Theor Appl Genet (2006) 113:1271–1281
DOI 10.1007/s00122-006-0381-4

ORIGINAL PAPER

Development and genetic mapping of sequence-tagged microsatellites (STMs) in bread wheat (*Triticum aestivum* L.)

M. J. Hayden · P. Stephenson · A. M. Logojan · D. Khatkar · C. Rogers · J. Elsden · R. M. D. Koebner · J. W. Snape · P. J. Sharp

Received: 11 January 2006 / Accepted: 25 July 2006 / Published online: 25 August 2006

Abstract The density of SSRs on the published genetic map of bread wheat (*Triticum aestivum* L.) has steadily increased over the last few years. This has improved the efficiency of marker-assisted breeding and certain types of genetic research by providing more choice in the quality of SSRs and a greater chance of finding polymorphic markers in any cross for a chromosomal region of interest. Increased SSR density on the published wheat genetic map will further enhance breeding and research efforts. Here, sequence-tagged microsatellite profiling (STMP) is demonstrated as a rapid technique for the economical development of anonymous genomic SSRs to increase marker density on the wheat genetic map. A total of 684 polymorphic sequence-tagged microsatellites (STMs) were developed, and 380 were genetically mapped in three mapping populations, with 296 being mapped in the International Triticeae Mapping Initiative W7984 × Opata85 recombinant inbred cross. Across the three populations, a total of 479 STM loci were mapped. Several technological advantages of STMs over conventional SSRs were also observed. These include reduced marker deployment costs for fluorescent-based SSR analysis, and increased genotyping throughput by more efficient electrophoretic separation of STMs and a high amenability to multiplex PCR.

Communicated by M. Morgante

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-006-0381-4 and is accessible for authorized users.

M. J. Hayden (✉) · J. Elsden
Plant Functional Genomics Center, University of Adelaide, Hartley Grove, Urrbrae, SA 5064, Australia
e-mail: matthew.hayden@adelaide.edu.au

M. J. Hayden
Molecular Plant Breeding CRC, PMB 1, Glen Osmond, SA 5064, Australia

P. Stephenson · A. M. Logojan · C. Rogers · R. M. D. Koebner · J. W. Snape
Crop Genetics Department, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

D. Khatkar · P. J. Sharp
Plant Breeding Institute, University of Sydney, PMB11, Camden, NSW 2570, Australia

P. J. Sharp
Value Added Wheat CRC, LB1345, North Ryde, NSW 2113, Australia

Introduction

Bread wheat (*Triticum aestivum* L em. Thell.) is an allohexaploid ($2n = 6x = 42$) composed of three distinct but related genomes (ABD), each with seven chromosomes. It has an extremely large genome of 16×10^9 bp per chromosome, of which more than 80% is repetitive DNA (Arumuganathan and Earle 1991). Over the past two decades, detailed molecular genetic maps have been developed for all 21 chromosomes (Nelson et al. 1995a, b, c; Van Deynze et al. 1995; Marino et al. 1996; Roder et al. 1998; Somers et al. 2004; Song et al. 2005). These maps have proved particularly useful for a broad range of applications including the discovery of marker–trait associations, positional cloning of important genes, characterization

of germplasm, and evolutionary studies (Jian et al. 2002). An important use has also been for marker-assisted selection in early generations of breeding programs (Koebner et al. 2001). However, the usefulness of molecular markers for different purposes is largely determined by their type. RFLPs are labor-intensive, and show a low level of polymorphism in wheat (Paull et al. 1998; Kim and Ward 2000). Single-nucleotide polymorphisms are proving slow to be developed and implemented in wheat because of lack of good sequence data. In contrast, microsatellites (SSRs) are easy to use and show a relatively high degree of polymorphism (Powell et al. 1996). As a consequence, SSRs are currently the preferred type of molecular marker in wheat.

SSRs are tandem repeats of short (1–6 bp) DNA sequences that are easily assayed by PCR. A high level of polymorphism is usually observed with these markers due to length variation resulting from changes in the number of repeat units (Lagercrantz et al. 1993). SSRs are highly amenable to automation and high-throughput capillary electrophoresis. For marker-assisted breeding, SSRs provide the potential to rapidly screen a large number of plants at low cost. However, the deployment of SSRs in genetic studies and marker-assisted breeding is most effective when genetic maps are available that are densely populated with these markers. This provides greater choice in the quality of markers, and a higher probability of finding polymorphic markers that tag chromosomal regions of interest (Somers et al. 2004).

A limitation to the development of a high-density SSR map in bread wheat has been the considerable cost and effort involved in marker development. Traditionally, this has required DNA sequencing of genomic clones containing microsatellites to enable the design of PCR primers for the amplification of the target sequence. Several strategies improving the efficiency of marker development focus on the enrichment of genomic libraries for SSR-containing clones (Edwards et al. 1996), targeting the development of SSRs to genomic regions of interest (Hayden and Sharp 2001a; Cregan et al. 1998), or mining publicly available EST databases (Varshney et al. 2002). The latter strategy was widely adopted in recent years, as it provides markers for transcribed regions of the genome. EST-SSRs generally exhibit good cross-species transferability, and are particularly useful for comparative genomics studies (Zhang et al. 2005). However, EST-SSRs typically reveal a much lower level of polymorphism compared to markers derived from anonymous genomic sequences (Eujayl et al. 2001; Leigh et al. 2003).

Sequence-tagged microsatellite profiling (STMP) is a technique for the economical development of SSRs from genomic and cDNA (Hayden and Sharp 2001b). This is achieved by substantially reducing the DNA sequencing effort that is required to characterize microsatellite clones, and essentially halving the number of primers that must be synthesized. STMP employs the principles of serial analysis of gene expression (Velculescu et al. 1995) to generate a library of nucleotide sequence tags that correspond to SSR-containing amplicons present within a pool of PCR-amplified restriction fragments. In essence, the technique generates from the DNA sequence flanking one side of each microsatellite repeat array a short (27-bp), but positionally defined, nucleotide sequence tag. Each tag contains sufficient DNA sequence to uniquely identify the corresponding SSR fragment, since a 27-bp sequence can distinguish more than 1.8×10^{16} (4^{27}) fragments. Each nucleotide tag also contains sufficient flanking DNA sequence to allow the design of an oligonucleotide primer [referred to as a sequence-tagged microsatellite (STM) primer] that can be used in anchored PCR to amplify the corresponding SSR from genomic DNA. In anchored PCR, microsatellite amplification is achieved using an STM primer in combination with a primer that anchors to the microsatellite repeat motif itself (Hayden et al. 2004). As the microsatellite anchoring primer is common to all markers targeting the same repeat motif, STMP halves the number of primers required compared to conventional SSR amplification, which is achieved using a pair of specific primers flanking each repeat array assayed.

