Theor Appl Genet (2000) 100:328–336

S. Van Campenhout · R.M.D. Koebner · G. Volckaert

The applicability of consensus PCR primers across species and genera: the use of wheat Em sequences to develop markers for orthologues in rye

Received: 7 December 1998 / Accepted: 21 July 1999

Abstract Two wheat consensus primer sets, directed to "early-methionine-labelled" (Em) gene sequences, were tested for their ability to amplify beyond their original source. A range of widely diverse templates, including other Triticeae species and sample monocot and dicot species, was assayed. Primer set EMC5/EMC3, amplifying the entire coding region with its intron and part of the 3' untranslated region, targets Triticeae and sorghum Em sequences. The other set, EMC5/EMCO31, directed to the coding region and its intron, amplifies templates from all the grass species. Both primer sets fail to amplify Em sequences from more distant monocots and the dicots. Using set EMC5/EMC3, we isolated and sequenced ten members of the rye Em gene family from five different rye sources. Significant DNA sequence variation between wheat and rye sequences in the non-coding regions was found, and this was used to develop seven sequence-specific primers. Twelve primer combinations were analysed, 7 of which were *Em-R1*-specific, amplifying a product in at least one of the tested rye or rye-carrying genotypes but not in wheat. Four sets exhibited clear amplification length polymorphisms which allowed discrimination between and within the rye sources. The primers also discriminated between wheat-rye recombinants with proximal 1RL rye chromatin and those carrying distal 1RL rye chromatin. These results show that wheat consensus primer sets can be used to isolate orthologous sequences, especially from species that are used for alien gene transfer in wheat. Subsequently, species-specific assays can be designed that are useful tools for this application.

Communicated by G.E. Hart

S. Van Campenhout (✉) · G. Volckaert
Katholieke Universiteit Leuven, Laboratory of Gene Technology,
K. Mercierlaan 92, B-3001 Leuven, Belgium
Fax. +32–16–321965
e-mail: steven.campenhout@agr.kuleuven.ac.be

R.M.D. Koebner
John Innes Centre, Norwich Research Park,
Colney NR4 7UH, UK

Key words Early-methionine-labelled protein multigene family · Heterologous PCR primers · Grasses · Sequence-tagged-sites · Wheat-rye recombinants

Introduction

Sequence-tagged-sites (STSs) represent polymerase chain reaction (PCR) tags for unique sequences from known chromosomal locations (Olson et al. 1989). In wheat (*Triticum aestivum*), STS markers have been developed by converting mapped restriction fragment length polymorphism (RFLP) clones (Talbert et al. 1994). Demonstrated applications of these assays have included their use as tools for intergeneric gene transfer (Blake et al. 1996) and for the prediction of progeny variance (Burkhamer et al. 1998). The transferability of STSs between closely related species is variable (Erpelding et al. 1996), particularly with respect to sequence-tagged-microsatellite-sites (Whitton et al. 1997; Peakall et al. 1998). We have described elsewhere how PCR primers can be designed from wheat consensus sequences in order to specifically amplify each of the relevant orthologous loci of hexaploid wheat (Van Campenhout and Volckaert 1997; Van Campenhout et al. 1998). These consensus primers are directed to 'conserved' genic regions, and since it is clear that, at the level of hybridisation, cDNA sequences are generally conserved across a wide species range, we sought to test the generality of the consensus primer concept across the plant kingdom.

Of particular interest in wheat improvement are the many related species which can be sexually hybridised with wheat and which are therefore regarded as resources for alleles absent in the cultivated wheat gene pool. Among these species, rye (*Secale cereale*), offers particular potential to increase the genetic variability and to introduce desirable characters for wheat improvement. Rye chromatin can be introgressed into wheat by the induction of homoeologous chromosome pairing, as demonstrated by Koebner and Shepherd (1985, 1986). Because only low levels of homoeologous recombination can be

induced, the screening of large populations is necessary to identify rare recombinants. As a result, rye locus-specific PCR markers, targeted at proximal and distal regions of the relevant target chromosome, are a desirable tool to enable the efficient selection of rare recombinants.

