Theor Appl Genet (1995) 90: 247–252

K. M. Devos · G. J. Bryan · A. J. Collins · P. Stephenson
M. D. Gale

Application of two microsatellite sequences in wheat storage proteins as molecular markers

Received: 3 March 1994 / Accepted: 30 May 1994

Abstract In eukaryotes, tandem arrays of simple-sequence repeat sequences can find applications as highly variable and multi-allelic PCR-based genetic markers. In hexaploid bread wheat, a large-genome inbreeding species with low levels of RFLP, di- and trinucleotide tandem repeats were found in 22 published gene sequences, two of which were converted to PCR-based markers. These were shown to be genome-specific and displayed high levels of variation. These characteristics make them especially suitable for intervarietal breeding applications.

Key words Wheat · Microsatellite markers

Introduction

The availability of restriction fragment length polymorphism (RFLP) markers has led to the rapid development of detailed genetic maps with a vast potential for strategic and practical plant breeding applications. Inbreeding species, such as hexaploid wheat (*Triticum aestivum*), are characterised by low levels of RFLP, and the choice of wide or interspecific mapping populations is essential to the construction of high density maps. The low level of detected variation and the unwieldiness and expense of the technology are major drawbacks to the use of RFLP in commercial breeding programmes employing adapted genotypes and limit map applications mainly to interspecific purposes.

A growing realisation that eukaryotic genomes are interspersed with tandem arrays of di-, tri-, and tetranucleotide repeats that exhibit extensive length variation was the basis for the development of a new generation of polymerase chain reaction (PCR)-based markers. These simple sequence repeat or microsatellite sequences are usually present within stretches of unique DNA that can be specifically amplified using oligonucleotide primers constructed to the microsatellite flanking regions. This type of marker often shows a codominant Mendelian inheritance and has already found widespread use by mammalian geneticists (Weissenbach et al. 1992). In plants, microsatellite markers have been reported in soybean (Akkaya et al. 1992; Morgante and Olivieri 1993), rice (Zhao and Kochert 1993; Wu and Tanksley 1993), maize (Senior and Heun 1993), *Brassica* (Lagercrantz et al. 1993) and *Arabidopsis* (Bell and Ecker 1994). In this paper, we report a detailed analysis of the specificity and variation revealed by two microsatellites present in genes of wheat storage proteins.

Communicated by G. E. Hart

K. M. Devos (✉) · G. J. Bryan · A. J. Collins · P. Stephenson
M. D. Gale
Cambridge Laboratory, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK

Materials and methods

Search of the EMBL and GenBank databases

Using the programme 'FASTA' (Genetics Computer Group, Madison), 327 wheat sequences in the EMBL (release 36) and GenBank (release 79) databases were searched for the presence of microsatellite arrays with a minimum length of 10 and 7 repeat units for 2-bp and 3-bp repeats, respectively.

Length analysis of the microsatellites

Primers

Two forward primers, F1 (GCAGACCTGTGTCATTGGTC) and F2 (GATCTGGCCACAAAGCGC), and one reverse primer, R1 (GATATAGTGGCAGCAGGATACG), were constructed to the microsatellite flanking regions in the γ-gliadin pseudogene (Em_P1:Taglgap; Rafalski 1986). One primer set (P1=TCCCGCCATGAGTCAATC; P2=TTGGGAGACACATTGGCC) was developed to amplify the microsatellite array in the low-molecular-weight (LMW) glutenin gene (Em_P1:Taglut; Pitts et al. 1988).

Genotypes

Length variation of the microsatellites was assessed in a sample of eight wheat varieties, including 'Chinese Spring' (CS), 'Timgalen', 'RL4137', 'Cappelle-Desprez', 'Herzog', 'Hereward', 'Soisson',

'Brigadier', and a synthetic hexaploid (IPSR1190903, McFadden and Sears 1946; Sears 1976). The nullisomic-tetrasomic (NT) lines N1AT1D, N1BT1A, and N1DT1B (Sears 1954) were used to determine the chromosomal location of the microsatellites in the γ-gliadin and LMW-glutenin genes. Mapping was carried out in a population of 120 F_2 or bulked F_3 plants from the wide cross 'Chinese Spring'×'Synthetic'.

