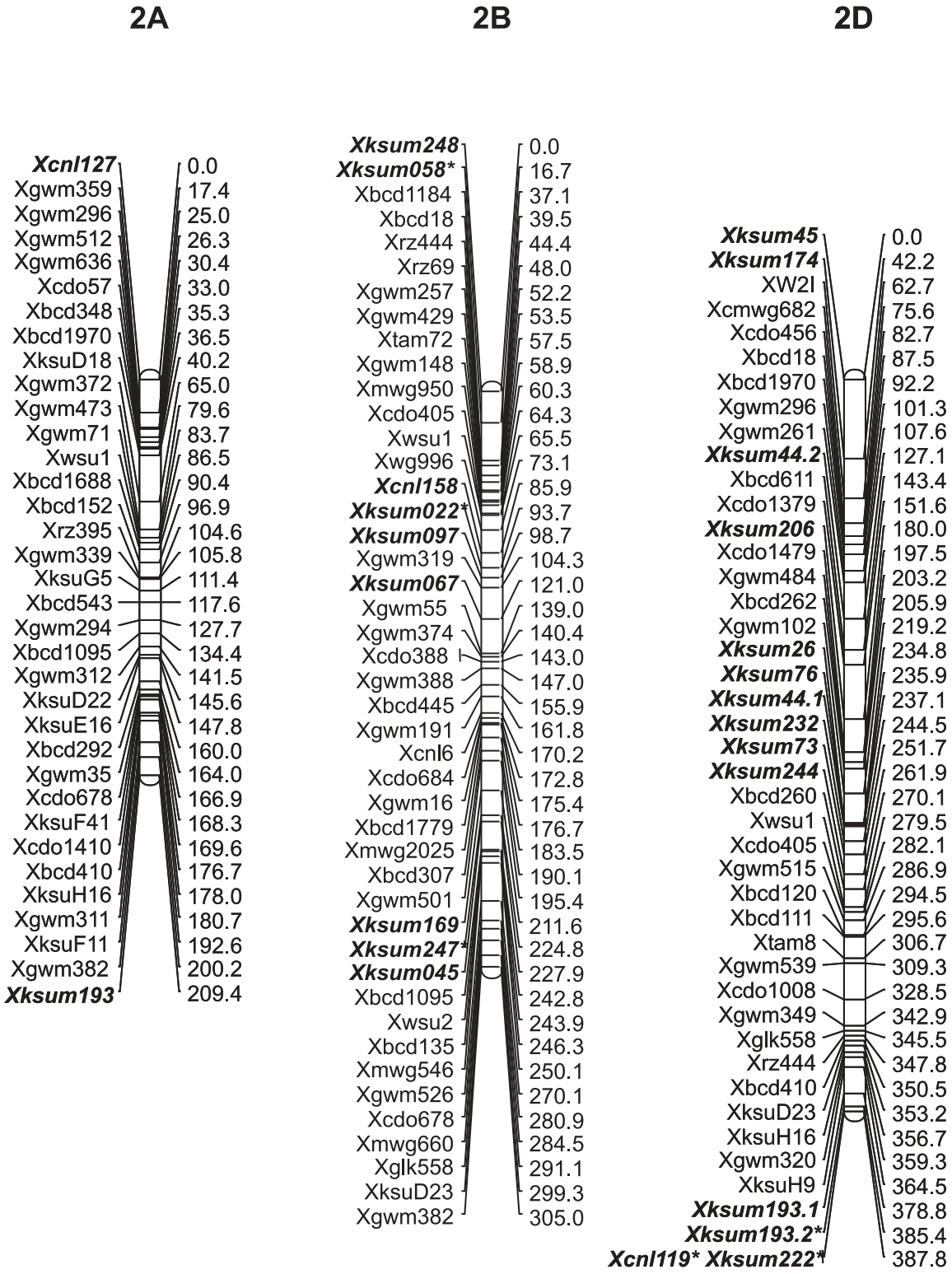


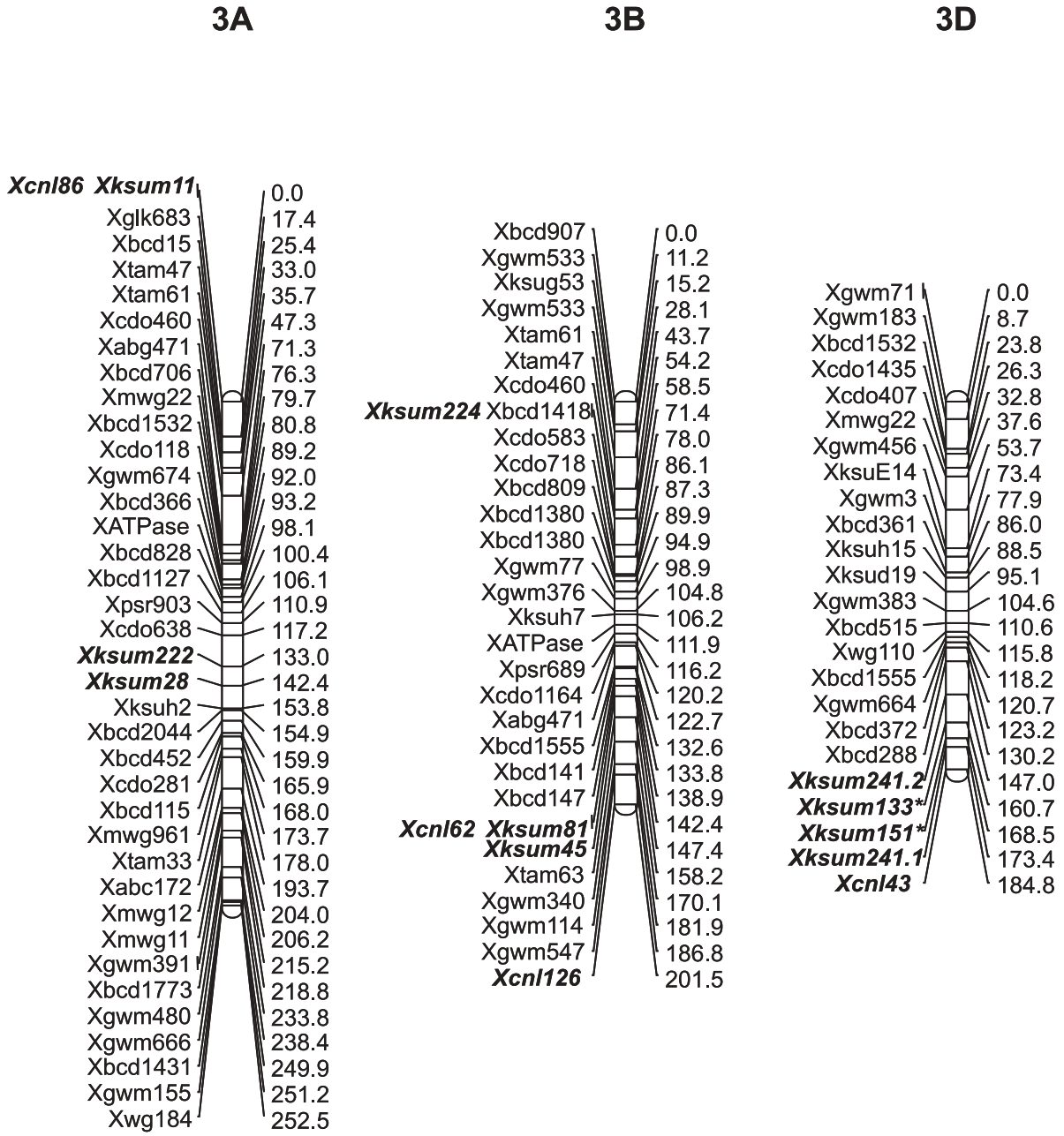
Fig. 2 (*continued*).



were located in the non-translating regions (81%), which may enhance their polymorphism compared with the location of trinucleotides (Thiel et al. 2003). Markers from UTR regions have been reported to be more polymorphic because sequences from UTRs do not participate directly in gene function (Scott et al. 2000; Temnykh et al. 2001; Thiel et al. 2003). However, in this study, EST-SSRs from both coding and 5′ UTR regions yielded a similar proportion of polymorphic SSRs (61% and 60%, respectively) and both were significantly more polymorphic than those in 3′ UTRs (49%).

Most primers from the same supercluster produced similar banding patterns in the mapping parents. In contrast, there

Fig. 2 (*continued*).

