Theor Appl Genet (2005) 111: 677–687
DOI 10.1007/s00122-005-2041-5

ORIGINAL PAPER

L. Y. Zhang · M. Bernard · P. Leroy · C. Feuillet
P. Sourdille

High transferability of bread wheat EST-derived SSRs to other cereals

Received: 23 February 2005 / Accepted: 14 April 2005 / Published online: 21 July 2005

Abstract The increasing availability of expressed sequence tags (ESTs) in wheat (*Triticum aestivum*) and related cereals provides a valuable resource of non-anonymous DNA molecular markers. In this study, 300 primer pairs were designed from 265 wheat ESTs that contain microsatellites in order to develop new markers for wheat. Their level of transferability in eight related species [*Triticum durum*, *T. monococcum*, *Aegilops speltoides*, *Ae. tauschii*, rye (*Secale cereale*), barley (*Hordeum vulgare*), *Agropyron elongatum* and rice (*Oryza sativa*)] was assessed. In total, 240 primer pairs (80%) gave an amplification product on wheat, and 177 were assigned to wheat chromosomes using aneuploid lines. Transferability to closely related Triticeae species ranged from 76.7% for *Ae. tauschii* to 90.4% for *T. durum* and was lower for more distant relatives such as barley (50.4%) or rice (28.3%). No clear putative function could be assigned to the genes from which the simple sequence repeats (SSRs) were developed, even though most of them were located inside ORFs. BLAST analysis of the EST sequences against the 12 rice pseudo-molecules showed that the EST-SSRs are mainly located in the telomeric regions and that the wheat ESTs have the highest similarity to genes on rice chromosomes 2, 3 and 5. Interestingly, most of the SSRs giving an amplification product on barley or rice had a repeated motif similar to the one found in wheat, suggesting a common ancestral origin. Our results indicate that wheat EST-SSRs show a high level of transferability across distantly related species, thereby providing additional markers for comparative mapping and for following gene introgressions from wild species and carrying out evolutionary studies.

Communicated by J.W. Snape

L. Y. Zhang · M. Bernard (✉) · P. Leroy · C. Feuillet
P. Sourdille
UMR INRA-UBP Amélioration et Santé des Plantes,
234 Avenue du Brézet, 63100 Domaine de Crouël,
Clermont-Ferrand, Cedex 2, France
E-mail: michel.bernard@clermont.inra.fr
Tel.: +33-4-73-62-43-07
Fax: +33-4-73-62-44-53

Introduction

Simple sequence repeats (SSRs), also referred to as microsatellites, are tandem arrays of short DNA repeats that range rom 1 bp to 6 bp in length. Polymorphism, which is based on the differences in the number of DNA repeats at any given locus, is easily detected by PCR and can be classified into two classes based on origin: genomic SSR markers, which are developed from enriched genomic DNA libraries, and expressed sequence tag (EST)-SSRs, which are derived from EST sequences originating from the expressed region of the genome.

Genomic SSR markers were developed during the early 1990s. They are distributed throughout the genome and have been used extensively for genome mapping, DNA fingerprinting and a wide range of genetic diversity, population and evolutionary studies in both plant and animal species (Liu et al. 1996; Senior et al. 1996; McCouch et al. 1997; Röder et al. 1998; Gupta and Varshney 2000; Prasad et al. 2000; Ramsay et al. 2000). The numerous advantages of this type of marker, including their abundance and dispersion throughout the entire genome, high information content, co-dominant inheritance, reproducibility and genomic specificity, are well-documented (Morgante and Olivieri 1993; Rafalski and Tingey 1993; Powell et al. 1996). However, most genomic SSRs have neither a genic function nor close linkage to coding regions (Metzgar et al. 2000), they are very time- and cost-expensive to develop and they show only a limited transferability to related species (Sourdille et al. 2001; Guyomarc'h et al. 2002a, b).

EST-SSRs have received much attention recently because of the increasing amounts of ESTs being deposited in databases for various economically important plants, such as rice (*Oryza sativa*), maize (*Zea mays*), wheat (*Triticum aestivum*), sorghum (*Sorghum*

bicolor), barley (*Hordeum vulgare*) and rye (*Secale cereale*). EST-SSRs can be rapidly developed from the in silico analysis of EST databases at low cost, and due to their presence in expressed regions, they can lead to the development of gene-based maps which may increase the efficiency of marker-assisted selection (MAS) through the use of candidate genes. Assessments of the polymorphism, diversity and transferability of EST-SSRs have been carried out in rice (Cho et al. 2000), grape (Scott et al. 2000), sugarcane (Cordeiro et al. 2001; Grivet et al. 2003), tomato (Areshchenkova and Ganal 2002), loblolly pine (Liewlaksaneeyanawin et al. 2004), Alpine Lady-fern (Woodhead et al. 2003), pasture grass endophytes (Jong et al. 2002), barley (Thiel et al. 2003) and rye (Hackauf and Wehling 2002). In wheat, in silico analysis showed that the frequency of EST-SSRs is 1 at every 6.2 kb of EST sequence (Varshney et al. 2002). With the rapid increase in bread wheat ESTs in the databases (561,731; http://www.ncbi.nlm.nih.gov/dbEST), EST-SSRs have become an attractive alternative to complement existing SSR collections, and 101 new EST-SSRs loci from bread wheat have recently been added to the wheat genetic map (Gao et al. 2004). Comparisons between genomic-SSRs and EST-SSRs have revealed that wheat EST-SSR markers have a lower level of polymorphism but produce higher quality patterns (Eujayl et al. 2001, 2002; Leigh et al. 2003). The genetic diversity has also been assessed in a collection of 52 elite exotic wheat genotypes (Gupta et al. 2003), and the results suggest that EST-SSRs can be successfully used for a variety of purposes and may be superior to genomic SSRs for diversity estimation.

In addition to the advantages of genomic SSR markers mentioned earlier, EST-SSRs show a high level of transferability to closely related species because they originate from conserved transcribed regions that are better conserved between the genomes; this consequently facilitates their use in comparative mapping (Yu et al. 2004a), as do restriction fragment length polymorphism (RFLP) markers derived from cDNA (for a review, see Gupta and Rustgi, 2004). The transferability of bread wheat EST-SSRs across 18 wild relatives and five cereal species (barley, rye, oat, rice and maize) was recently studied with 78 EST-SSR markers (Gupta et al. 2003). More than 80% of cross-species transferability was observed with wild relatives; this rose to as high as 90% with at least one of the cereal species. Similarly, a relatively high level of transferability (55%) of EST-SSRs was found from barley to wheat (Holton et al. 2002). In another study, 368 EST-SSRs derived from five different grass species (barley, maize, rice, sorghum and wheat) were developed and 149 loci integrated into a reference wheat genetic map; 80 of these were subsequently assigned to chromosomes using nullisomic-tetrasomic lines (Yu et al. 2004b).