In the present study, STMP was used to develop 684 polymorphic wheat STMs targeting compound repeat motifs, which consist of two or more repeat motifs in tandem.

Materials and methods

Plant materials and mapping populations

The 12 wheat varieties used to assess STMs for amplification and polymorphism were Opata85, W7984, Chinese Spring, Synthetic (Devos et al. 1992), Rialto, Spark, Halberd, Cranbrook, Sunco, Tasman, Carnamah and WAWHT2046. These varieties represent the parental lines of the mapping populations that were used to genetically map polymorphic STMs, and parents of other important Australian and European crosses. The mapping populations were the International Triticeae Mapping Initiative (ITMI) W7984 × Opata85 recombinant inbred cross, Chinese

Spring × Synthetic F_2 intercross (Devos et al. 1992) and Cranbrook × Halberd doubled haploid population (Kammholz et al. 2001). In all instances, only the first 94 progeny of each cross were used for genetic mapping. Each mapping population was selected because it had been previously used for genetic mapping. The set of 12 Australian wheat varieties used to assess the informativeness and allele size range of STMs and published microsatellite markers were Cadoux, Cascades, Carnamah, CD87, Cranbrook, Halberd, Janz, Katewpa, Kukri, Spear, Sunco and Tasman.

STMP library construction, primer design and marker nomenclature

Sequence-tagged microsatellite libraries were developed for ten compound repeat motifs (Table 1) from a pool of preamplified *Pst*I–*Mse*I restriction fragments prepared from genomic DNA using the STMP technique (Hayden and Sharp 2001b). The STMP procedure is described fully in the Electronic Supplementary Materials. Ninety-six plasmid clones were sequenced from each microsatellite library, and nucleotide sequence tags were extracted from the raw sequence data files using the STMP software (Hayden and Sharp 2001b). STM profiles were generated for each library to determine the relative abundance and frequency of SSRs within the pool of *Pst*I–*Mse*I restriction fragments from which the sequence tags were derived. Nucleotide sequence tags with a low ($n \leq 3$) copy number were assumed to correspond to unique (or low copy) SSRs in the bread wheat genome, and were selected for PCR primer design.

Sequence-tagged microsatellite primers were designed with NetPrimer (Premier Biosoft International) from low-copy nucleotide sequence tags using the parameters: primer length 20–25 bp, 3′-end stability –5.5 to –9.0 kcal/mol, oligomer T_m 55–70°C, GC content 30–70% and primer rating >80. The PCR specificity of each primer toward all tags was tested in silico using Amplify (Engels 2004). Briefly, the number of nucleotide sequence tags with similar, but not identical sequence, to which each STM primer could efficiently hybridize under PCR conditions was determined. STM primers were discarded when they hybridized strongly to more than three nucleotide sequence tags within the microsatellite library from which the tag primer was derived.

Sequence-tagged microsatellites were named using the prefix ‘‘stm,’’ following by a unique number, and a suffix which corresponded to the microsatellite anchoring primer that should be used in combination with the STM primer to amplify the target locus. For example, stm25tgag denotes a marker that is amplified using the primer *stm25tgag* and the microsatellite anchoring primer *tgag* (Table 2).

Uniplex and multiplex STM amplification

Sequence-tagged microsatellites were amplified in 10 µl reaction mixtures containing 0.2 mM dNTP, 1× Platinum Taq PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 0.2 mM each of STM (forward) and appropriate microsatellite anchoring (reverse) primer (Table 2), 50 ng genomic DNA and 0.25 U Platinum Taq DNA polymerase (Invitrogen). Each microsatellite anchoring primer was labeled at its 5′-end with hexachloro-6-carboxylfluorescein. PCR was performed for 47 cycles with the touchdown profile: 60 s at 92°C, 60 s at 62°C and 30 s at 72°C. Following the first cycle, the annealing temperature was reduced by 1°C per cycle for the next seven cycles. The PCR products were

Table 1 Statistics for ten STM profiles developed for compound repeat motifs in bread wheat

Repeat Motif	Total number tags sequenced	Copy number of unique tags				Total number of low-copy tags ($n \leq 3$)	Proportion of duplicated tags[a] (%)	Number of STM primers synthesized
		1	2	3	4			
$(Tg)_m(Ag)_n$	1,538	1,195	106	10	25	1,311	14.8	325
$(Tc)_m(Ac)_n$	1,750	1,015	129	70	6	1,214	30.7	386
$(Tc)_m(Tg)_n$	1,521	1,058	95	33	24	1,186	22.0	357
$(Ac)_m(Ag)_n$	1,620	858	73	62	38	993	38.8	212
$(Ag)_m(Ac)_n$	1,682	607	225	72	53	904	46.3	393
$(Ag)_m(Tg)_n$	1,466	742	40	11	63	793	45.9	327
$(Tg)_m(Tc)_n$	1,270	202	50	17	82	269	78.9	48
$(Ac)_m(Tc)_n$	1,421	122	29	19	41	170	88.0	135
$(Ac)_m(AT)_n$	1,429	86	35	14	42	135	90.5	97
$(Tg)_m(TA)_n$	1,758	78	20	6	35	104	94.1	24
Total	15,455	5,963	802	314	409	7,079		2,304

[a] Calculated as 100% minus percent of low-copy tags as a proportion of total number of tags sequenced

Table 2 Sequences of microsatellite anchoring primers used to construct the ten STM profiles, and STM primers used for six-plex PCR in Fig. 1