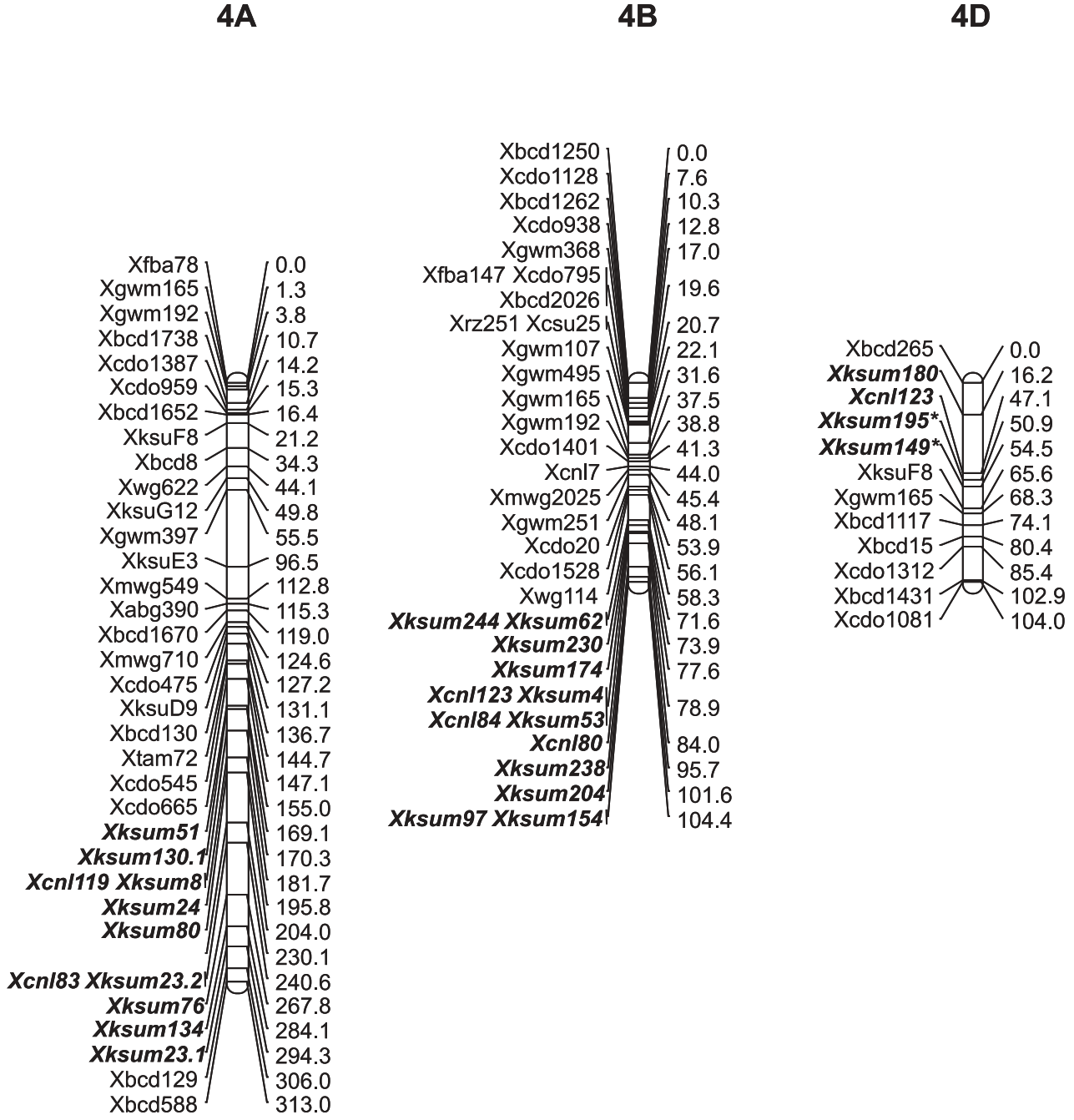


were primer pairs from 10 superclusters where one generated polymorphism and the other did not. For these superclusters, there was no relationship between SSR location (e.g., coding region or UTR) and polymorphism. Of the 160 EST-SSR primer pairs (representing 69 superclusters and 67 singletons) polymorphic for the ITMI population parents (Fig. 1), 124 (representing 55 superclusters and 57 singletons) were used for mapping. The other 36 generated complicated banding patterns, weak band intensity, or extremely large or small band size.

SSR marker transferability

The 368 functional primer pairs were tested on 16 genotypes from four species (wheat, rice, barley, and maize) and 227 (98 superclusters and 88 singletons) successfully amplified DNA