The purpose of the project reported here was to investigate 300 primer pairs designed from bread wheat ESTs containing at least one SSR and to study their transferability to durum wheat and other related diploid species carrying the A, B and D genomes as well as rye, *Agropyron*, rice and barley. Their chromosomal assignment was assessed and compared to the one obtained by in silico mapping on the rice pseudo-molecule. Homology relationships with barley and rice ESTs and the evolution of SSRs within ESTs are also discussed.

Materials and methods

Primer design and SSR amplification and detection

SSRs were detected among the 46,510 Génoplante EST contigs (version 1 02/2002) following the method of Nicot et al. (2004). A microsatellite was defined as a sequence containing a minimum of three repeats of a motif comprising from one to six nucleotides, with a total length of at least 12 nucleotides. Primers were designed using PRIMER software (version 0.5, Whitehead Institute for Biomedical Research, Cambridge, Mass.) based on the following criteria: (1) primer length ranging from 18 to 22 bp with 20 bp as the optimum; (2) product size ranging from 100 to 400 bp; (3) melting temperature (T_m) between 57°C and 63°C with 60°C as the optimum; (4) a GC% content between 20% and 80%; (5) maximum acceptable primer self-complementarity of five bases; (6) maximum acceptable 3′ end primer self-complementarity of three bases. Primers were selected on the basis of their containing as few as possible repeated sequences and with the 3′-end of the two primers ending with a C or G when possible. Primer sequences were subjected to BLAST analysis against an in-house database to avoid redundancy with those that already exist. On this basis, 300 primer pairs were selected and designated as CFE (primer sequences available on Graingenes: http://wheat.pw.usda.gov/index.shtml). Each forward primer was M13-tailed [M13: 5′-CACGACGTTGTAAAACGAC-3′, synthesis MWG (Germany)]. PCR analyses using the M13 protocol were performed twice as described by Nicot et al. (2004) with an annealing temperature of 60°C for 30 cycles (30 s 94°C, 30 s 60°C, 30 s 72°C) and 56°C for eight cycles. Polymorphisms were visualised using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, Calif.).

Plant material

Eight hexaploid wheat lines corresponding to the parents of five mapping populations were used for polymorphism screening of the microsatellites: W7984 (synthetic wheat; Van Deynze et al. 1995) and cultivars Opata, Courtot, Chinese Spring, Eurêka, Renan, Arche and Récital. Eight other species were used to study the transferability of EST-SSRs: *Triticum durum* cv. Bidi 17, *T. monococcum* accession 68212, *Aegilops speltoides* accession 37, *Ae. tauschii* accession 15, rye cv. Dankovski Nove (*Secale cereale*), barley cv. Plaisant

(*Hordeum vulgare*), *Agropyron elongatum*, and rice cv. IR64 (*Oryza sativa*). A set of 20 nulli-tetrasomic (NT) and three ditelosomic (DT) wheat lines (kindly provided by Dr. Steve Reader, John Innes Centre, UK) was used for chromosomal assignment of the markers. DNA extraction was performed from fresh leaves using a CTAB protocol as previously described (Tixier et al. 1998).

Sequence analysis and in silico mapping

To assign putative functions to ESTs containing microsatellites, we compared the sequences of the ESTs to the SwissProt and TrEMBL protein databases using BLASTX algorithms (Altschul et al. 1990), with expected value of $1\ e^{-5}$ as a significant homology threshold. Putative functions were attributed according to the definitions given at http://www.godatabase.org/cgi-bin/amigo/go.cgi. TBLASTX searches were also performed against rice and barley NCBI unigene sets (National Center for Biotechnology Information, Washington, D.C.; http://www.ncbi.nlm.nih.gov/) to study the degree of conservation of the repeated motif between the three species. In addition, the EST sequences were compared to the rice pseudo-molecules (http://www.tigr.org/tdb/e2k1/osa1/) using TBLASTX and BLASTN to identify putative orthologues on rice chromosomes. *E*-Values of less than $1\ e^{-5}$, $1\ e^{-10}$, $1\ e^{-25}$, $1\ e^{-50}$ and $1\ e^{-100}$ were selected. Potential bias in the distribution of wheat EST orthologues on rice pseudomolecules was investigated using the relative error (E_r). This value was computed as follows: the rice gene proportion for each chromosome was calculated as the ratio between the number of genes on the rice chromosome and the total number of genes on the rice genome (values at http://www.tigr.org/tdb/e2k1/osa1/). The expected number of hits was then evaluated as the product of the number of wheat EST blasted (251) by the rice gene proportion. Observed hit values were compared to the expected values using a classical binomial test from SPLUS, and only significant values at the 0.05 or 0.01 threshold were retained.

Results

Amplification of EST-SSRs in wheat

The 300 primer pairs were selected among a collection of 46,510 wheat ESTs and were tested on genomic DNA of eight reference hexaploid wheat cultivars in order to assess amplification and polymorphism; 80% (240/300) of these led to an amplification product. In some cases, differences were observed between the expected and the observed size of the amplification product. Among the 240 primer pairs, 51 (21.3%) gave an amplification product either larger than expected, suggesting the presence of an intron in the genomic sequence, or smaller, suggesting the occurrence of deletions within the genomic sequences or a lack of specificity which may lead to amplification of another copy of a gene family.

In general, the EST-SSR markers produced high-quality patterns (Fig. 1). Overall, 198 markers—representing about 80% of total number of primer pairs giving an amplification (240)—produced strong and clear bands; 120 primer pairs gave a single amplification product on wheat, while 120 gave complex patterns with more than one band. Among the latter group, 38 gave three bands likely corresponding to the three homoeologous (A, B, D) copies of the gene. As expected, single-copy EST-SSRs presented a lower level of polymorphism in bread wheat (56%) than multiple-copy (3) EST-SSRs (79%).