DNA extractions

DNA extractions were carried out as described in Sharp et al. (1988)

PCR amplifications

Polymerase chain reaction (PCR) was carried out in a DNA Thermal Cycler (Perkin Elmer Cetus) on 20 ng of template DNA in 50 µl reaction buffer [10 m*M* TRIS.HCl pH 8.3, 1.5 m*M* $MgCl_2$, 50 m*M* KCl, 100 µg/ml gelatine, 200 n*M* primer, 200 µ*M* each of dATP, dCTP, dGTP and TTP, 0.8 U Taq DNA-polymerase (Boehringer)]. After an initial denaturation step at 94°C for 5 min, the samples were subjected to 30 cycles of 1 min denaturation at 94°C, 1 min annealing at 55°C (F1/R1 and F2/R1) or 60°C (P1/P2), and 2 min extension at 72°C. Amplification products were visualised with ethidium bromide on 3% agarose gels (1.5% NuSieve agarose - Flowgen, 1.5% ultrapure agarose - BRL) in TAE buffer and on 6% polyacrylamide (PA) gels in TBE buffer. For reamplification of specific products, the fragments were excised from the gel, eluted in 50 µl H_2O, and 1 µl of a 1/1000 dilution was used as template in the PCR, carried out as above.

Sequencing of the amplification products

Individual amplification products were electro-eluted from the PA gels onto DEAE-paper and cloned into pGEM-T Vector (Promega)

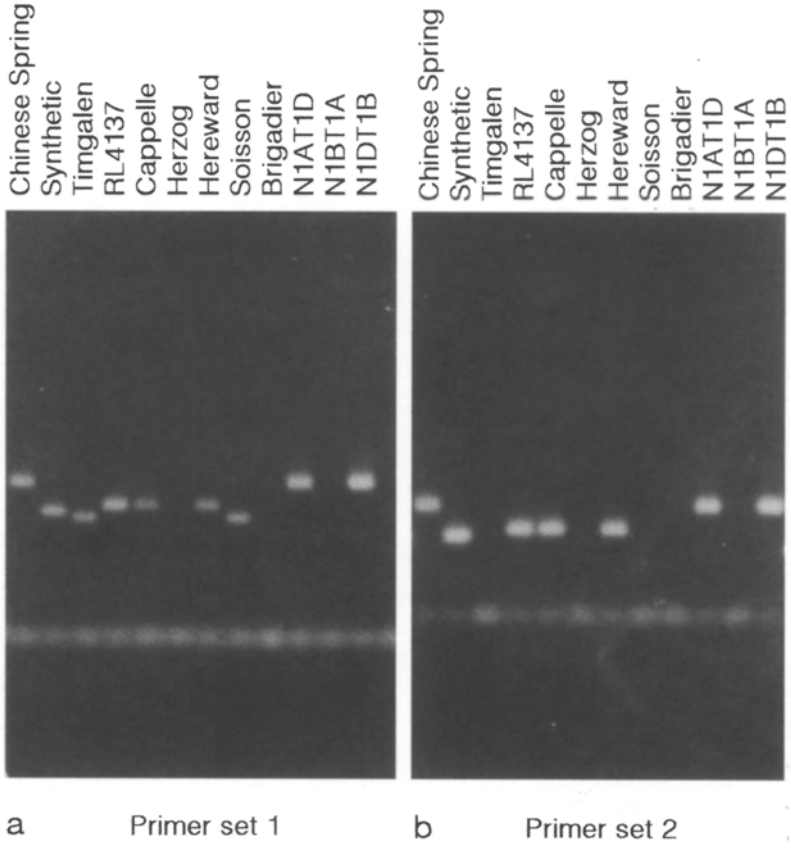


Fig. 1 Chromosomal location and length variation of the amplification products obtained with primers F1/R1 (**a**) and F2/R1 (**b**) in a sample of nine wheat varieties

Table 1 Wheat genes containing microsatellite arrays as present in the EMBL and GenBank databases