from two or more species (Fig. 1 and http://wheat.pw.usda.gov/ITMI/EST-SSR/Cornell/). Among these, 49 primer pairs (23 superclusters and 21 singletons) amplified DNA in all four species (http://wheat.pw.usda.gov/ITMI/EST-SSR/Cornell/). Based on the 227 successful primer pairs, 67% from coding regions produced PCR amplicons in two or more species (22% were from 5′ UTR and 11% from 3′ UTR). To construct superclusters from a previous study, Kantety et al. (2002) used an 85% sequence similarity threshold to cluster EST sequences from two or more species. In the present study, primers designed from superclusters showed a slightly higher frequency of transferability than those from singletons, 74% and 69%, respectively. In

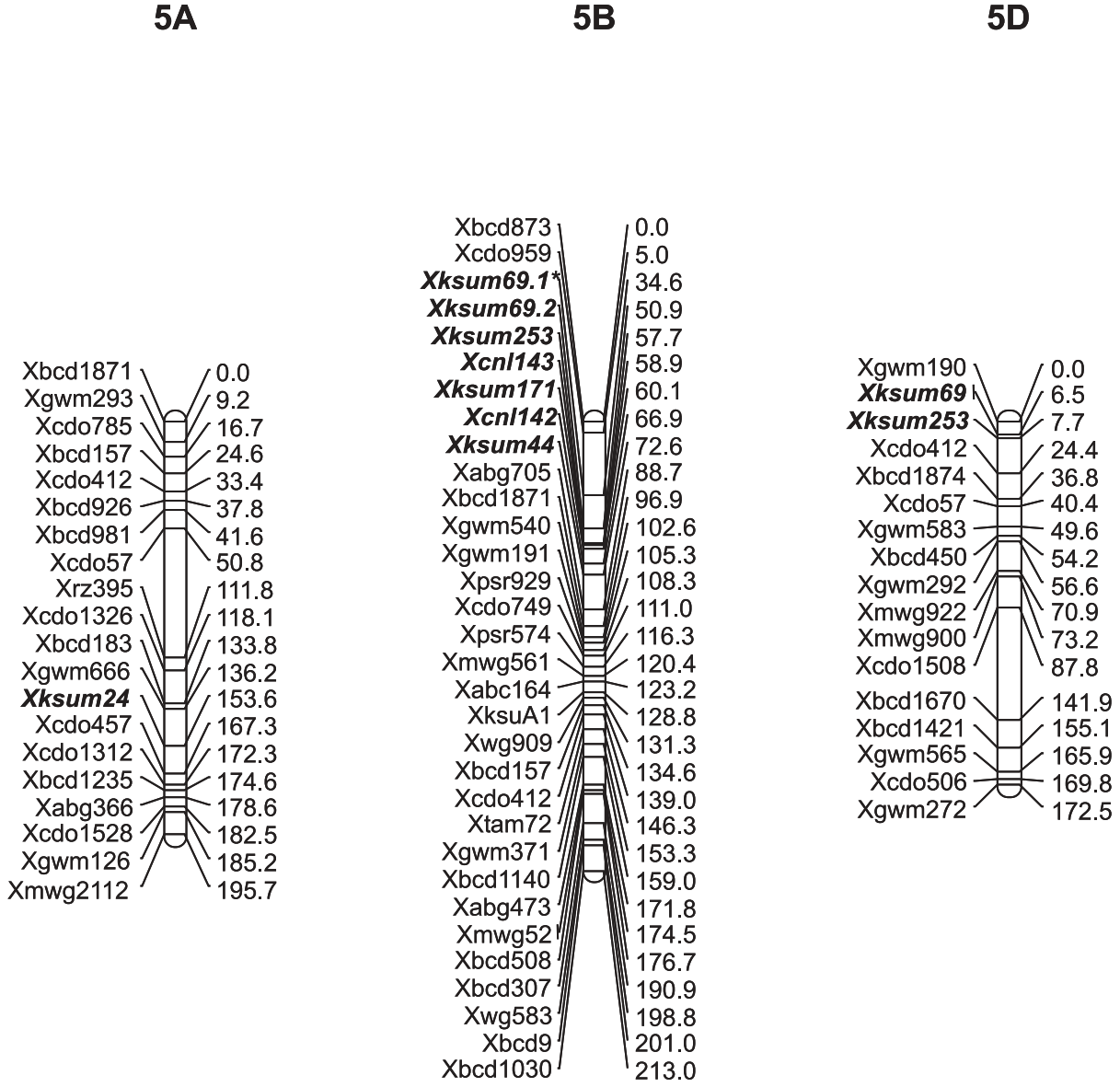
Fig. 2 (*continued*).