Microsatellite anchoring primers (5 → 3′)	
tgag	TGTGTGTGTGTGTGAGAGAGAG
tcac	TCTCTCTCTCTCTCACACACAC
tctg	TCTCTCTCTCTCTCTGTGTGTG
acag	ACACACACACACACAGAGAGAG
agac	AGAGAGAGAGAGAGACACACAC
agtg	AGAGAGAGAGAGAGTGTGTGTG
tgtc	TGTGTGTGTGTGTGTCTCTCTC
actc	ACACACACACACACTCTCTCTC
acat	ACACACACACACACACATATATA
tgta	TGTGTGTGTGTGTGTGTATATAT
STM primers (5 → 3′)	
stm5tgag	AATTGCCCTTTAAGGCTAGACA
stm511tgag	CACAGTCTCAGATTCATCTATTCA
stm25tgag	GCATTCTACTTCTAGGATCTTCTG
stm15tgag	GAGTAGGTCAAGCACCAATGAGG
stm549tgag	CTGTTGCTTGCTCTGCACCCTT
stm552tgag	ATGGAGGGGTATAAAGACAGCG

mixed with an equal volume of gel loading buffer (98% formamide, 10 mM EDTA, 0.25% xylene cyanol as tracking dye), heated at 95°C for 3 min, chilled on ice and separated on a Gel Scan 2000 DNA fragment analyzer (Corbett Research) using a 5% sequencing gel (Sambrook and Russell 2001).

For multiplex STM amplification, PCR was performed essentially as described above except that up to six STM primers were included in a single reaction, each at a concentration of 0.2 mM. The selection of STMs for multiplex PCR required that each marker had the same microsatellite anchoring primer, and different PCR product sizes to avoid allele overlap. PCR was performed for 40 cycles with the profile: 60 s at 92°C, 60 s at 55°C and 30 s at 72°C. PCR products were separated and detected on a Gel Scan 2000 (Corbett Research) as described above.

Genetic mapping

Preliminary linkage analysis and the construction of genetic maps was performed through an iterative process using MapManager QTX (Manly 1998) and JoinMap v3.0 (Van Ooijen and Voorrips 2001). Initially, the chromosomal assignment of STMs was determined by integrating (P<0.001) polymorphic markers into the genetic maps of appropriate mapping populations using MapManager. Final marker loci orders were determined using JoinMap. The genetic map of the ITMI W7984 × Opata85 cross into which STMs were integrated consisted of 620 loci, including 353 RFLPs (Nelson et al. 1995a, b, c; Marino et al. 1996; Van Deynze et al. 1995) and 267 microsatellites (Roder et al. 1998). Marker data was obtained from the Graingenes website (http://www.wheat.pw.usda.gov/ggpages/maps.html). The marker loci used to construct the ITMI genetic map were those reported by Roder et al. (1998), supplemented with 218 RFLPs present in the published genetic maps of the other mapping populations. The latter RFLP loci were included to facilitate map alignment between crosses, and assist with the construction of a consensus STM map in the future. The genetic maps of the Cranbrook × Halberd and Chinese Spring × Synthetic crosses were published by Chalmers et al. (2001) and on the Graingenes website (http://www.wheat.pw.usda.gov/ggpages/maps.html), respectively. Map data for the crosses was kindly provided by R. Appels (WA Department of Agriculture), or obtained from the Graingenes website.

Polymorphism information content

Polymorphism information content (PIC) values were calculated following Anderson et al. (1993), assuming that the wheat lines were homozygous:

$$\mathrm{PIC}_i = 1 - \sum_{j=1}^{n} P_{ij}^2$$

where P_{ij} is the frequency of the jth pattern for marker i and the summation extends over n patterns. This value provides an estimate of the discriminatory power of a SSR locus by taking into account not only the number of alleles per locus, but also their relative frequencies in the germplasm studied.

Results

Generation of STM profiles

Sequence-tagged microsatellite profiles were developed for ten different compound repeat motifs (Table 1). As the creation of nucleotide sequence tags is a unidirectional process (see Electronic Supplementary Materials and Methods), the sequence tags for each profile are expected to represent unique SSRs in the bread wheat genome, despite complementarity of some of the compound repeat motifs targeted, for example $(Tg)_m(Ag)_n$ and $(Tc)_m(Ac)_n$. Each profile consisted of at least 1,270 nucleotide sequence tags corresponding to DNA fragments that contained a compound SSR in the pool of *Pst*I–*Mse*I restriction fragments from which the nucleotide tags were

prepared. The average number of nucleotide sequence tags per profile was 1,546 (Table 1). As 96 plasmid clones containing concatenated nucleotide sequence tags were sequenced for each microsatellite profile, each DNA sequencing reaction allowed the characterization of about 16 nucleotide tags (or SSRs).

In most instances, the high frequency of low-copy (single) nucleotide tags in the STM profiles indicated that insufficient plasmid clones were sequenced to fully characterize each library (Table 1). As a consequence, the ability to accurately quantify the frequency and distribution of target microsatellites in the *Pst*I–*Mse*I fraction of the bread wheat genome was limited. The lowest number of low-copy tags was observed for the $(Tg)_m(TA)_n$ profile, which had a nucleotide tag redundancy of 94.1% (Table 1). This result indicated that this class of compound repeat motif was either poorly represented in the *Pst*I–*Mse*I genomic fraction or consisted of a relatively small number of highly repetitive sequences. In contrast, the microsatellite profile for $(Tg)_m(Ag)_n$ repeat motifs consisted mostly of low-copy tags, with only 14.8% of nucleotide sequence tags showing redundancy (Table 1). Based on the frequency of low-copy nucleotide sequence tags observed in the microsatellite profiles, their order of usefulness for the design of STM primers was $(Tg)_m(Ag)_n > (Tc)_m(Ac)_n > (Tc)_m(Tg)_n > (Ac)_m(Ag)_n > (Ag)_m(Ac)_n > (Ag)_m(Tg)_n >> (Tg)_m(Tc)_n > (Ac)_m(Tc)_n > (Ac)_m(AT)_n > (Tg)_m(TA)_n$. Despite containing only 27 bp of DNA sequence, PCR primers could be designed for 68% of the low-copy nucleotide sequence tags present in the ten microsatellite profiles.