Assignment to wheat chromosomes

The 240 primer pairs that yielded products were used for amplification on 20 NT and three DT (to balance the missing NT) lines derived from Chinese Spring. Overall, 177 primer pairs gave rise to PCR products that could be assigned to wheat chromosomes. The remainder (63) of the PCR products could not be assigned because either the profile produced was too indistinct or too complex (42) or because the primer pair amplified a product in all of the NT and DT lines (21). Among the 177 primer pairs, 99

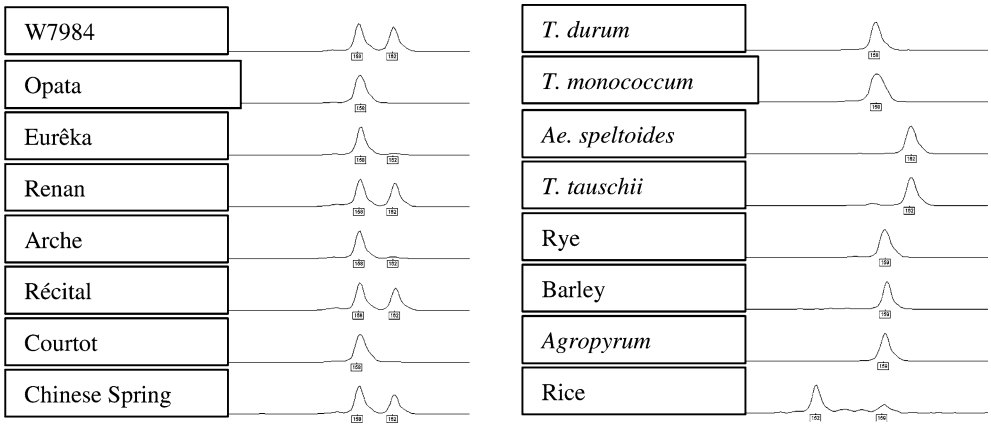

Fig. 1 Example of a high-quality pattern obtained on the ABI 3100 capillary system after M13 PCR amplification of the EST-SSR CFE154 on the eight wheat varieties (*left*) and the eight related species (*right*)

Table 1 Distribution of EST-SSR loci according to their assignment to wheat chromosomes and homoeologous groups

	Homoeologous group							
	1	2	3	4	5	6	7	Total
Chromosome A	14	9	16	17	16	17	12	101
Chromosome B	16	15	15	10	13	14	14	97
Chromosome D	8	12	19	15	15	11	16	96
Total	38	36	50	42	44	42	42	294

were assigned to a single locus, while 78 others were assigned to a maximum of four loci. Most of these latter (70) were found at loci located on the same homoeologous group. Thus, 294 EST-SSR loci distributed across all 21 wheat chromosomes (Table 1) were assigned, with 101, 97 and 96 loci on the A, B and D genomes, respectively. No significant bias in the chromosome or in the genome location was observed, and the distribution ranged from 8 (on chromosome 1D) to 19 loci (3D).

Transferability of wheat EST-SSRs to closely related species

The 300 primer pairs were used on genomic DNA of *T. durum*, *T. monococcum*, *Ae. speltoides* and *Ae. tauschii* accessions. As expected, a high transferability to *T. durum* (AABB; 90.4%) was found. Even the EST-SSRs loci assigned to the D genome showed a good transferability (93.7%). Transferability to the three diploid species carrying the A (*T. monococcum*), B (*Ae. speltoides*) and D (*Ae. tauschii*) genomes was also very good (85.3%, 79.2% and 76.7%, respectively).

The relationships between the copy number in wheat, the degree of polymorphism and the transferability to diploid genomes were also investigated (Table 2). Among the 240 primer pairs giving an amplification product on wheat, 98 were selected and classified into two sets based on the number of bands revealed and their chromosomal location: one set consisted of primer pairs giving a single amplification product (69) that corresponded to one locus, while the other set produced three amplification products (29), which corresponded to the three homoeologous copies. All multiple-copy as well as single-copy markers gave an amplification product on either *T. durum* or on wheat diploid species carrying the A, B and D genomes, indicating a high level of gene-sequence similarity between these species and wheat and, thus, a high level of transferability.

The relationship between the genomic location of the EST-SSR loci on the A, B or D genomes of bread wheat and the transferability to the three diploid ancestral genomes was assessed. Most of the 101 EST-SSR loci assigned to the A genome of bread wheat amplified the very least on *T. monococcum* (97%, Table 3). Only two primer pairs were specific for the B and D diploid genomes and one was specific for the A hexaploid genome. More than 70% amplified on all three diploid genomes, while about 9% amplified on the A and B genomes and 4% on the A and D genomes. The remaining 11% were specific for the A diploid genome. All except one of the 96 EST-SSR loci assigned to the D genome of hexaploid wheat amplified on *Ae. tauschii* (99%); most (85.4%) amplified on the three diploid genomes, while only 5.2% amplified on both the D and A genomes and 6.3% on both the D and B genomes. Only 2.1% remained specific to the D diploid genome. On the contrary, 84% of the 97 loci assigned to the B genome of hexaploid wheat amplified on *Ae. speltoides*, and only 1% were specific to the B diploid genome; 9.3% amplified on both the B and A genomes, 1% on the B and D genomes, while 73.2% amplified on all three genomes. Interestingly, 5% were specific for the hexaploid genome, while 6% amplified on both the A and D diploids and 4% on the A genome species only. Thus, our results show a better conservation between the A and D genomes of bread wheat and the corresponding genomes of the diploid species than between the B genome of bread wheat and the B genome of *Ae. speltoides*, the latter supposedly being the closest species to the ancestral donor.

Transferability of wheat EST-SSRs to other grass species

We used the same set of 98 EST-SSRs giving a single amplification product (69) that corresponded to one

Table 2 Relationships[a] between the number of bands revealed by EST-SSRs and the results of amplification and polymorphism

Number of bands	Polymorphism in common wheat		Transferability to *Triticum monococcum*, *Aegilops speltoides*, *Ae. tauschii*, *T. durum*		Transferability to rye, barley, *Agropyrum* or rice	
1 band	+	39	+	39	+	29
					−	10
	−	30	+	30	+	23
					−	7
3 bands	+	23	+	23	+	23
	−	6	+	6	+	6

[a] +, Indicates either polymorphism was observed on at least one of the eight wheat varieties studied or amplification on at least one related species; −, indicates either the absence of polymorphism on the eight wheat varieties or the absence of amplification on all related species

Table 3 Transferability of EST-SSRs to the A, B and D diploid species as a function of their chromosomal location

Hexaploid genome[a]	ABD	AB	AD	BD	A	B	D	NA	Total
A genome	74 (73.3%)	9 (8.9%)	4 (4%)	2	11 (10.9%)	–	–	1	101
B genome	71 (73.2%)	9 (9.3%)	6	1 (1%)	4	1 (1%)	–	5	97
D genome	82 (85.4%)	1	5 (5.2%)	6 (6.3%)	–	–	2 (2.1%)	–	96