addition, BLASTN analysis with an E value $< 1 \times 10^{-10}$ was performed to cluster multiple alignments and conserved regions flanking SSRs. These criteria allowed the development of EST-SSR primers that amplify orthologous loci in multiple species, even when using individual species-specific primers. Of 337 primer pairs derived from wheat EST sequences, 180 (53%), 151 (45%), and 57 (17%) produced amplicons from barley, rice, and maize DNA, respectively (http://wheat.pw.usda.gov/ITMI/EST-SSR/Cornell/). In addition, out of the 57 wheat-derived primer pairs amplifying maize DNAs, 55 primer pairs produced amplicons on more than three species, suggesting that these markers were developed in relatively conserved gene sequences. Several studies have reported that SSRs were highly conserved between barley and wheat but less conserved relative to the more distantly related rice and maize (Holton et al. 2002; Kantety et al. 2002; Thiel et al. 2003).

A total of 162 primer pairs (from 66 superclusters and 61 singletons) generated polymorphism (77% of 165 rice markers; 95 superclusters and 70 singletons) between the rice mapping parents (IR64 and Azucena; Temnykh et al. 2001) and 95% of 127 polymorphic rice markers were developed based on wheat sequences. A total of 86 primer pairs (40 superclusters and 37 singletons) was polymorphic for both

Fig. 2 (*continued*).



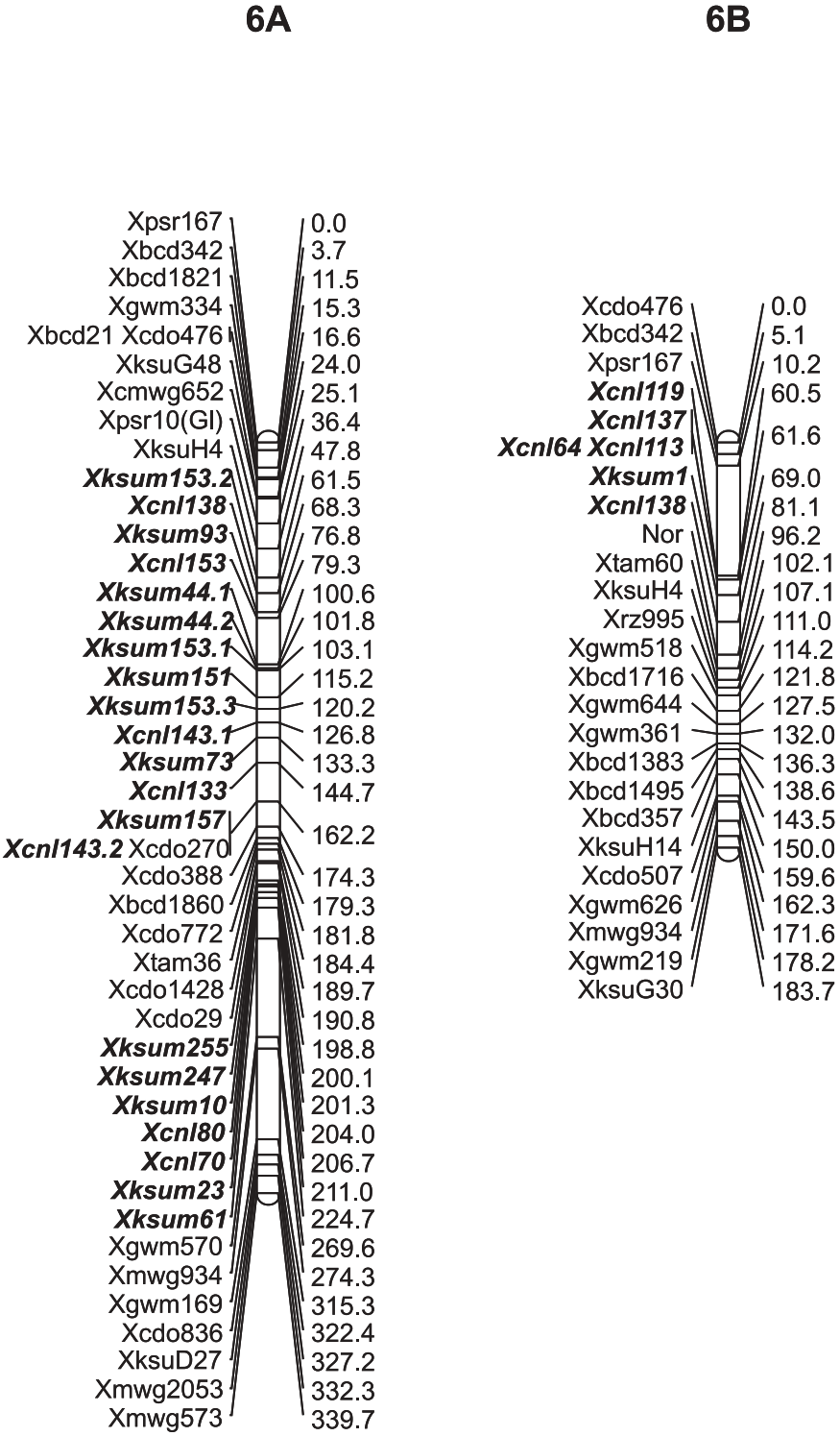
wheat and rice mapping population parents resulting in 77 EST-SSR loci for enhancing comparative maps between wheat and rice.