Amplification of STMs

A total of 2,304 STM primers were synthesized for low-copy nucleotide sequence tags present in the ten microsatellite profiles. Each primer was tested on the 12 parental bread wheat varieties to assess PCR amplification and polymorphism. Overall, 1,289 (56%) of the STM primers amplified a scorable microsatellite fragment with a level of PCR specificity typical of that observed for published SSR markers. Of the 1,289 scorable STMs, 62% amplified a single SSR locus, 31% amplified two or three loci, and 7% gave complex patterns with more than three bands. Fifty-three percent of the functional STMs revealed polymorphism among the 12 parental wheat varieties. The overall success rate for the development of polymorphic STMs was 30% (684/2,304). No significant correlation was found between STMs targeting different classes of compound microsatellites, and the success rate for PCR amplification and level of polymorphism detected (data not shown). Primer sequences for polymorphic markers are provided in the Electronic Supplementary Table S1.

An initial observation from the screening of STMs for polymorphism on the 12 parental bread wheat varieties was the occurrence of a relatively narrow range of allele sizes amongst polymorphic markers. However, this did not interfere with marker scoring as most STMs revealed only a small amount of microsatellite stuttering, typically one to three stutter bands (Fig. 1). To investigate these observations further, the number of alleles, PIC and allele size range was determined for 12 Australian bread wheat varieties using 96 randomly selected polymorphic STMs, and 96 published Gatersleben SSRs (Roder et al. 1998). The average PIC and number of alleles observed for the STMs was significantly ($P < 0.05$) lower than that of the published SSRs (0.57 and 0.69 PIC, and 3.6 and 5.0 alleles per marker, respectively). However, the average allele size range for the STMs was half that observed for the published SSRs (9 bp versus 18 bp, respectively) (Table 3).

To investigate the cause of the reduced allele size range observed the STMs, the PCR products amplified from the bread wheat variety Halberd were sequenced. Figure 2 shows the distribution of the microsatellite repeat lengths observed in the sequenced STMs. More than half (57%) of the microsatellites were less than 10 dinucleotide repeat units in length, 29% had between 10 and 24 repeat units and 14% had more than 24 repeat units (average 13). In contrast, the average microsatellite repeat length reported for the published SSRs was 28 units (Fig. 2). This data revealed that the STMs typically contained shorter repeat sequences than the Gatersleben SSRs. The microsatellite repeat length distribution also showed that the microsatellite anchoring primers used for STM amplification were faithfully anchoring to the repeat junction of the compound SSRs, as most (95/96) of the STMs contained repeat motifs that were longer than that of the microsatellite anchoring primer, that is four dinucleotide repeat units.

Genetic mapping

A total of 380 STMs were genetically mapped in at least one of the three mapping populations, with 296, 72 and 94 markers being mapped in the ITMI W7984 × Opata85, Chinese Spring × Synthetic and Cranbrook × Halberd crosses, and 366, 84 and 110 loci, respectively. The mapped STMs showed a relatively even distribution along the individual linkage groups of the W7984 × Opata85 cross, except in the centromeric

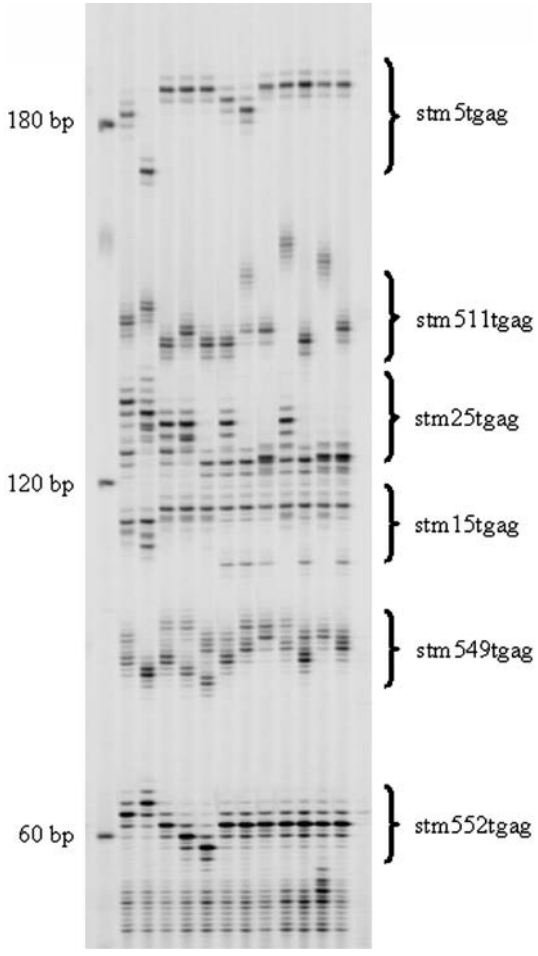

Fig. 1 Six-plex PCR performed on the 12 parental lines of the mapping populations. The primer sequences of STMs used in this assay are listed in Table 2

regions of some chromosomes (Fig. 3). Across the three mapping populations, a similar number of loci were mapped to the A and D genomes, while twice as many mapped to the B genome. The STMs showed a relatively even distribution across the seven homoeologous chromosomes, although chromosomes 3B and 7B were overrepresented. In contrast, relatively few STMs were mapped to chromosomes 1D, 4D and 5A (Table 4). Ten percent (36/380) of the STMs amplified two or more homoeoloci. Thirty-four STMs were mapped to

Table 3 Number of alleles, allele size range (bp) and PIC observed for 96 published Gatersleben SSRs and 96 randomly selected polymorphic STMs

	STMs	Gatersleben SSRs
Average number of alleles	3.62	4.97
Average PIC	0.57	0.69
Average allele size range	9.1	17.6
Median allele size range	6.0	13.5