[a]ABD, Amplification in all the three diploid species; AB, AD, BD, amplification in A and B, A and D, and B and D diploid species, respectively; A, B, D, amplification, respectively, in the A, B and D diploid species only; –, no amplification

locus or to three amplification products (29) that corresponded to the three homoeologous copies described previously to study the transferability to related grass species. Overall, 67.9%, 50.4% and 55.8% of the markers gave an amplification product in rye, barley and *Agropyrum*, respectively. The transferability to rice was lower, accounting for 28.3%. About 74% of the single-copy SSRs (Table 2) showed an amplification on at least one of the related species tested (rye, barley, *Agropyrum* or rice). Among these, those located on the A and D genomes were the most transferable (81.5% and 81%, respectively), while those from the B genome exhibited a lower level of transferability (61.9%). All of the multiple-copy SSRs gave an amplification product on at least one of the species. This confirmed that the primers producing numerous bands are less specific than those giving single amplification products and, consequently, are more transferable.

In order to study the relationships between genomic location and transferability between wheat and rice, we investigated on which rice chromosomal regions (telomeric or centromeric) homologues of the wheat EST-SSRs were located. The 265 EST-SSR sequences were BLASTed against the 12 rice pseudomolecules with threshold *E*-values of 1 e^{-5}, 1 e^{-10}, 1 e^{-25}, 1 e^{-50} and 1 e^{-100} (http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules). The results are illustrated in Fig. 2 and Table 4. They show that the rice homologues were not uniformly distributed on the whole rice genome (Table 4), despite the fact that the wheat EST-SSRs were randomly chosen. Rice chromosomes 2 and 3 were significantly overrepresented (E_r = 40.6%, and 50.3%; significant at $P = 0.05$ and 0.01, respectively), while chromosome 11 was significantly underrepresented ($E_r = 65.9\%$; significant at $P = 0.01$). In addition, the wheat ESTs corresponding to those located on rice chromosome 3 showed a better similarity, indicating a higher level of conservation between rice and wheat genes in this region. Interestingly, the wheat EST-SSRs which were transferable to rice were mainly located in telomeric regions (Fig. 2) where the highest sequence similarity was found, indicating a clear relationship between the degree of sequence conservation and the level of transferability. A significant bias at $P = 0.05$ was observed on rice chromosome 5 (Table 4), indicating that the distal part of the long arm of this chromosome was best conserved between wheat and rice at the level of the flanking sequences of the SSRs.

Relationship between the putative function of the ESTs and the level of transferability to rice

The 265 ESTs containing a SSR were BLASTed against the SwissProt and TrEMBL protein databases in order to identify their putative functions. Using an *E*-value of 1 e^{-5} as the threshold, 168 (63.4%) EST sequences had one hit with one sequence from SwissProt, while 228 (86%) had one hit with the TrEMBL database. We classified the 168 EST sequences that gave a hit on the SwissProt database according to the criteria given on the gene ontology database (AmiGO, http://www.godatabase.org/cgi-bin/amigo/go.cgi). Three different classes of function are proposed on this basis: (1) biological processes, which means phenomena marked by changes that lead to a particular result; (2) cellular components, including gene products that are parts of macromolecular complexes; (3) molecular function, which means elemental activities, such as catalysis or binding. Most of the ESTs (107, 63.7%) were involved in a molecular function, while 23 (13.7%) played a role in biological processes and 19 (11.3%) were cellular components (Fig. 3). The 19 remaining ESTs had no putative function. With respect to those with a molecular function, almost 50% of them (53) were involved in catalytic activity, about 15% in binding activity, and some in transporter (8) or chaperone activity (8). This distribution corresponds to the one observed on the whole EST library (about 170,000 sequences, data not shown), indicating a random distribution of the EST-SSRs among the different classes of genes.

To study further the relationship between transferability to rice and EST function, we selected the 69 EST-SSRs giving an amplification product on rice and examined their putative function. Among the 44 showing a hit to the SwissProt database, similar proportions were observed between the three classes of function, indicating no significant bias between gene function and the level of transferability to rice.

With respect to gene location, 123 of the 168 EST-SSRs sequences found in SwissProt occurred inside open reading frames (ORFs), while only 45 were in non-coding regions (41 in 3′-untranslated regions (UTRs) and 4 in 5′-UTRs), suggesting that SSRs occurred more frequently in ORFs than in untranscribed regions. The proportions of di-, tetra- and pentanucleotides that cannot change without mutational consequences inside the ORFs were not different from those observed for the whole library (Nicot et al. 2004).