Genetic mapping of hexaploid wheat

For the ITMI population mapping parents, 124 (112 non-redundant markers; 55 superclusters and 57 singletons) out of 160 polymorphic primer pairs were used to map 213 loci. We included redundant primer pairs for initial screening of mapping progeny for two reasons, (*i*) to evaluate the parameters used in sequence matching for supercluster analysis in a laboratory-based experiment and (*ii*) to identify as many polymorphic loci from the same superclusters as possible. However, one polymorphic primer pair was chosen for mapping from each supercluster except for superclusters 1 and 109. Seventy-five primer pairs (61%) amplified a single locus and 48 (39%) amplified two to six loci. As expected, the proportion of multi-loci detecting markers was twice that of markers derived from genomic DNA (20%) (Röder et al. 1998), possibly owing to sequence conservation in coding regions, polyploidy, and 25%–30% gene duplication (Anderson et al. 1992). Loci amplified by all but eight primer pairs detecting multiple loci were mapped in non- homoeologous groups.

A total of 213 segregating loci from 124 primer pairs (112 non-redundant markers) was integrated into a published framework map of RFLP and SSR on the ITMI population spanning all 21 linkage groups (Nelson et al. 1995*a*, 1995*b*, 1995*c*; Van Deynze et al. 1995; Marino et al. 1996; Röder et al. 1998). The assigned loci were not uniformly distributed across all chromosomes. A total of 149 loci from 90 markers (46 superclusters and 44 singletons) was assigned to all chromosomes except 1D, 6D, and 7D (Figs. 1 and 2). Twenty-three loci from 20 primer pairs representing 8 superclusters (223, 224, 406, 426, 536, 1021, 3p2879, and 3p3804) and 12 singletons were not linked at LOD 3. The marker loci from the same superclusters were mapped at the same loci except for those from superclusters No. 1 and 109.

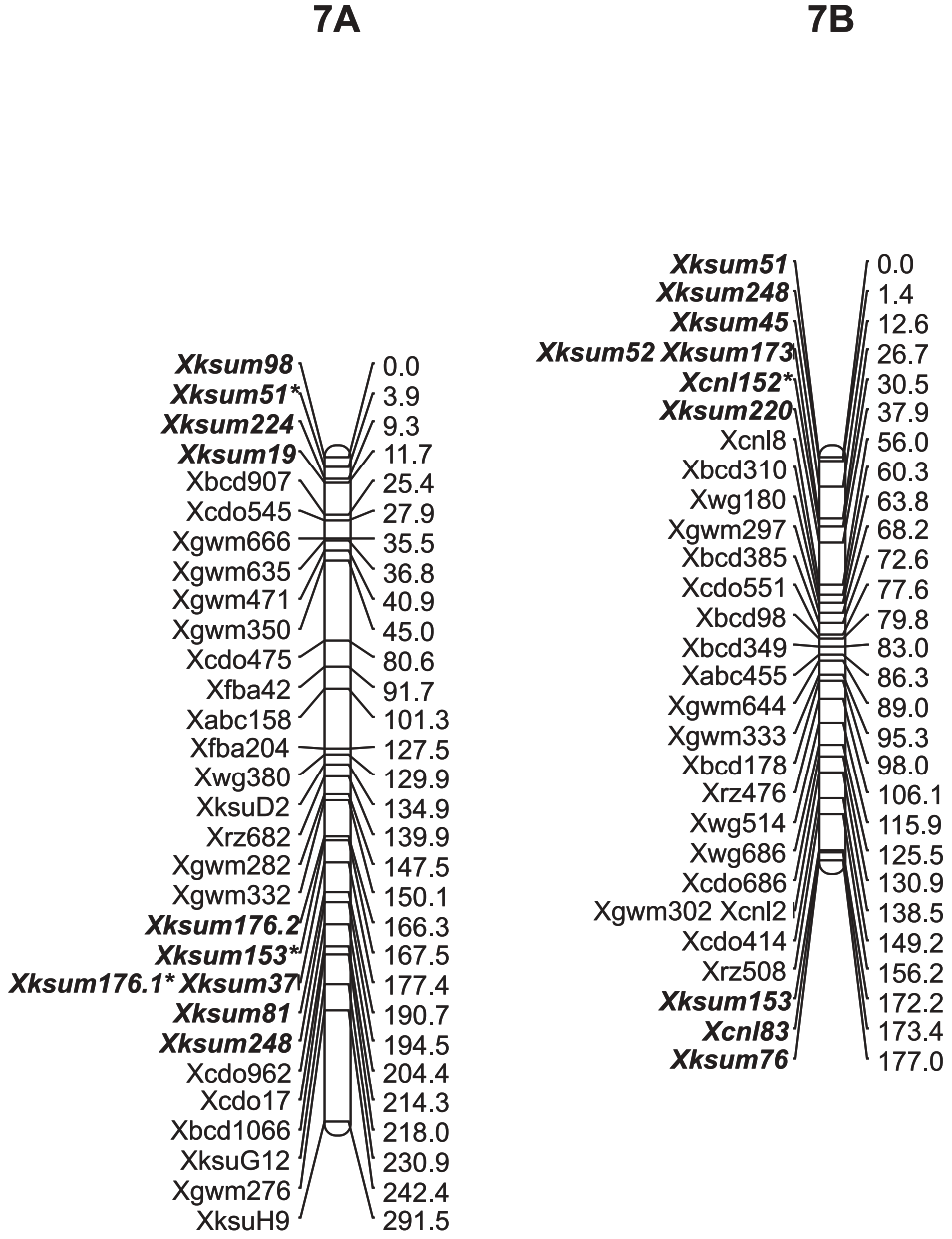
Fig. 2 (*continued*).