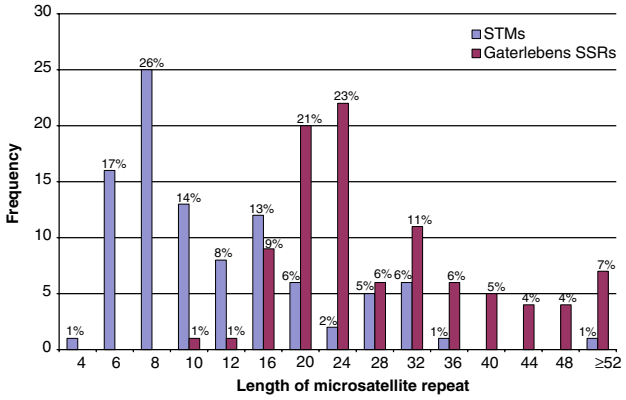

Fig. 2 Length distribution of microsatellite repeats observed in the sequenced STMs, and reported for the published Gatersleben SSRs

two homoeologous chromosomes, and two STMs mapped to all three homoeologous chromosomes. In all other instances, STMs amplifying more than one locus mapped to non-homologous (paralogous) sites.

Multiplex PCR

An anticipated advantage of STMs was high amenability to multiplex PCR, in which several markers are co-amplified under identical conditions, in the same reaction. To determine the amenability of STMs to multiplex PCR amplification, 74 three-plex and 63 six-plex assays were performed using the 12 parental lines of the mapping populations. To construct multiplex marker panels, STMs that required the same microsatellite anchoring primer and had non-overlapping allele sizes were selected. Redundant use of STMs to construct the marker panels was allowed to demonstrate the combinability of different markers in multiplex PCR. Of the multiplex PCR assays performed, SSR fragments of expected size were amplified with a high level of PCR specificity for 88 and 80% of the STMs in three-plex and six-plex reactions, respectively

Table 4 Genomic distribution of STM loci mapped in the ITMI W7984 × Opata85, Chinese Spring × Synthetic and Cranbrook × Halberd crosses

Homeologous group	A-genome	B-genome	D-genome	Total
1	17	28	10	55
2	18	31	25	74
3	30	56	21	107
4	17	24	11	52
5	10	28	15	53
6	13	31	13	57
7	19	37	25	81
Total	124	235	120	479

(Fig. 1). Unsuccessful multiplex PCR assays mostly resulted from the failed amplification of one or more of the STMs, rather than the amplification of detectable non-specific fragments. However, the substitution of the failed STM with a different marker nearly always resulted in a successful multiplex PCR. This showed that the development of robust multiplex STM assays could be rapidly achieved without optimizing PCR conditions or adjusting primer concentrations (data not shown).

Discussion

The success rate typically reported for the development of polymorphic SSRs from genomic DNA of bread wheat is quite low. For example, Bryan et al. (1997), Roder et al. (1998), Pestova et al. (2000) and Song et al. (2005) reported that on average only about 30% (28, 24, 32 and 36%, respectively) of all primer pairs developed from the sequences of anonymous genomic microsatellite clones were polymorphic. In the present study, STMP was used for the large-scale development of wheat SSRs. The results of this study show that the technique can substantially reduce the cost of marker development, whilst maintaining a success rate (30%) for the recovery of polymorphic SSRs similar to that reported for other methods. Overall, we believe that STMP provided at least a twofold cost saving for the development of anonymous genomic SSRs. This cost reduction results from several aspects of the STMP procedure. First, STMP uses anchored PCR to directly generate a genomic DNA library highly enriched for microsatellite fragments, thereby eliminating the requirement to screen genomic DNA libraries via hybridization for microsatellite clones. Second, the DNA sequencing effort required to characterize each microsatellite library was reduced by 16-fold. This was achieved by sequencing for each microsatellite only a characteristic 27-bp nucleotide tag, which was derived from a defined position in the DNA sequence flanking the repeat motif. Concatenization of sequence tags to form long linear chains prior to cloning allowed on average 16 nucleotide tags to be characterized in a single DNA sequencing reaction. And third, the use of anchored PCR to amplify a target microsatellite locus from genomic DNA halved the number of primers that was needed for marker development compared to conventional SSRs, which require a pair of primers flanking each repeat.

The development of STM profiles for ten different repeat motifs revealed that compound SSRs are an abundant class of microsatellite in the bread wheat genome. This was indicated by the high frequency of low (single) copy nucleotide sequence tags in the microsatellite profiles (Table 1). This result implied a high potential for the STMP technique to develop a very large number of wheat STMs. Additional DNA sequencing of the STM profiles would be expected to provide a richer source of markers, as the rate of identification of new nucleotide tags was not saturated at the sequencing depth reached. In-depth sequencing of the nucleotide tag profiles would also improve marker quality. A more accurate profile of the abundance and frequency of microsatellites in the bread wheat genome would allow more efficient selection of low-copy nucleotide tags for primer design. It would also increase the ability to determine in silico the predicted PCR specificity of a primer derived from a low-copy nucleotide sequence tag. In practice, however, there would be a trade-off between the additional cost of sequencing a nucleotide tag profile to depth, the rate of identification of new tags, and the ability to more accurately determine in silico the predicted PCR specificity of a STM primer.

Sequence-tagged microsatellites revealed a genomic distribution similar to that reported for conventional SSRs mapped in the same populations (Roder et al. 1998; Song et al. 2005). Marker clustering observed around the centromeres of some chromosomes is consistent with published cereal genetic maps, and is likely the result of suppressed recombination around the centromeres (Ramsay et al. 2000; Kunzel et al. 2000). Low recombination in crosses made between certain genetic backgrounds can also cause apparent marker clustering in affected chromosomal regions, such as chromosome 4B in the ITMI W7984 × Opata85 cross (Fig. 3). While there was clustering of several STMs targeting the same compound repeat motif in a few chromosomal regions this is unlikely to have resulted from the amplification of the same microsatellite locus, since each STM primer was derived from a unique nucleotide sequence tag. Across the three mapping populations, the STMs showed a non-random distribution among the wheat genomes with almost twice as many markers mapped to the B genome (Table 4). This non-random distribution follows a pattern reported for most types of genetically mapped wheat markers including RFLPs, ESTs and both genomic and EST-derived SSRs, and is attributed to high polymorphism in the B genome (Peng and Lapitan 2004; Qi et al. 2004; Roder et al. 1998; Somers et al. 2004).