The Em genes, which encode hydrophilic proteins that may play a role in water-stress tolerance during seed desiccation and early germination (McCubbin et al. 1985; Morris et al. 1990), are a good model to test consensus primer transferability since (1) the gene family is present in a wide range of plant species, including *Arabidopsis thaliana* (Gaubier et al. 1993) and rice (Litts et al. 1992), the models for, respectively, dicotyledonous and monocotyledonous plants; (2) it is a small gene family offering several targets, thereby improving the probability of detecting an Em gene sequence; (3) size polymorphisms are known to occur among Em genes, both as a result of the presence of an intron of variable length (Gaubier et al. 1993; Van Campenhout and Volckaert 1997; Calvo et al. 1997) and due to polymorphism with respect to the number of hydrophilic repeats (Espelund et al. 1992; Gaubier et al. 1993; Calvo et al. 1997); (4) as the Em family members are clustered at one locus (*Em-1*) mapped adjacent to the centromere on the long arms of the homoeologous group-1 chromosomes (Dubcovsky et al. 1995; Van Deynze et al. 1995), the reliability and value of the *Em-R1*-specific assays can be tested on established recombinants between wheat chromosomes and rye chromosome 1RL (Koebner and Shepherd 1985).

In the work described here, we tested two wheat consensus primer sets, directed to the Em gene family, on a range of diverse monocotyledonous and dicotyledonous plant species. Detailed analysis by cloning and sequencing was carried out in rye, given our interest in the development of PCR assays for the identification and characterisation of wheat-rye recombinants. Seven rye-specific primer sets were developed, and their specificity and polymorphicity were assayed on templates representing several rye sources, including varieties and addition, substitution, translocation and introgression lines.

Materials and methods

Plant material

The test array for the consensus primers included DNA extracted from 12 Triticeae grass species, three grasses belonging to other tribes, two more distantly related monocot species and seven dicot species. These were, respectively: wheat (*Triticum aestivum*), rye (*Secale cereale*), barley (*Hordeum vulgare*), oats (*Avena sativa*), *Aegilops umbellulata*, *Agropyron repens*, *Dasypyrum villosum*, *Elymus giganteus*, *Hordeum chilense*, *Pseudoroegneria strigosa*, *Psathyrostachis stoloniformis*, *Thinopyrum bessarabicum*; maize (*Zea mays*), sorghum (*Sorghum bicolor*), rice (*Oryza sativa*); banana (*Musa* - AAB group), leek (*Allium porrum*); carrot (*Daucus carota*), chicory (*Cichorium intybus*), *Arabidopsis thaliana*, beet (*Beta vulgaris*), cucumber (*Cucumis sativus*), pea (*Pisum sativum*) and tomato (*Lycopersicon esculentum*).

For the detailed analysis of *Em-R1*, five distinct sources of rye chromosome 1R were assayed. These consisted of: (1) an individual of the inbred rye variety 'Imperial'; (2) the 'Chinese Spring' (CS)/'Imperial' rye addition line 1R (Driscoll and Sears 1971) and some derivatives, including the wheat-rye recombinants with proximal (RL-2, RL-3 and RL-6) and distal (R1, R2, R3, R4, R5, R7 and R8) 1RL chromatin (Koebner and Shepherd 1985), and the substitution line 1R(1B), in which 1R replaces wheat chromosome 1B (Shepherd 1973); (3) individuals of cross-pollinated rye cultivars 'Petkus' and 'King II'; (4) wheat cultivars 'Neuzucht' and 'Salzmünder Bartweizen' (SB), in which chromosome 1B is replaced by 1R originating from a 'Petkus' population (Zeller 1973); (5) An F_3 individual from a cross between the two experimental Polish inbred rye lines DS2 and RXL. Non-rye carrying wheat DNA was obtained from cvs 'Cadenza', 'Cappelle-Desprez', 'Hobbit', CS and its group-1 nulli-tetrasomic stocks. Genomic DNA was isolated from fresh leaves of single plants according to Sharp et al. (1988) or Van Campenhout et al. (1995).

PCR, cloning and sequencing

Two consensus primer combinations for the wheat Em gene family, EMC5/EMC3 and EMC5/EMCO31 (Van Campenhout and Volckaert 1997), were used to test for transferability across the range of templates. Sequence-specific primers were designed from multiple alignment comparisons (Genmon v4.3, GBF, Braunschweig, Germany). PCR reactions were conducted as described (Van Campenhout and Volckaert 1997; Van Campenhout et al. 1998) and were electrophoretically analysed on 1% agarose gels (Biozym, Landgraaf, The Netherlands).