The marker loci derived from duplicated primer pairs were not presented in this map because of cosegregation.

Out of the 149 loci, 43% mapped to the B genome, 40% to the A genome, and 17% to the D genome (Fig. 2). This is in agreement with Röder et al. (1998) where the highest number of genomic SSR marker loci was on the B genome and the lowest was on the D genome. There is generally a low level of polymorphism in wheat relative to other cereal species; the D genome in particular is less diverse in bread wheat than the A and B genomes. Only 13% of the loci in this study showed segregation distortion compared with 30% for previously published ITMI maps (Röder et al. 1998). Two loci (*Xksum* 202 and *Xksum* 229) out of the 23 that failed to integrate into any of the linkage groups showed dis-

Fig. 2 (*concluded*).



torted segregation ($p < 0.05$). Half of the distorted loci (9 of 19) were distributed in the distal regions of chromosomes (Fig. 2).

The mapped marker loci showed significant clustering in the distal regions of chromosome arms (Fig. 2), which is consistent with the deletion line-based physical mapping of single or low copy number cDNA-RFLP clones in wheat and other members of the Triticeae (Gill et al. 1996; Moore 2000; Sandhu and Gill 2002). Chromosomes 1B, 4A, 4B, and 6A each have a small cluster comprising 15, 12, 13, and 14 loci, respectively (Fig. 2). The loci from these four clusters contained 36% of the total loci, but these high-density gene regions covered only 12% of the genome. Interestingly, clusters of loci from 15 of 19 mapped chromosomes were placed into the distal regions of chromosomes in contrast to genomic SSR maps (Röder et al. 1998) where loci generally clustered in proximal regions (Fig. 2). In addition, EST-SSR marker loci tended to map in relatively large gaps. Recombination in the telomeric region (distal 25%) of the chromosome 5 long arm was reported to be more than 20 times the recombination in the centromeric region (proximal 25%) (Gill et al. 1996) and from a maize study, the regions of active recombination tended to be gene-rich regions (Yao et al. 2002; Ware and Stein 2003). The present study suggests that EST-SSRs tend to map to regions of high recombination where markers are less likely to be identified using genomic DNA-based markers. The preferential localization in recombination hot spots enhances the value of these markers.

The use of EST-SSRs is an efficient method for mapping expressed genes. Using BLASTX queries (E value $< 1 \times 10^{-20}$) putative functions were assigned to 22 superclusters (http://wheat.pw.usda.gov/ITMI/EST-SSR/Cornell/). For example,

two loci derived from KSUM81 from supercluster 1p2382, for which the consensus EST sequences showed high homology with *thioredoxin* h (*E* value, 2×10^{-27}), were mapped on wheat chromosomes 3B and 7A. In cereals, *thioredoxin* h is abundant in seeds (Serrato et al. 2002). Because *thioredoxin* h inactivates an inhibitor of hydrolytic enzymes and activates serine proteases at early stages of germination, it has been proposed that they function in signaling in cereal seed germination (Jiao et al. 1993; Besse et al. 1996; Rodriguez-Lopez et al. 2000). One could speculate that because *Xksum81* loci on 3L and 7L coincide with QTL for resistance to pre-harvest sprouting that have been mapped to the long arms of chromosome 3 (Groos et al. 2002) and chromosome 7D (Roy et al. 1999), this gene may be a candidate for control of this trait. Therefore, these putative germination-related loci and EST-SSR markers may be useful for marker-assisted selection for pre-harvest sprouting resistance.

A total of 104 loci from 80 EST-SSR markers, including 18 monomorphic markers, 14 unlinked markers, and 48 genetically mapped markers, was mapped to chromosome using nullisomic-tetrasomic lines (Fig. 1; http://wheat.pw.usda.gov/ITMI/EST-SSR/Cornell/). As mentioned earlier, no EST-SSR loci were mapped on D genome chromosomes 1, 6, or 7 because of a lack of polymorphism. However, using these aneuploid stocks, *Xcnl132* was mapped to 1D, *Xcnl142* and *Xksum40* to 6D, and three markers, *Xcnl71*, *Xcnl130*, and *Xksum108*, to 7D. Ten primer pairs amplified loci that mapped to chromosomes different from those genetically mapped. This is because some bands that were monomorphic for the mapping population parents were not genetically mapped but could be assigned a chromosome location using the aneuploids. Not all of the EST-SSRs used in the present study were mapped using the aneuploids, but this information may be useful for other studies.