An interesting observation from the screening of STMs on the 12 parental lines of the mapping populations was the occurrence of a relatively narrow allele size range, and a reduced level of polymorphism

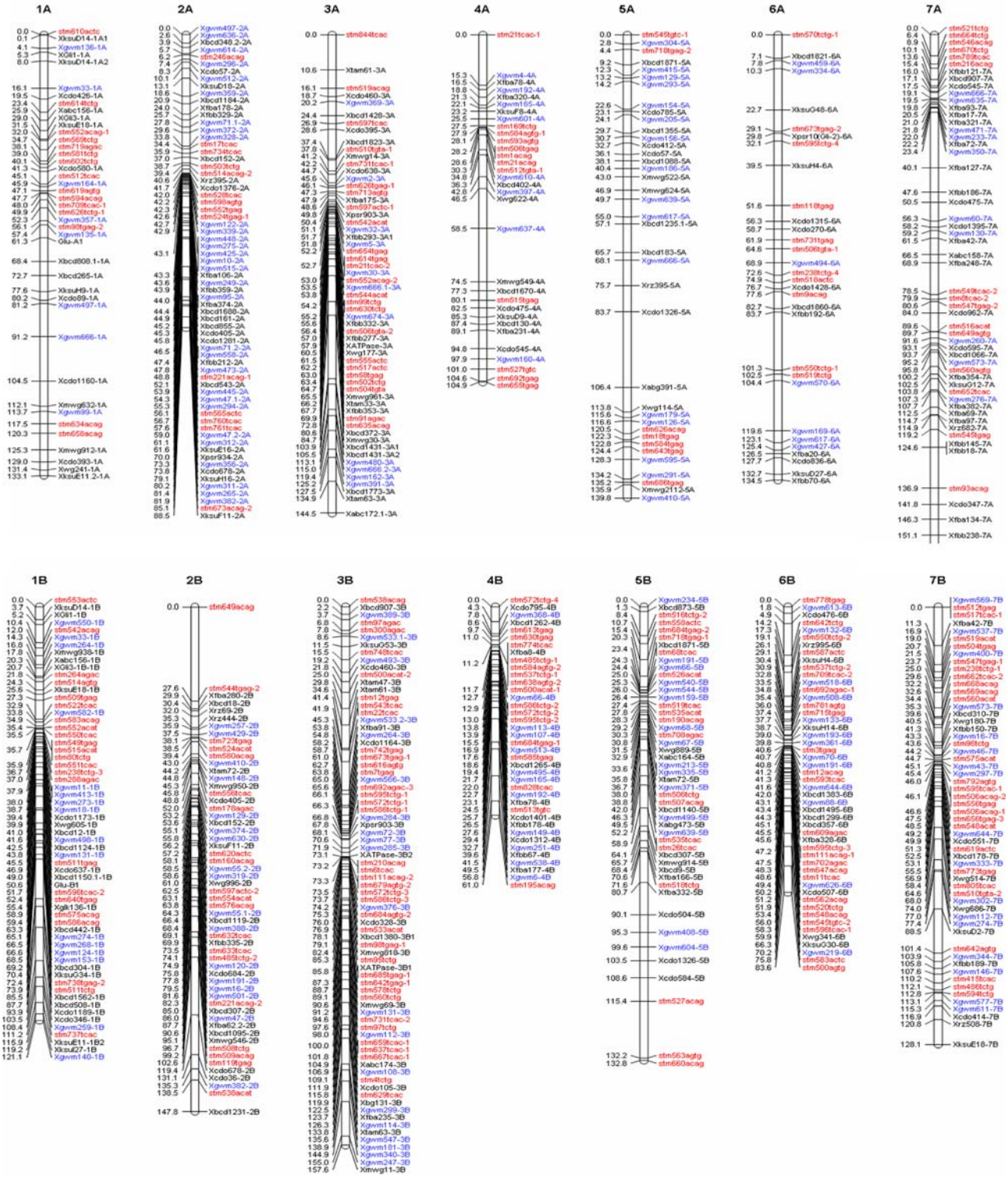

Fig. 3 Genetic map of the ITMI W7984 × Opata85 cross. Cumulative centimorgan distances are indicated on the *left hand side* of each chromosome. The short arm of each chromosome is at the *top*. RFLP and Gatersleben SSR loci derived from GrainGenes data are in *normal* and *blue font*, respectively. STM loci are in *red*. STMs amplifying more than one locus are denoted with a *dash* (–) at the end of their name, followed by an *integer* indicating the *number of loci* that were genetically mapped in the population