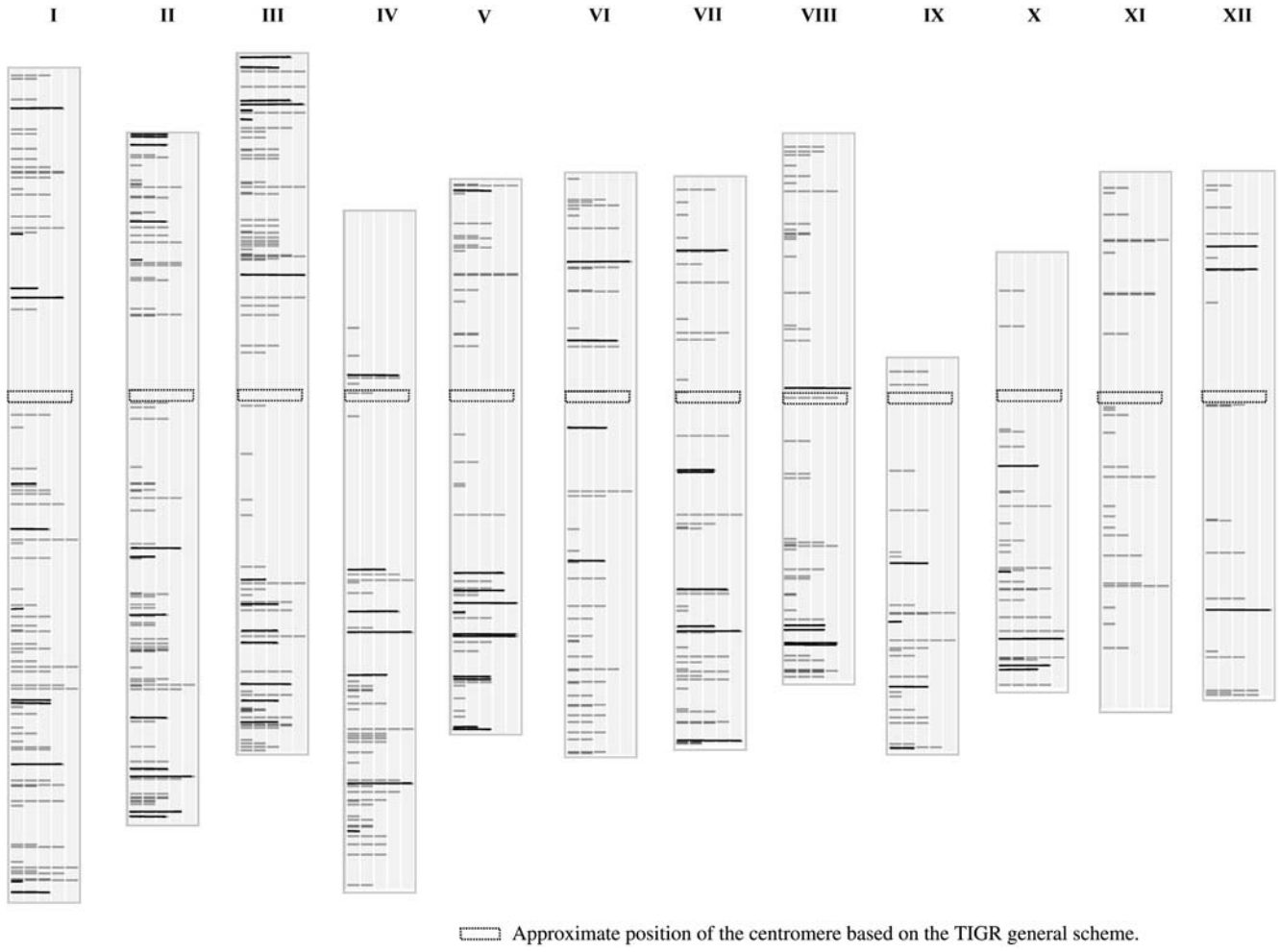

Fig. 2 Results of the BLAST analysis of the wheat EST-SSRs on the 12 rice pseudo-molecules. Five different expected values were selected: 1 e^{-5}, 1 e^{-10}, 1 e^{-25}, 1 e^{-50} and 1 e^{-100} . Approximate position of the hit for each locus is indicated with *solid bars* according to an e-value for one bar of 10^{-5} up to an e-value for five bars of 10^{-100} . The EST-SSRs that were transferable to rice are indicated by *black bars* and all others are indicated by *grey bars*. Approximate position of the centromere is indicated with a *dotted rectangle*

Conservation of the SSR motif between wheat, barley and rice EST-SSRs

Wheat EST sequences were BLASTed against rice and barley EST unigene sets (http://www.ncbi.nlm.nih.gov/) in order to determine if a repeated motif existed within homologous rice and barley sequences. Of the 265 EST

Table 4 Distribution of 251 wheat EST loci on the 12 rice pseudomolecules

	Rice pseudomolecules												
	1	2	3	4	5	6	7	8	9	10	11	12	Total
Chromosome length (%)	11.8	9.7	9.9	9.6	7.9	8.2	8.1	7.8	5.8	6.2	7.6	7.4	100
Expected gene proportion[a] (%)	12.9	10.2	11.4	8.9	7.9	8.2	8.1	7.3	5.4	6.2	7.0	6.5	100
Expected no. of genes[b]	32.4	25.6	28.6	22.3	19.8	20.6	20.3	18.3	13.6	15.6	17.6	16.3	251
Observed no. of genes[c]	36	36	43	21	20	21	19	15	12	12	6	10	251
E_r[d] (%)	11.1	40.6*	50.3**	5.8	1.0	1.9	6.4	18.0	11.8	23.1	65.9**	38.7	
Expected Transferability[e]	14.1	14.1	16.8	8.2	7.8	8.2	7.4	5.8	4.7	4.7	2.3	3.9	98
Observed Transferability[f]	14	14	15	9	13*	5	8	6	4	6	1	3	98

*, **Significant at the 0.05 and 0.01 levels, respectively

[a]Expected gene proportion, Number of genes on the rice chromosome/total Nb of genes on the rice genome

[b]Expected number of genes, number of rice homologue loci with a hit with wheat EST (251) × expected gene proportion

[c]Observed number of genes, Number of hits observed on each rice pseudomolecule

[d]Er (relative error) = expected value − observed value / expected value × 100 (%)

[e]Expected transferability, Expected transferability between wheat and rice = [observed no. of genes × no. of transferable homologue loci (98)]/no. of rice homologue loci with a hit with wheat EST (251)

[f]Observed transferability, Observed transferability between wheat and rice

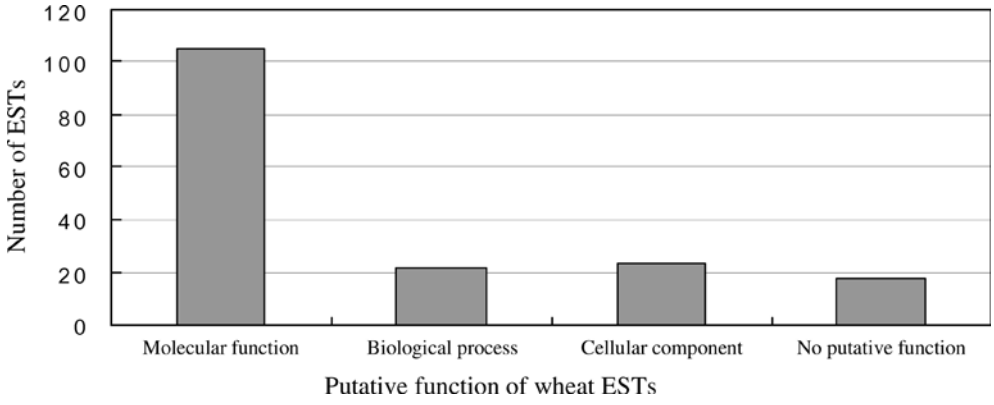

Fig. 3 Putative function annotation of ESTs containing microsatellites according to the criteria given on the gene onthology database (AmiGO, http://www.godatabase.org/cgi-bin/amigo/go.cgi)

sequences, 236 had at least one hit to rice ESTs, and 228 had a hit to barley ESTs using an *E*-value of 1 e^{-5}. Fifty-four ESTs showing a hit to both barley and rice EST databases and having large SSR-flanking regions were studied in detail (Table 5). Among these, 27 amplified in wheat, barley and rice simultaneously, 11 in both wheat and barley, three in both wheat and rice and 13 in wheat only. For those amplifying in the three species, most of the sequences (24/27 for barley, 20/27 for rice) presented similar repeated motifs (perfect or non-perfect). An example is given in Fig. 4a. For EST-SSR CFE1, a trinucleotide motif $(cgc)_4$ was observed in wheat. The same motif was also present in barley and rice, but it was repeated five times in the former and two times in the latter. Repeats were not present in three sequences in barley and seven in rice, with two of the barley sequences being common with the seven rice sequences. For the EST-SSRs amplifying in wheat and barley only, 9 of the 11 presented a similar repeated motif in barley, while no repeats were present for the two others. For those amplifying in wheat and rice only, a repeated motif was always detected in the rice sequences. For the EST-SSR CFE59 (Fig. 4b), the motif was different in wheat and barley (ccgt) from rice (cgc),