In this study, (*i*) 368 functional EST-SSR primer pairs from 141 superclusters and 128 singletons were developed, (*ii*) 227 EST-SSR primer pairs were found to be useful as cross-species transferable markers, (*iii*) 149 EST-SSR loci from 90 marker sets were genetically mapped in the ITMI wheat population, (*iv*) putative functions were assigned to 22 mapped ESTs, and (*v*) 80 EST-SSR markers amplified 104 loci located to chromosome arms using nullisomic-tetrasomic lines. In conclusion, EST databases provide a valuable source of sequences for SSR marker development that corresponds to transcribed genes. An enhanced ITMI map with EST-SSR loci can provide a basis for comparative mapping using orthologous markers and possibly for identifying expressed genes associated with agronomically important traits in wheat and other grass species. In addition, this map may be used as a reference for enhancing the collaborative effort of researchers working towards developing a more comprehensive wheat genetic map using public EST databases.

Acknowledgements

The authors would like to thank Ramesh V. Kantety, Mauricio La Rota, and Qi Sun for their assistance in EST sequence analysis. We also thank David E. Matthews and Hugh Edwards for presentation of our data in the GrainGenes Web site (www.wheat.pw.usda.gov). This publication is based upon work supported by USDA–NRI project No. 2001-35301-10612 and Hatch Project No. 149419.

References

Anderson, J.A., Ogihara, Y., Sorrells, M.E., and Tanksley, S.D. 1992. Development of a chromosomal arm map for wheat based on RFLP markers. Theor. Appl. Genet. **83**: 1035–1043.

Besse, I., Wong, J.H., Kobrehel, K., and Buchanan, B.B. 1996. Thiocalsin: a thioredoxin-linked, substrate-specific protease dependent on calcium. Proc. Natl. Acad. Sci. USA. **93**: 3169–3175.

Chen, X., Salamini, F., and Gebhardt, C. 2001. A potato molecular-function map for carbohydrate metabolism and transport. Theor. Appl. Genet. **102**: 284–295.

Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S.R., Park, W.D., Ayers, N., and Cartinhour, S. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100**: 713–722.

Cordeiro, G.M., Casu, R., McIntyre, C.L., Manners, J.M., and Henry, R.J. 2001. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci. **160**: 1115–1123.

Decroocq, V., Fave, M.G., Hagen, L., Bordenave, L., and Decroocq, S. 2003. Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor. Appl. Genet. **106**: 912–922.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., and Mattick, J.I. 1991. "Touchdown" PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. **19**: 4008.

Edwards, K.J., Barker, J.H., Daly, A., Jones, C., and Karp, A. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques, **20**: 758–760.

Endo, T.R., and Gill, B.S. 1996. The deletion stocks of common wheat. J. Hered. **87**: 295–307.

Eujayl, I., Sorrells, M.E., Baum, M., Wolters, P., and Powell, W. 2002. Isolation of EST-derived microsatellite markers for phenotyping the A and B genomes of wheat. Theor. Appl. Genet. **104**: 399–407.

Gill, K.S., Gill, B.S., Endo, T.R., and Boyko, E.V. 1996. Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. Genetics, **143**: 1001–1012.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science (Washington, D.C.), **296**: 92–100.

Groos, C., Gay, G., Perretant, M.R., Gervais, L., Bernard, M., Dedryver, F., and Charmet, G. 2002. Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white × red from bread–wheat cross. Theor. Appl. Genet. **104**: 39–47.

Gupta, P.K., and Varshney, R.K. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica, **113**: 163–185.

Gupta, P.K., Varshney, R.K., Sharma, P.C., and Ramesh, B. 1999. Molecular markers and their applications in wheat breeding. Plant Breed. **118**: 369–390.

Gupta, P.K., Balyan, H.S., Edwards, K.J., Isaac, P., Korzun, V., et al. 2002. Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor. Appl. Genet. **105**: 413–422.

Hackauf, B., and Wehling, P. 2002. Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed. **121**: 17–25.

Heun, M., Kennedy, A.E., Anderson, J.A., Lapitan, N.L.V., Sorrells, M.E., and Tanksley., S.D. 1991. Construction of an RFLP map for barley (*Hordeum vulgare* L.). Genome, **34**: 437–446.

Holton, T.A., Christopher, J.T., McClure, L., Harker, N., and Henry, R.J. 2002. Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Mol. Breed. **9**: 63–71.

Holloway, J.L., and Knapp, S.J. 1993. GMendel 3.0 Users guide. Department of Crop and Soil Science, Oregon State University, Corvallis, Ore.

Jiao, J., Yee, B.C., Wong, J.H., Kobrehel, K., and Buchanan, B.B. 1993. Thioredoxin-linked changes in regulatory properties of barley α-amylase – subtilisin inhibitor protein. Plant Physiol. Biochem. **31**: 799–804.

Kantety, R.V., Rota, M.L., Matthews, D.E., and Sorrells, M.S. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum, and wheat. Plant Mol. Biol. **48**: 501–510.

Kosambi, D.D. 1944. The estimation of map distance from recombination values. Ann. Eugen. **12**: 172–175.