Consensus PCR products from rye templates were reamplified to generate blunt-end PCR products in 50-µl reactions consisting of commercially supplied buffer, 200 µ*M* of each dNTP, 1 µ*M* of each primer, 2 U *Pfu* DNA polymerase polymerase (Stratagene, La Jolla, Calif.) and one-tenth of the original PCR product as template. After 30 cycles of 94°C for 1 min, 53°C for 1 min and 72°C for 90 s, PCR products were purified with the Qiaquick PCR purification kit (Qiagen, Hilden, Germany), phosphorylated by T4 polynucleotide kinase (Eurogentec, Seraing, Belgium) in the presence of 1 m*M* rATP and repurified. For plasmid ligation, 100 ng PCR product, 100 ng of *Sma*I-cut and dephosphorylated pUC18 vector (Pharmacia Biotech, Uppsala, Sweden), 1 m*M* rATP and 2 U of T4 DNA ligase in the appropriate buffer (Promega, Madison, Wis.) were incubated in a volume of 10 µl overnight at 22°C and transformed into *E. coli* strain DH5α by a modified calcium chloride procedure (Volckaert 1987). Putative positive clones (blue/white selection) were screened for the presence of inserted DNA of the expected length by colony PCR following the protocol of Verhasselt et al. (1993). Plasmid DNA from positive clones was prepared using the QIAprep Spin Plasmid kit (Qiagen), and inserts were sequenced with the ABI PRISM Big Dye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, Calif.) using vector-specific primers. Direct sequencing of PCR products was performed with the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) using Qiaquick purified amplification products as the template. Sequence analysis was done on a 377 DNA sequencer (Applied Biosystems).

Results

Transferability of consensus primers

PCR amplifications using primer set EMC5/EMC3 [designed to amplify the entire coding region with its intron and part of the 3' untranslated region (3' UTR)] resulted in clear, specific products from all the Triticeae templates (wheat, rye, barley, oat, *Aegilops umbellulata*, *Agropyron repens*, *Dasypyrum villosum*, *Elymus giganteus*, *Hordeum chilense*, *Pseudoroegneria strigosa*,

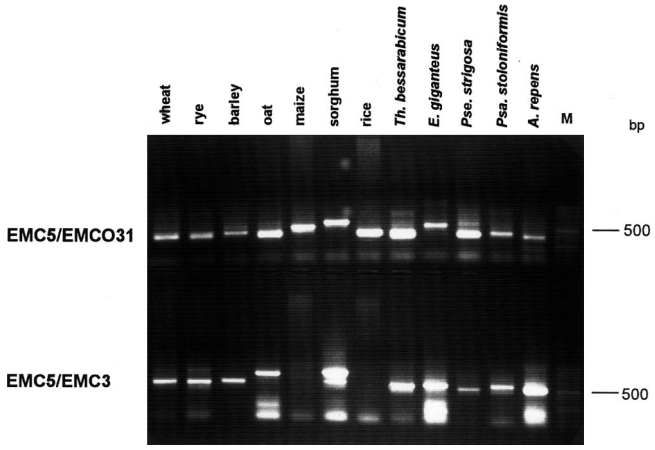


Fig. 1 Amplification with consensus primer combinations EMC5/EMCO31 (*top*) and EMC5/EMC3 (*bottom*) of sample grass species. The first amplifies the coding region and the intron of Em gene sequences, giving products in the range of 350–550 bp. The latter amplifies the coding region, the intron and part of the 3' UTR of Em gene sequences, resulting in amplicons of size 490–750 bp. The 500-bp band of the size marker (Biozym Low Ladder) is indicated

Psathyrostachis stoloniformis, *Thinopyrum bessarabicum*), and from sorghum (Fig. 1). A weak, lower molecular-weight product was formed from banana, while there was no amplification from templates of maize, rice, leek or any of the dicot species. The other primer set, EMC5/EMCO31 (directed to the coding region and its intron only), amplified successfully from all the grass species but formed no visible product from any of the other templates.