Table 5 Degree of conservation of the SSR motifs in rice and barley EST sequences

Amplification in	Number of markers	Repeated motif in barley				Repeated motif in rice			
		Existence			No existence	Existence			No existence
		Perfect	Non-perfect	Different		Perfect	Non-perfect	Different	
Wheat, barley and rice	27	15	9	0	3	9	11	0	7
Wheat and barley	11	5	4	0	2	2	3	0	6
Wheat and rice	3	1	0	1	1	0	2	1	0
Wheat	13	4	2	0	7	1	1	0	11
Total	54	24	15	1	13	12	17	1	24

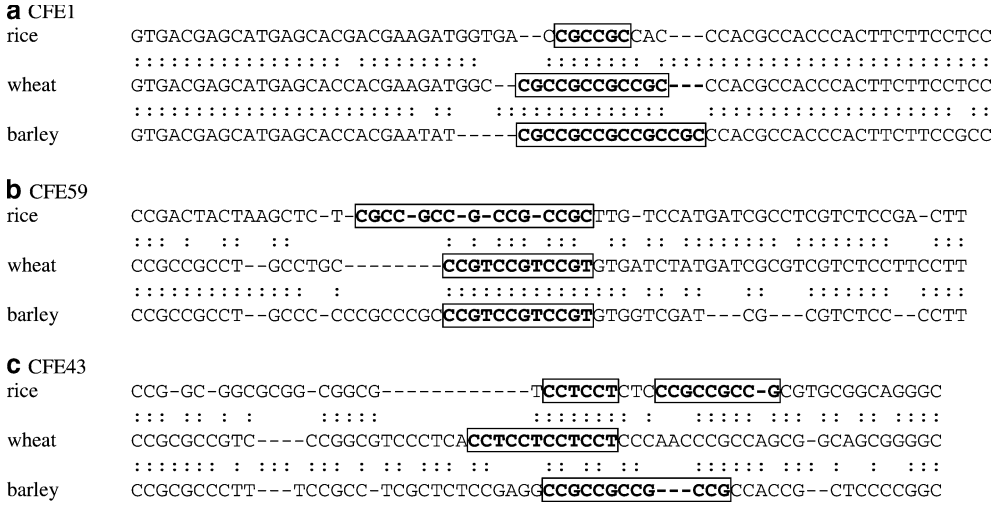

Fig. 4 Sequence comparison between wheat, barley and rice EST sequences for the EST-SSRs CFE1 (**a**), CFE59 (**b**) and CFE43 (**c**). In each case, the SSR is indicated by *boxed bold characters*. Identical nucleotides are related with a *double dot*

suggesting a common origin for the two triticeae species and a different one for rice. For EST-SSR CFE43 (Fig. 4c), the motif was composite in rice [$(cct)_2(ccg)_3$], while it was unique but different in wheat [$(cct)_4$] and barley [$(ccg)_4$], suggesting that a motif similar to the one present in rice was present in the common ancestor and then evolved differently after the divergence of the three species. For the EST-SSRs amplifying in wheat only, repeats did not exist in the corresponding region for most of those for rice (11/13) and, to a lower extent, for barley (7/13). In some cases, the repeated motifs did exist in one or in both barley and rice species, but no amplification occurred. This is likely due to mispairing at the primer level because of the presence of mutations in the flanking regions. One possible explanation for the absence of repeated motifs is that the SSRs arose after the divergence between wheat and rice or wheat and barley. However, we cannot exclude the possibility that we are not comparing orthologues but paralogues of the genes that followed different evolutionary histories.

Discussion

In our study, 80% of the primer pairs successfully amplified EST-SSR products, and over 80% of these produced strong and clear profiles in wheat. About 64% of the primer pairs yielded fragments of the expected size; this is in sharp contrast to results obtained with genomic SSRs, for which only 36% yielded fragments of the expected size, with many resulting in a smear (Röder et al. 1995). This results confirms that EST-SSRs give better profiles than genomic SSRs (Eujayl et al. 2001, 2002; Leigh et al. 2003).

With respect to chromosomal assignment, 177 EST-SSRs revealed 291 loci that were randomly distributed along all the 21 wheat chromosomes, indicating no prevalence for the location of EST-SSRs. Up to 45% of the EST-SSRs identified more than one locus, suggesting an amplification of either the homoeologous or homologous copies. Some bands were not assigned because the products were amplified in all NT and DT lines, indicating that these EST-SSRs produced co-migrating bands of the same size for more than one copy of the gene. If necessary, such markers can be converted in locus-specific markers by cloning and sequencing each amplification product, identifying specific mutations by comparing the sequences of the different gene copies and designing new pairs of primers specific for each copy. Such specific EST-SSRs would represent excellent markers for phylogenetic studies.

EST-SSR markers are more transferable across closely related genera than genomic SSRs because they originate from the coding regions which have a higher level of sequence conservation than intergenic regions. Our results confirmed that wheat EST-SSRs show high simultaneous transferability (86.6%) to at least two of the A-, B- and D-related diploid genomes, while fewer than 15% were only specifically transferable to one genome. These results contrast with those observed with genomic SSRs which are more genome-specific and thus less transferable to related species (Sourdille et al. 2001). Similar results were reported in soybean where genomic SSRs are mostly restricted to congenic species (Peakall et al. 1998). In addition, our results show that EST-SSRs in wheat may be used to detect the three homoeologous copies at the same time, as confirmed by the chromosomal assignment of EST-SSR loci using the set of Chinese Spring compensating nullisomic-tetrasomic lines (Sears 1966).