Kruglyak, S., Durrett, R., Schug, M.D., and Adqadro, C.F. 1998. Equilibrium distribution of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc. Natl. Acad. Sci. USA. **95**: 10 074 – 10 078.

Marino, C.L., Nelson, J.C., Lu, Y.H., Sorrells, M.E., and Leroy, P. 1996. Molecular genetic maps of the group 6 chromosomes of hexaploid wheat (*Triticum aestivum* L. em. Thell). Genome, **39**: 359–366.

Miller, R.T., Christoffels, A.G., Gopalakrishnan, C., Burke, J., Ptitsyn, A.A., Broveak, T.R., and Hide, W.A. 1999. A comprehensive approach to clustering of the expressed human gene sequence: the sequence tag alignment and consensus knowledge base. Genome Res. **9**: 1143–1155.

Moore, G. 2000. Cereal chromosome structure, evolution, and pairing. Annu. Rev. Plant Physiol. Plant Mol. Biol. **51**: 195–222.

Morgante, M., and Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J. **3**: 175–182.

Nelson, J.C., Van Deynze, A.E., Autrique, E., Sorrells, M.E., Lu, Y.H., Merlino, M., Atkinson, M., and Leroy, P. 1995*a*. Molecular mapping of wheat: Homoeologous group 2. Genome, **38**: 516–524.

Nelson, J.C., Van Deynze, A.E., Autrique, E., Sorrells, M.E., Lu, Y.H., Negre, S., Bernard, M., and Leroy, P. 1995*b*. Molecular mapping of wheat: homoeologous group 3. Genome, **38**: 525–533.

Nelson, J.C., Sorrells, M.E., Van Deynze, A.E., Lu, Y.H., Atkinson, M., Bernard, M., Leroy, P., Faris, J.D., and Anderson, J.A. 1995*c*. Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. Genetics, **141**: 721–731.

Pestsova, E., Ganal, M.W., and Röder, M.S. 2000. Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome, **43**: 689–697.

Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, N.-H., Leroy, P., and Ganal, M.W. 1998. A microsatellite map of wheat. Genetics, **149**: 2007–2023.

Rodriguez-Lopez, M., Baroja-Fernandez, E., Zandueta-Criado, A., and Pozueta-Rpmero, J. 2000. Adenosine diphosphate glucose pyrophosphatase: a plastidial phosphodiesterase that prevents starch biosynthesis. Proc. Natl. Acad. Sci. USA. **97**: 8705–8710.

Roy, J.K., Prasad, M., Varshney, R.K., Balyan, H.S., Blake, T.K., Dhaliwal, H.S., Singh, H., Edwards, K.J., and Gupta, P.K. 1999. Identification of a microsatellite on chromosomes 6B and a STS on 7D of bread wheat showing an association with preharvest sprouting tolerance. Theor. Appl. Genet. **99**: 336–340.

Sandhu, D., and Gill, K.S. 2002. Gene-containing regions of wheat and the other grass genomes. Plant Physiol. **128**: 803–811.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., and Henry, R.J. 2000. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. **100**: 723–726.

Serrato, A.J., Perez-Ruiz, J.M., and Cejudo, F.J. 2002. Cloning of thioredoxin h reductase and characterization of the thioredoxin reductase – thioredoxin h system from wheat. Biochem. J. **367**: 491–497.

Song, Q.J., Fickus, E.W., and Cregan, P.B. 2002. Characterization of trinucleotide SSR motifs in wheat. Theor. Appl. Genet. **104**: 286–293.

Stephenson, P., Bryan, G., Kirby, J., Collins, A., Devos, K., Busso, C., and Gale, M. 1998. Fifty new microsatellite loci for the wheat genetic map. Theor. Appl. Genet. **97**: 946–947.

Taylor, C., Madsen, K., Borg, S., Moller, M.G., Boelt, B., and Holm, P.B. 2001. The development of sequence-tagged sites (STSs) in *Lolium perenne* L.: the application of primer sets derived from other genera. Theor. Appl. Genet. **103**: 648–658.

Temnykh, S., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho, Y.G., Ishii, T., and McCouch, S.R. 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100**: 697–712.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. **11**: 1441–1452.

Thiel, T., Michalek, W., Varshney, R.K., and 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.

Van Deynze, A.E., Dubcovsky, J., Gill, K.S., Nelson, J.C., Sorrells, M.E., Dvorak, J., Gill, B.S., Lagudah, E.S., McCouch, S.R., and Appels, R. 1995. Molecular-genetic maps for group 1 chromosomes of *Triticeae* species and their relation to chromosomes in rice and oat. Genome, **38**: 45–59.

Varshney, R.K., Thiel, T., Stein, N., Langridge, P., and 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.

Ware, D., and Stein, L. 2003. Comparison of genes among cereals. Curr. Opin. Plant Biol. **6**: 121–127.

Yao, H., Zhou, Q., Li, J., Smith, H., Yandeau, M., Nikolau, B.J., and Schnable, P.S. 2002. Molecular characterization of meiotic recombination across the 140-kb multigenic *a1–sh2* interval of maize. Proc. Natl. Acad. Sci. USA. **99**: 6157–6162.

Yu, J.K., Mangor, J., Thompson, L., Edwards, K.J., Slabaugh, M.B., and Knapp, S.J. 2002*a*. Allelic diversity of simple sequence repeat markers among elite inbred lines in cultivated sunflower. Genome, **45**: 652–660.

Yu, J., Hu, S., Wang, J., Wong, G.K.-S., Li, S., et al. 2002*b*. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science (Washington, D.C.), **296**: 79–92.