Gene family	Database accession	Number of sequences	Microsatellite	Location of microsatellite
α-Amylase 3	Taamya	1	$(TA)_{12}CA(TA)_9$	Upstream of coding region
α-Amylase 2	Taaam234	1	$(GCT)_7$	Coding region
α,β-Gliadin	Taagcnn16, Taagcnn17, Taagcnn35, Tagliabc, Tagliabd, Tagliabh, Tagliag1, Tagliag3, Taglia01, Tagliabb	10	$(CAA)_n$	Coding region
	Taglia, Tagliaa, Tagliaba, Tagliabf, Tagliag2, Taglib, Taabglia, Tagliabg	8	$(CAG)_n(CAA)_n$	Coding region
γ-Gliadin	Taglgap	1	$(CAA)_{15}$	Coding region
LMW-Glutenin	Taglut	1	$(CAG)_5(CAA)_8$	Coding region

according to the manufacturer's recommendations. Sequencing reactions were carried out using the AutoRead Sequencing kit (Pharmacia) and analysed with an Automated Laser Fluorescent DNA Sequencer (Pharmacia).

Results

Twenty-two gene sequences, belonging to five gene families, contained microsatellite arrays (Table 1), of which two were compound. Two microsatellite arrays, one in a γ-gliadin pseudogene and one in a LMW-glutenin gene, were chosen for further study.

γ-gliadin pseudogene

Amplification of the (CAA)-microsatellite using forward primer F1 in combination with reverse primer R1 generated a single amplification product in seven of the nine wheat lines, while no product was obtained in var 'Herzog' and 'Brigadier'. Four length variants were identified among the amplified products (Fig. 1a). Similarly, primer combination F2/R1 gave single, multi-allelic amplification products of comparable size, with null-alleles in wheat var 'Herzog', 'Brigadier', 'Timgalen' and 'Soisson' (Fig. 1b). The sequences are given in Fig. 2. The microsatellite sequence originated from chromosome 1B, as shown by NT-analysis (Fig. 1), and mapped 1.5 cM from *Xpsr11* (*Glu-3*)-*1B*, one of the homoeoloci detected by the LMW-glutenin clone pTag544.

LMW-glutenin gene

Primers constructed to the flanking regions of the (CAG)(CAA) compound microsatellite in the LMW-glutenin gene generated a set of one or two amplification products in each wheat variety, and at least five patterns (Fig. 3a) were observed in our sample of nine wheat lines. Nul-