The high transferability to rye (67.9%) that we observed contrasted with the low transferability of wheat genomic SSRs to rye reported by Kuleung et al. (17%; 2003) and by Röder et al. (6.75%; 1995). Transferability to barley (55.8%) and rice (28.3%) was also relatively good under our conditions. Our results with barley were in total agreement with those of Yu et al. (2004a), who reported 53% of transferability between these two species. However, the same authors mentioned a higher level of transferability to rice (45%). This was due to the selection of primers that was done in well-conserved regions between wheat and rice, while our primers were designed on the wheat EST sequences only. Our results can thus be improved by using homology information from rice and barley databases, and they largely confirmed the higher power of EST-SSRs compared to genomic SSRs for transferability studies (Eujayl et al. 2003; Gupta et al. 2003; Thiel et al. 2003).

The analysis of the relationships between the level of transferability and the number of bands revealed that EST-SSRs showing three bands had a higher transferability than those with only one band, even to more distantly related species such as barley, rye and rice. This is in accordance with the results reported by Guyomarc'h et al (2002a), who studied genomic SSRs developed from *Ae. tauschii*. The single-band markers may amplify more conserved regions, which would explain the fact that most of them yielded one product compared to multiple-banded markers. This suggests that EST-SSR markers giving three bands are more useful for comparative mapping.

Recent comparisons of cytogenetic maps with genetic linkage maps based on chromosome deletion lines have revealed the existence of gene-rich regions in the wheat genome, most of which are located in the distal parts of the chromosomes (Werner et al. 1992; Gill et al 1996; Faris et al. 2000; Weng et al. 2000; Sandhu et al. 2001, Sourdille et al. 2004). BLASTing the 265 wheat ESTs against rice's pseudo-molecules showed that despite the fact that EST-SSRs were randomly chosen, the distribution was uneven. Chromosomes 2 and 3 from rice were overrepresented while chromosome 11 was underrepresented. The results also showed that our set of wheat ESTs had a better sequence similarity with genes on rice chromosome 3 compared to the others. This suggests that in the course of evolution, a relatively higher level of gene conservation was maintained between rice chromosome 3 and the corresponding

homoeologous group 4 in wheat (La Rota and Sorrells 2004; Yu et al. 2004b). This may partly explain the lack of polymorphism that is frequently observed on homoeologous group 4, especially on chromosome 4D (Nelson et al. 1995; Cadalen et al. 1997; Paillard et al. 2003). EST-SSRs located on wheat group 4 chromosomes would be more effective for comparative mapping studies in the future.

In silico analysis demonstrated that 73.2% of the EST-SSRs were located in ORFs in wheat, which was similar to observations made on other plants (Gupta et al. 2003). Many studies have shown that SSRs are prevalently found in 5′-UTR (Cho et al. 2000; Scott et al. 2000; Gupta et al. 2003; Thiel et al. 2003; Yu et al. 2004a). In contrast, our results showed that SSRs were more often located in 3′-UTRs than in 5′-UTRs. This is due to a bias in our sampling of EST-SSRs, which mainly originate from sequences of 3′-end cDNAs (Génoplante collection) and also from the fact that 5′-UTRs are less frequent in the databases since they are more difficult to obtain. We were not able to correlate the presence of a SSR in a coding sequence to a specific cellular function. SSRs seem to be randomly distributed among gene classes since the EST-SSRs' function distribution was consistent with the distribution of gene function in the EST libraries.

We further investigated 54 wheat ESTs which were found in both barley and rice databases and were long enough for significant BLAST analysis. A distinct correlation was found between the rate of success of amplification in all three species (wheat, barley and rice) and the degree of conservation between the repeated motifs (89% and 74% for barley and rice, respectively). However, the PCR yield was sometimes weaker than in wheat, suggesting that flanking sequences were subjected to mutations at one or several bases. More wheat repeated motifs remained identical in barley than in rice. This was expected since wheat ESTs show a higher degree of homology with barley ESTs than with rice ESTs as the former diverged from each other more recently during evolution (10 mya between wheat and barley compared to 50 mya between wheat and rice). Most of the conserved motifs between the three species were trinucleotides (77%), which is in agreement with the size distribution of the motif among ESTs (Nicot et al. 2004). Interestingly, one-half of the conserved motifs were (ccg), and all of them contained at least one g (or one c). When the SSR motifs were identical between the three species, no significant difference was observed with respect to the number of repeats even if rice had frequently fewer repeats, suggesting that the evolution rate of the EST-SSRs could be slower in rice than in wheat and barley.

In some cases, repeated motifs exist at a proto-microsatellite state in rice; this also occurs on occasion in barley. In these cases, our results suggest that the wheat microsatellites evolved following the divergence between the Bambusoideae and the Pooideae, or even later. For some of the motifs that are proto-microsatellites in rice but appeared as microsatellites in barley, the differentiation can be refined at the level of the Triticodeae. Several types of mutation may have transformed proto-microsatellites into microsatellites: point mutations, frameshift mutations or the transposition of mobile elements. The frequency of such events is not different between the three species. If the proto-microsatellites have not evolved in rice, this may be due by chance or be associated with selection pressure. Since wheat is a polyploid which bears most of the genes in triplicate, the constraint is probably lower and proto-microsatellites can evolve faster than in rice or barley.

Finally, two hypotheses can be raised for those motifs that are absent from rice and barley. The first one is that they could have arisen in the *Triticum* lineage only. Such a hypothesis can be confirmed by testing a large number of additional grass species in order to identify the divergence points for these EST-SSRs. This could be a valuable contribution to the reorganisation of the phylogeny of the grass species. We would also be able to date some of the differentiations using strategies such as the one described in Thuillet et al. (2002), who estimated the mutation rate of ten SSR loci in durum wheat. Such a programme is currently being pursued in our laboratory. The second hypothesis is that the wheat and rice ESTs that are compared are not true orthologues. Some of the ESTs we used had several hits on the rice pseudomolecules, and it is possible that only one or few of them have the SSR, whereas the one expressed and found in the databases does not carry the microsatellite.

Thus, we can conclude that wheat EST-SSR markers show a high transferability across a large range of species. This transferability make them a powerful tool to work on orphan wild species such as *Agropyrum* where less effort has been expended to develop genomic resources such as molecular markers. These species are a very important source of both abiotic and biotic resistance genes (Doussinault et al. 1983), and molecular markers are precious tools to use and reduce the introgressions of genes (location, size of the introgression) from these species. EST-SSRs are thus excellent molecular markers that can now be used extensively in breeding programmes by MAS methods.