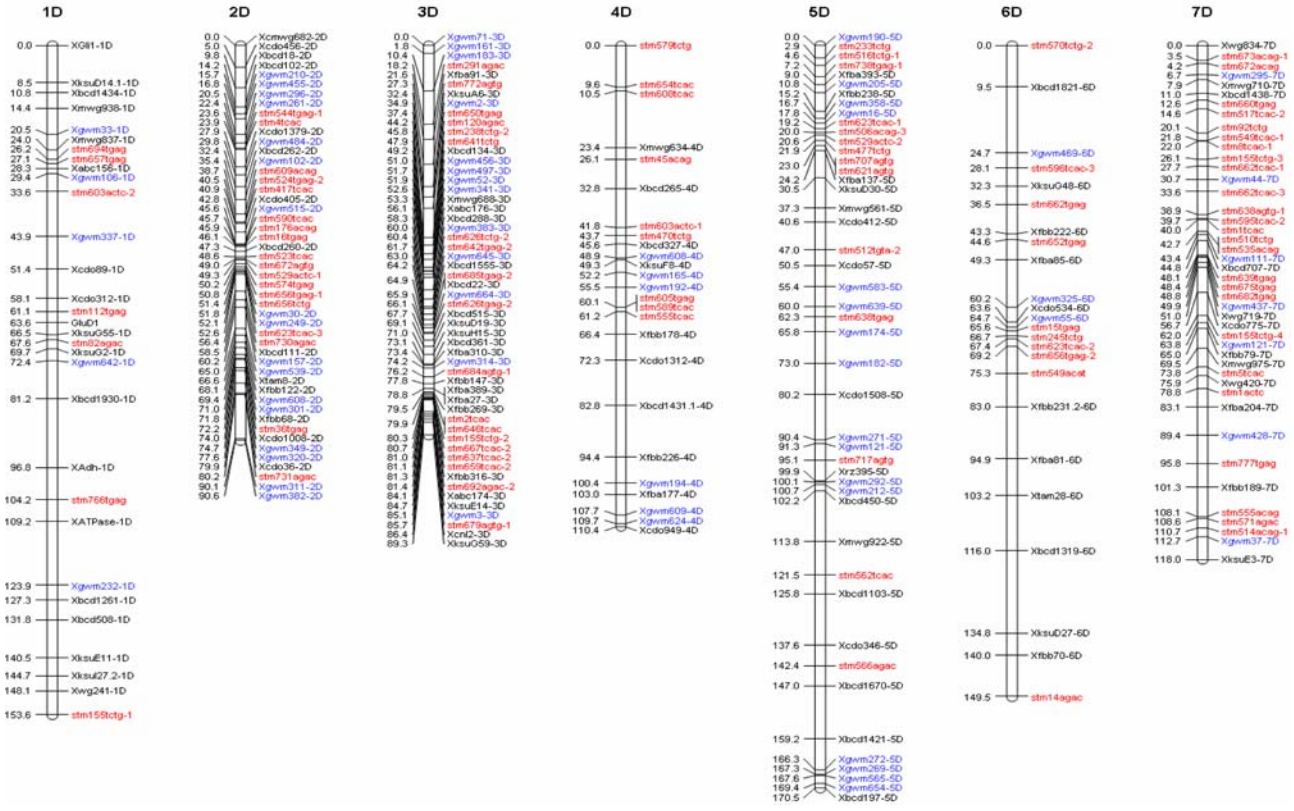

Fig. 3 continued

compared to published SSRs (Table 3). One possible explanation for this observed difference is the use of anchored PCR to generate STM libraries, rather than a hybridization-based method such as those typically used to develop conventional SSRs. Comparison of the microsatellite repeat length distribution for 96 Gatersleben SSRs and 96 randomly selected polymorphic STMs revealed an abundance of short microsatellites sequences in STMs (Fig. 2). A predominance of SSRs with short (<11) repeat sequences has been reported for PCR-enriched microsatellite libraries in several species including barley (Davila et al. 1999) and pine (Fisher et al. 1998). In contrast, SSRs developed using hybridization-based methods typically have longer repeat sequences. For example, the average repeat length in Gatersleben SSRs published by Roder et al. (1998) was 21 dinucleotide units. A positive correlation between microsatellite repeat length and polymorphism has been reported in some species (Bryan et al. 1997; Smulders et al. 1997), but not others (Taramino and Tingey 1996). In the present study, the reduced informativeness of STMs may indicate a positive correlation between microsatellite repeat length and polymorphism in bread wheat. Nevertheless, STMs reveal a useful (moderate) level of polymorphism for genetic analysis, perhaps more comparable to that reported for EST-SSRs. Moreover, STM libraries may prove useful for developing markers for an apparently abundant class of short SSRs that is less accessible to hybridization-based methods.

Sequence-tagged microsatellites provide several technological advantages that can help to increase marker assay throughput and reduce genotyping costs. The amplification of STMs with a single microsatellite anchoring primer that is common to all markers targeting the same repeat motif provides substantial cost savings for fluorescent-based SSR analysis because dye-labeling of only the microsatellite anchoring primer is needed. Hence, the cost of deploying STMs on an automated DNA fragment analyzer is substantially reduced. In addition, the small allele size range observed for STMs allows for more efficient capillary electrophoresis since a greater number of markers can be separated at a time without risk of allele overlap (Table 3). Further increasing genotyping throughput is the high amenability of STMs to multiplex PCR, a technique that enables the co-amplification of several markers under identical conditions, in the same reaction. The multiplex compatibility of STMs is believed to arise from the reduced number (n+1) of primers in a

multiplex STM assay, compared to the number ($2n$) of primers in a conventional multiplex PCR assay. This reduces the probability of undesirable primer–primer interactions that can lead to assay failure (Elnifro et al. 2000). These advantages suggest that STMs will strongly complement other high-throughput marker technologies that have been developed recently for wheat, such as the array-based DArT technology (Jaccoud et al. 2001). Indeed, the co-mapping of STMs and DArT markers has commenced (Akbari et al. 2006).

In conclusion, the present study has demonstrated STMP as an economical technique for the development of STMs in complex polyploid genomes, such as bread wheat. STMP reduced the cost of marker development by at least twofold compared to traditional methods, and provided markers with several technological advances over published SSRs amplified in conventional PCR assays. These advantages include reduced marker deployment costs for fluorescent-based SSR analysis, high amenability to multiplex PCR and more efficient electrophoretic separation since a greater number of markers can be separated at a time. The increased marker density provided by the addition of 479 STM loci to the published wheat genetic map will facilitate the mapping of genes and QTL, and support studies of genetic diversity, pedigree analysis and association mapping. The STMs also provide greater opportunity to select markers along chromosomal regions of interest to improve selection efficiency in marker-assisted breeding, the recovery of recurrent genetic background, and to assemble complex genotypes for breeding purposes.

Acknowledgments This research was supported by the Grains Research and Development Corporation, Australia, the Value Added Wheat CRC, the Biotechnology and Biological Sciences Research Council, UK and the International Atomic Energy Agency (IAEA RCA-11023).