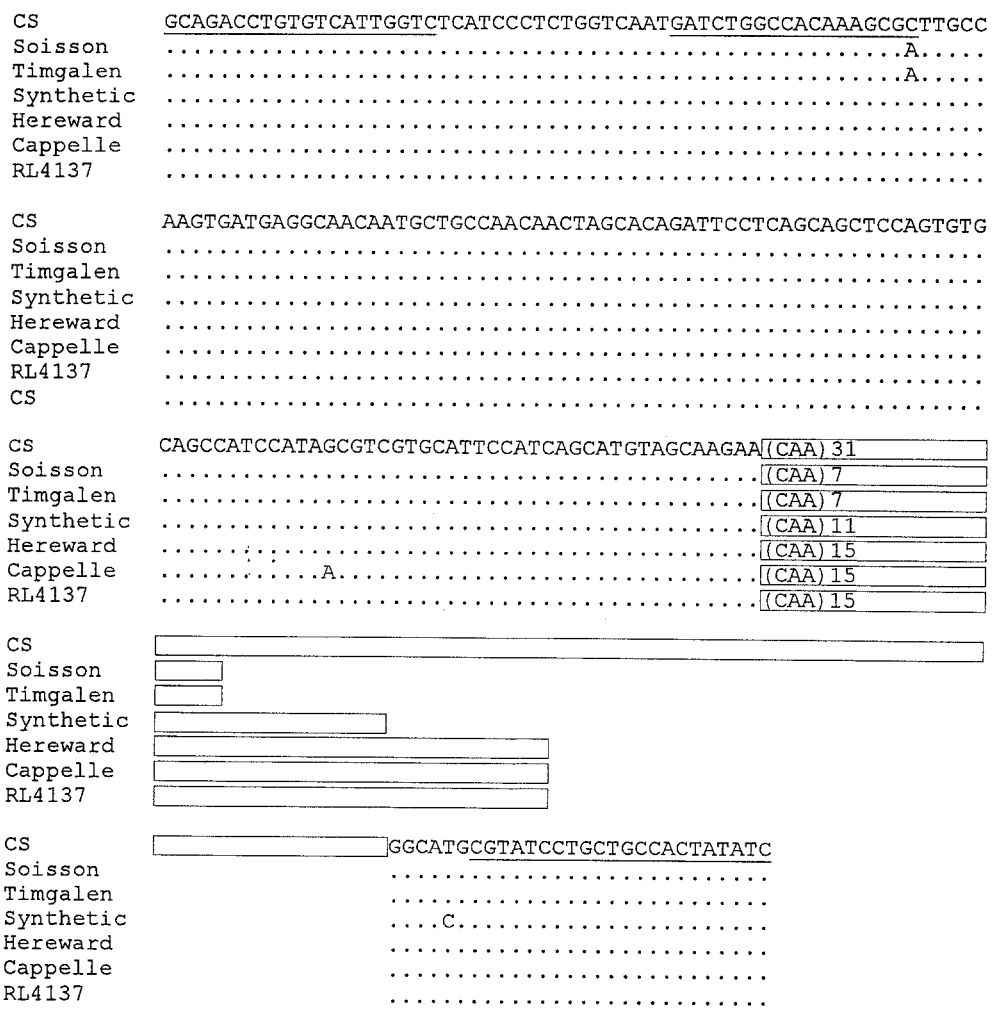
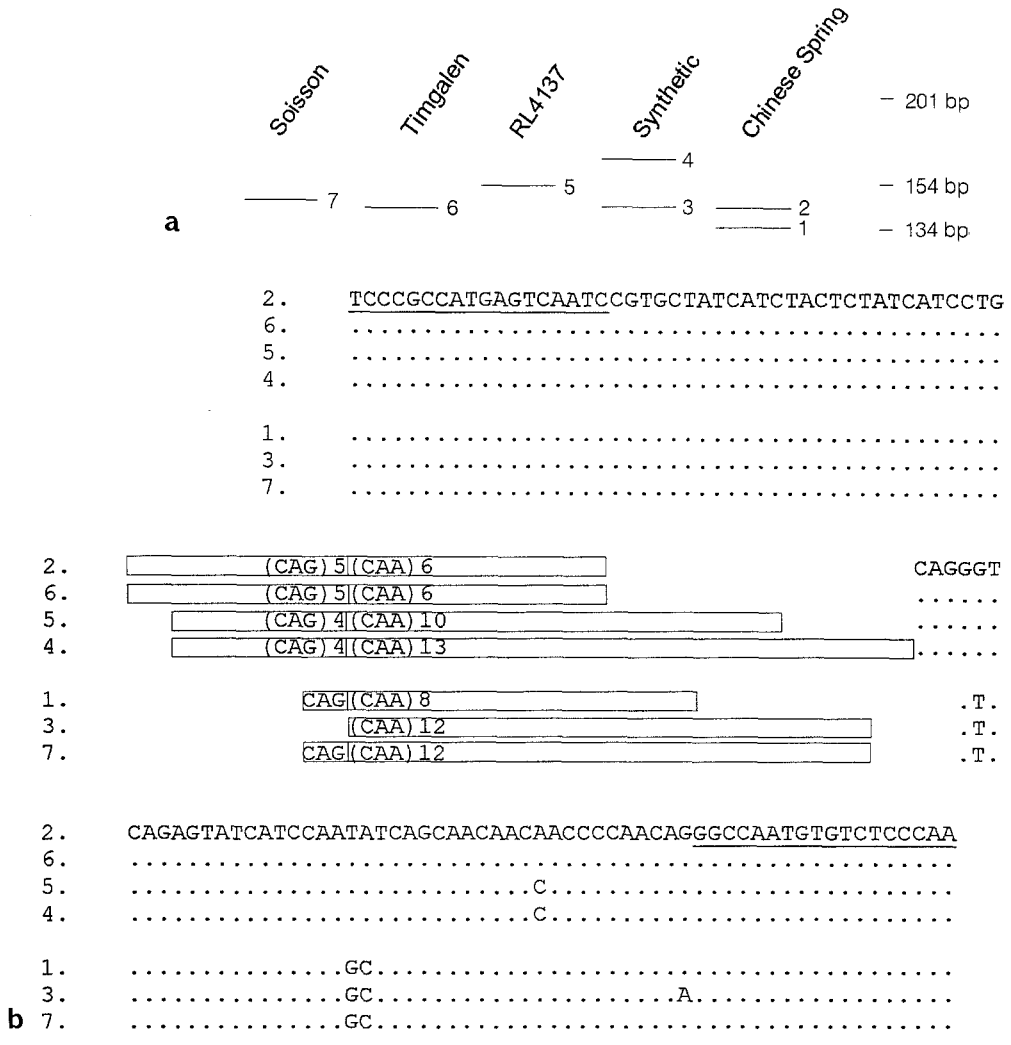