Acknowledgements The authors thank gratefully G. Gay and A. Loussert for growing the plants and S. Reader for providing the aneuploid lines. Gilles Charmet is also acknowledged for the statistical analyses. This work was funded by the China Scholarship Council (CSC). All experiments comply with the current laws of France.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Areshchenkova T, Ganal MW (2002) Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. Theor Appl Genet 104:229–235

Cadalen T, Boeuf C, Bernard S, Bernard M (1997) An intervarietal molecular marker map in *Triticum aestivum* Em. Thel and comparison with a map from cross. Theor Appl Genet 94:367–377

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

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci 160:1115–1123

Doussinault G, Delibes A, Sanchez-Monge R, Garcia-Olmedo F (1983) Transfer of a dominant gene for resistance to eyespot disease from a wild grass to hexaploid wheat. Nature 303:698–700

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2001) Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Euphytica 119:39–43

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

Eujayl I, Sledge MK, Wand L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR (2003) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–422

Faris JD, Hean KM, Gill BS (2000) Saturation mapping of a gene-rich recombination hot spot region in wheat. Genetics 154:823–835

Gao LF, Jing RL, Huo NX, Li Y, Li PX, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392–1400

Gill KS, Gill BS, Endo TR, Boyko EV (1996) Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. Genetics 143:1001–1012

Grivet L, Glaszmann JC, Vincentz M, Silva F, Arruda P (2003) ESTs as a source for sequence polymorphism discovery in sugarcane: example of the *Adh* genes. Theor Appl Genet 106:190–197

Gupta PK, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomics 4:139–162

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genomics 270:315–323

Guyomarc'h H, Sourdille P, Charmet G, Edwards KJ, Bernard M (2002a) Characterisation of polymorphic microsatellite markers from *Aegilops tauchii* and transferability to the D-genome of bread wheat. Theor Appl Genet 104:1164–1172

Guyomarc'h H, Sourdille P, Edwards KJ, Bernard M (2002b) Studies of the transferability of microsatellites derived from *Triticum tauschii* to hexaploid wheat and to diploid related species using amplification, hybridization and sequence comparisons. Theor Appl Genet 105:736–744

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

Jong E, van Z, Guthridge KM, Spangenberg GC, Forster JW (2002) Development and characterization of EST-derived simple sequence repeat (SSR) markers for pasture grass endophytes. Genome 46:277–290

Kuleung C, Baenziger PS, Dweikat I (2003) Transferability of SSR markers among wheat, rye and triticale. Theor Appl Genet 108:1147–1150

La Rota M, Sorrells ME (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. Funct Integr Genomics 4:34–46

Leigh F, Lea V, Law J, Wolters P, Powell W, Donini P (2003) Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. Euphytica 133:359–366

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet 109:361–369

Liu ZW, Biyashev RM, Maroof MAS (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor Appl Genet 93:869–876

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T, Blair M (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol 35:89–99

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10:72–80

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Nelson JC, Sorrells ME, Van Deynze AE, Lu YH, Atkinson M (1995) Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. Genetics 141:721–731

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy F, Bernard M, Sourdille P (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109:800–805

Paillard S, Schnurbusch T, Winzeler M, Messmer M, Sourdille P, Abderhalden O, Keller B, Schachermayr G (2003) An integrative genetic linkage map of winter wheat (*Triticum aestivum*). Theor Appl Genet 107:1235–1242

Peakall R, Gilmore S, Keys W, Morgants M, Rafalski A (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Mol Biol Evol 15:1275–1287

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 7:215–222

Prasad M, Varshney RK, Roy JK, Balyan HS, Gupta PK (2000) The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. Theor Appl Genet 100:584–592

Rafalski S, Tingey A (1993) Genetics diagnostics in plant breeding: RAPDs, microsatellite and machines. Trends Genet 9:275–280

Ramsay L, Macaulay M, degli Ivannissevich S, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, Maestri A, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997–2005

Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246:327–333

Röder MS, Korsun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998) A microsatellite map of the wheat genome. Genetics 149:2007–2023

Sandhu D, Champoux JA, Bondareva SV, Gill KS (2001) Identification and physical location of useful genes and markers to a major gene-rich region on wheat group 1S chromosomes. Genetics 157:1738–1747

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Sears ER (1966) Nullisomic-tetrasomic combinations in hexaploid wheat. In: Riley KR, Lewis R (eds) Chromosome manipulations and plant genetics. Oliver and Boyd, Edinburgh, pp 29–45

Senior ML, Chin ECL, Lee M, Smith JSC (1996) Simple sequence repeat markers developed from maize found in the GenBank database: map construction. Crop Sci 36:1676–1683

Sourdille P, Tavaud M, Charmet G, Bernard M (2001) Transferability of wheat microsatellites to diploid Triticeae species carrying the A, B and D genomes. Theor Appl Genet 103:346–352

Sourdille P, Singh S, Cadalen T, Brown-Guedira GL, Gay G, Qi L, Gill BS, Dufour P, Murigneux A, Bernard M (2004) Microsatellite-based deletion mapping system for the establishment of genetic map—physical map relationships in wheat. Funct Integr Genomics 4:12–25

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*). Theor Appl Genet 106:411–422

Thuillet AC, Bru D, David J, Roumet P, Santoni S, Sourdille P, Bataillon T (2002) Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp *durum* desf. Mol Biol Evol 19:122–125

Tixier MH, Sourdille P, Charmet G, Gay G, Jaby C, Cadalen T, Bernard S, Nicolas P, Bernard M (1998) Detection of QTLs for crossability in wheat using a doubled-haploid population. Theor Appl Genet 97:1076–1082

Van Deynze AE, Dubcovsky J, Gill KS, Nelson JC, Sorrels ME, Dvorak J, Gill BS, Lagudah ES, McCouch SR, 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 RK, Thiel T, Stein N, Langridge P, Grande A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546

Weng Y, Tulee NA, Hart GE (2000) Extended physical maps and a consensus physical map of the homoeologous group-6 chromosomes of wheat (*Triticum aestivum* L-em Thel). Theor Appl Genet 100:519–527

Werner JE, Endo TR, Gill BS (1992) Toward a Cytogenetically Based Physical Map of the Wheat Genome. Proc Natl Acad Sci USA 89:11307–11311

Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Cardle L, Ramsay L, Gibby M, Powell W (2003) Development of EST-SSRs from the Alpine Lady-fern, *Athyrium distentifolium*. Molecular Ecology Notes 3:287–290

Yu JK, La Rota M, Kantety RV, Sorrells ME (2004a) EST derived SSR markers for comparative mapping in wheat and rice. Mol Genet Genomics 271:742–751

Yu JK, Dake TM, Singh S, Benscher D, Li WL, Gill B, Sorrells ME (2004b) Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47:805–818