References

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes H, Sharp P, Vaughan P, Rathmell W, Huttner E, Kilian A (2006) Diversity Arrays Technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet (in press)

Anderson JA, Churchill GA, Autrique EJ, Tanksley SD, Sorrells ME (1993) Optimising parental selection for genetic linkage maps. Genome 36:181–186

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Bryan GK, Collins AJ, Stephenson P, Orry A, Smith JB, Gale MD (1997) Isolation and characterisation of microsatellites from hexaploid bread wheat. Theor Appl Genet 94:557–563

Chalmers KJ, Campbell AW, Kretschmer J, Karakousis A, Henschke PH, Pierens S, Harker N, Pallotta M, Cornish GB, Shariflou MR, Rampling LR, McLauchlan A, Daggard G, Sharp PJ, Holton TA, Sutherland MW, Appels R, Langridge P (2001) Construction of three linkage maps in bread wheat, *Triticum aestivum*. Aust J Agric Res 52:1089–1119

Cregan PB, Mudge J, Fickus EW, Marek LF, Danesh D, Denny R, Shoemaker RC, Matthews BF, Jarvik T, Young ND (1998) Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. Theor Appl Genet 98:919–928

Davila JA, Loarce Y, Ferrer E (1999) Molecular characterisation and genetic mapping of random amplified microsatellite polymorphism in barley. Theor Appl Genet 98:265–273

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

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:759–760

Elnifro EM, Ashshi AM, Cooper RJ, Klapper PE (2000) Multiplex PCR: optimisation and application in diagnostic virology. Clin Microbiol Rev 13:559–570

Engels WR (2004) Amplify v3.1.4. http://www.engels.genetics.wisc.edu/amplify/

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

Fisher PJ, Richardson TE, Gardner RC (1998) Characterisation of single- and multi-copy microsatellites from *Pinus radiata*. Theor Appl Genet 96:969–979

Hayden MJ, Sharp PJ (2001a) Targeted development of informative microsatellite (SSR) markers. Nucleic Acids Res 29:e44

Hayden MJ, Sharp PJ (2001b) Sequence tagged microsatellite profiling (STMP): a rapid technique for developing SSR markers. Nucleic Acids Res 29:e43

Hayden MJ, Stephenson P, Logojan AM, Khatkar D, Rogers C, Koebner RMD, Snape JW, Sharp PJ (2004) A new approach to extending the wheat marker pool by anchored PCR amplification of compound SSRs. Theor Appl Genet 108:733–742

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

Jian SM, Brar DS, Ahloowalia BS (2002) Molecular techniques in crop improvement. Kluwer, Dordrecht

Kammholz SJ, Campbell AW, Sutherland MW, Hollamby GJ, Martin PJ, Eastwood RF, Barclay I, Wilson RE, Brennan PS, Sheppard JA (2001) Establishment and characterisation of wheat genetic mapping populations. Aust J Agric Res 52:1079–1088

Kim HS, Ward R (2000) Patterns of RFLP-based genetic diversity in germplasm pools of common wheat with different geographical or breeding program origins. Euphytica 115:197–208

Koebner RMD, Powell W, Donini P (2001) The contribution of current and forthcoming DNA molecular marker technologies to wheat and barley genetics and breeding. Plant Breed Rev 21:181–220

Kunzel G, Korzun L, Meister A (2000) Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 154:397–412

Lagercrantz E, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111–1115

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

Manly KF (1998) Map Manager QTX. http://www.mapmgr.roswellpark.org/mmQTX.html

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

Nelson JC, Sorrells ME, van Deynze AE, Lu YH, Atkinson M, Bernard M, Leroy P, Faris JD, Anderson JA (1995a) Molecular mapping of wheat: major genes and rearrangements in homologous groups 4, 5, and 7. Genetics 141:721–731

Nelson JC, van Deynze AE, Autrique E, Sorrells ME, Lu YH, Merlino M, Atkinson M, Leroy P (1995b) Molecular mapping of wheat, homologous group 2. Genome 38:516–524

Nelson JC, van Deynze AE, Autrique E, Sorrells ME, Lu YH, Negre S, Bernard M, Leroy P (1995c) Molecular mapping of wheat, homologous group 3. Genome 38:525–533

Paull JG, Chalmer KJ, Karakousis A, Kretschmer JM, Manning S, Langridge P (1998) Genetic diversity in Australian wheat varieties and breeding material based on RFLP data. Theor Appl Genet 96:435–446

Peng JH, Lapitan NLV (2004) Characterisation of EST-derived microsatellites in the wheat genome and development of eSSR markers. Funct Integr Genomics 5:80–96

Pestova E, Ganal MW, Roder MS (2000) Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 43:689–697

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–221

Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvoral J, Linkiewicz AM, Ratnarisi A, Dubcovsky J, Bermudez-Kandiania CE, Greene RA, Kantety R, La Rota CM, Munkvold JD, Sorrells SF, Sorrells ME, Dilbirligi M, Sidhu D, Erayman M, Randhawa HS, Sandhu D, Bondareva SN, Gill KS, Mahmoud AA, Ma XF, Miftahudin A, Gustafson JP, Conley EJ, Nduati V, Gonzalez-Hernandez JL, Anderson JA, Peng JH, Lapitan NLV, Hossain KG, Kalavacharla V, Kianian SF, Pathan MS, Zhang DS, Nguyen HT, Choi DW, Fenton RD, Close TJ, McGuire PE, Qualset CO, Gill BS (2004) A chromosome bin map of 16,000 EST loci and distribution of genes among the three genomes of polyploidy wheat. Genetics 168:701–712

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Smulders MJM, Bredemeijer G, Rus-Korekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor Appl Genet 97:264–272

Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

Song QJ, Shi JR, Singh S, Fickus EW, Costa JM, Lewis J, Gill GS, Ward R, Cregan PB (2005) Development and mapping of microsatellite (SSR) markers in wheat. Theor Appl Genet 110:550–560

Stephenson P, Bryan G, Kirby K, Collins A, Devos K, Busso C, Gale M (1998) Fifty new microsatellite loci for the wheat genetic map. Theor Appl Genet 94:946–949

Taramino G, Tingey S (1996) Simple sequence repeats for germplasm analysis and genetic variation. Genome 39:277–287

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

Roder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Van Deynze AE, Dubcovsky J, Gill KS, Nelson JC, Sorrells 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

Van Ooijen JW, Voorrips RE (2001) JoinMap 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484–487

Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. Theor Appl Genet 111:677–687