Fig. 2 Sequence analysis of the individual amplification products showing length variation of the γ-gliadin pseudogene microsatellite sequence. *Dots* represent nucleotides that are identical to the γ-gliadin pseudogene in CS. The microsatellite arrays are *boxed*. The forward and reverse primers are *underlined*



lisomic-tetrasomic analysis showed the two CS fragments both to be located on chromosome 1A, and complete co-segregation of both CS and 'Synthetic' fragments in 120 F_2 plants indicated that these fragments identified only a single genetic locus, which co-segregated with *Xpsr11*(*Glu-3*)-*1A*. The sequences of the seven products are given in Fig. 3b.

Discussion

Primer-specificity

In order to design primer combinations that would specifically amplify the (CAA)-microsatellite in the γ-gliadin pseudogene, a sequence comparison of the pseudogene with all γ-gliadin genes present in the EMBL and GenBank databases was carried out (data not shown). Sequence variation, unique to the pseudogene, was chosen to coincide with the 3′ end of the forward and reverse primers. All primer combinations indeed gave single amplification products, shown by NT-analysis to be located on chromosome 1B (Fig. 1). Furthermore, the generation of a product from chromosome 1B, but not from its homoeologues 1A and 1D, demonstrates the genome-specific nature of the primer sequences. The genome-specificity of the primers was further supported by the fact that no amplification products were obtained from var 'Herzog' and 'Brigadier', which carry a 1BL.1RS translocation (unpublished results). The primer set P1/P2, designed to amplify the (CAG)(CAA) microsatellite in a LMW-glutenin gene, generated one or two amplification products from each of the nine varieties tested. Analysis of the NT-lines showed the origin of the two CS products to be chromosome 1A and that, with the PCR regime used, the primers were genome-specific.

The question of genome specificity of the primer sets was explored further. A database search revealed that se-

quences with more than 88% homology with both P1 and P2 and which matched perfectly at the 3′ end were present in a LMW-glutenin gene (*Glu-D3*, on chromosome 1D) and in a γ-gliadin gene (*Gli-1-1*, location unknown). Nevertheless, in the present experiment the primers showed chromosome- and genome-specificity. In an effort to further test the specificity of P1 and P2, we found that when the annealing temperature was lowered to 55°C and the number of cycles increased to 60, some new products were observed, but even then only in the absence of the perfect primer sites on chromosome 1A. Sequence analysis of a N1AT1D amplification product revealed 96.7% homology with both the *Glu-D3* and *Gli-1-1* genes. These results indicate that while in an allopolyploid primer pairs are likely to be genome-specific, in heterologous situations the primers may amplify sequences which may, or may not, be present in homoeoloci. This has been reported by Moore et al. (1991) in a microsatellite study in closely related mammals such as cattle and sheep. However, even with reduced annealing stringencies, in the presence of both the homologous and heterologous sequences, amplification products are likely to be generated only from the perfectly matching primer sites.

Sequence variation

The γ-gliadin primers detected a total of five (four length variants and a null) alleles in our sample of nine varieties, corresponding to a pairwise polymorphism value of 86%, and at least five amplification patterns, corresponding to a polymorphism level of 72%, were observed with the LMW-glutenin primers. This level of variation compares favourably to that obtained by D'Ovidio et al. (1990) in a sample of *T. durum* cultivars using PCR-amplification of a γ-gliadin gene sequence. In order to examine the basis of the variation, individual amplification products were cloned and sequenced (Figs. 2 and 3b). The LMW-glutenin amplification products can be classified in two groups, based on their sequence and the nature of the microsatellite (Fig. 3b). The first group contains a (CAG)(CAA) compound repeat and shows high homology to the LMW-glutenin gene sequence from the variety 'Yamhill' extracted from the database. The second group carries a (CAA) microsatellite and is characterised by three base pair mutations with respect to the group 1 sequences. As the LMW-glutenin family comprises some eight to ten genes in the A-genome (Sabelli and Shewry 1991), it is likely that the two sets of sequences, which were completely linked in our mapping population, represent two different members of the LMW-glutenin family. For the γ-gliadin microsatellite markers, a high sequence similarity was observed for all alleles, with most of the variation being due to a variable number of (CAA) repeats (Fig. 2). Interestingly, var 'Timgalen' and 'Soisson', which failed to give an amplification product with primer set F2/R1, carried a single base pair mutation (C→A) (Fig. 2) that marked the 3′ end of primer P2. Clearly, a single T:C base-pair mismatch between a template and 3′ end of a primer sequence can be adequate to inhibit primer extension. Kwok et al. (1990) showed that only A:G and C:C mismatches at the terminal 3′ end position of the primer led to a more than 100-fold reduction of the amplification efficiency. Other experiments, however, suggest that the effects of mismatches on amplification may vary with primer length, sequence context and reaction conditions (Kwok et al. 1990).

Conclusions

In this study, we have shown the potential of microsatellite sequences as a PCR-based alternative to RFLP markers in wheat. Even within multigene families, minor sequence variation in the microsatellite flanking regions can be exploited to construct highly specific primer sets, as demonstrated with the γ-gliadin gene family. Indeed, a single base-pair difference at the 3′ end of a primer can render primer pairs ineffective. Consequently, primer pairs are likely to generate genome-specific markers and thus have significant potential for application in wheat breeding programmes. In other applications, such as interspecific introgression studies, where amplification of homoeoloci from the alien donor parent cannot be predicted using primers derived from the cultivated species sequence, or in intergenomic comparative mapping experiments, where PCR-based microsatellite markers are unlikely to be transferable between species, these markers will lose much of their power as a molecular marker system.

Acknowledgements This work was undertaken as part of a programme jointly funded by the BBSRC, DTI, and ZENECA Plc, Plant Breeding International Cambridge Ltd., Nickerson BIOCEM Ltd. and Ciba-Geigy Plc. We thank J. B. Smith for the sequencing.

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132: 1131–1139

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144

D'Ovidio R, Tanzarella OA, Porceddu E (1990) Rapid and efficient detection of genetic polymorphism in wheat through amplification by polymerase chain reaction. Plant Mol Biol 15:169–171

Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nucleic Acids Res 18:999–1005

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

McFadden ES, Sears ER (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J Hered 37:81–89

Moore SS, Sargeant LL, King TJ, Mattick JS, Georges M, Hetzel DJS (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10: 654–660

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

Pitts EG, Rafalski JA, Hedgcoth C (1988) Nucleotide sequence and encoded amino acid sequence of a genomic gene region for a low-molecular-weight glutenin. Nucleic Acids Res 16:11376

Rafalski JA (1986) Structure of wheat gamma-gliadin genes. Gene 43:221–229

Sabelli PA, Shewry PR (1991) Characterization and organization of gene families at the *Gli-1* loci of bread and durum wheats by restriction fragment analysis. Theor Appl Genet 83:209–216

Sears ER (1954) The aneuploids of common wheat. Mo Agric Exp Stn Res Bull 572:1–59

Sears ER (1976) A synthetic hexaploid wheat with fragile rachis. Wheat Inf Serv 41–42:31–32

Senior ML, Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36:884–889

Sharp PJ, Kreis M, Shewry P, Gale MD (1988) Location of β-amylase sequences in wheat and its relatives. Theor Appl Genet 75:289–290

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second-generation linkage map of the human genome. Nature 359:794–801

Wu K-S, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241: 225–235

Zhao X, Kochert G (1993) Phylogenetic distribution and genetic mapping of a (GGC)n microsatellite from rice (*Oryza sativa* L.). Plant Mol Biol 21:607–614