All the PCR products were directly sequenced and this confirmed that Em gene sequences had been amplified in the grass species. The EMC5/EMC3-directed product from banana proved to be not a member of the Em gene family, as its sequence (outside of the primers) did not show any similarity with any one of the known Em gene sequences. Direct sequencing of the EMC5/EMC3 amplification products gave a complete, clean read for barley (completely homologous to that of barley cDNA sequence B19.1, Espelund et al. 1992), *H. chilense* (93% similarity to the wheat EM1B01 gene, Van Campenhout and Volckaert 1997), *Pse. strigosa* (87% similarity to EM1B01), *Psa. stoloniformis* (86%

Table 1 DNA sequences isolated by consensus primer set EMC from different rye sources and their structural features. Nucleotide sequences are deposited under EMBL accession numbers AJ011945–AJ011954

Sequence	Source	Total size PCR product	Size of exons	Size of intron	Size of 3′ UTR
EMRYE30	1 clone 1R(1B)CS 5 clones Salzmünder 1 clone KingII	590 bp	282 (118+164) bp	167 bp	137 bp
EMREC04	1 clone RL3	591 bp	282 (118+164) bp	167 bp	138 bp
EMDSR51	13 clones DS2×RXL	540 bp	282 (118+164) bp	117 bp	137 bp
EMPET77	1 clone Petkus	540 bp	282 (118+164) bp	117 bp	137 bp
EMRYE92	6 clones Petkus 1 clone DS2×RXL	540 bp	282 (118+164) bp	117 bp	137 bp
EMKII78	1 clone KingII	538 bp	282 (118+164) bp	85 bp	167 bp
EMKII77	1 clone KingII	538 bp	282 (118+164) bp	85 bp	167 bp
EMDSR24	1 clone DS2×RXL	509 bp	273 (109+164) bp	93 bp	139 bp
EMREC19	1 clone RL3	525 bp	282 (118+164) bp	101 bp	138 bp
EMKII91	1 clone KingII	509 bp	282 (118+164) bp	86 bp	137 bp

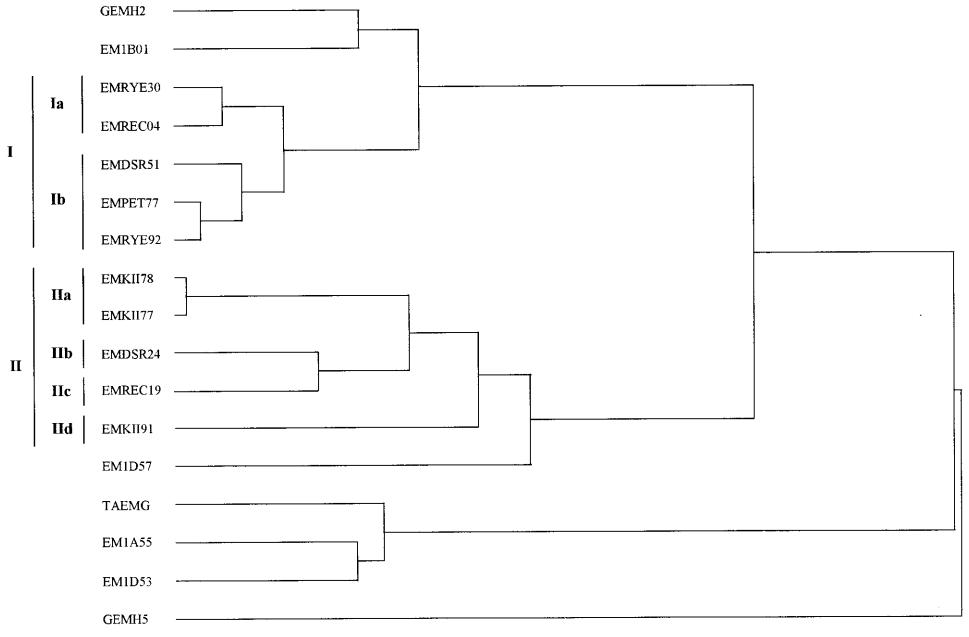


Fig. 2 Dendrogram showing inter-sequence relationships among the ten isolated *Em-R1* sequences and their relationship to the seven known wheat *Em-1* sequences as derived from the multiple DNA sequence alignment by the Genmon computer package. The wheat sequences GEMH2 and GEMH5 were previously published by Futers et al. (1993), TAEMG by Litts et al. (1991) and EM1B01, EM1D57, EM1A55 and EM1D53 by Van Campenhout and Volckaert (1997)

Fig. 3 A, B Multiple alignment of parts of the ten rye *Em-R1* DNA sequences and the related wheat *Em-1* sequences GEMH2, EM1B01 and EM1D57 (see Fig. 1). Identical positions are indicated by *asterisks*, positions marked by + are medium-conserved and *dots* represent aligning gaps. **A** Intron region of the Em genes. Primers EMROG52, EMROG51, EMRYE52 and EMRYE31 (Table 2) are *underlined*, intron borders are indicated and tandem repeats are *underscored* by *dotted lines*. **B** 3' UTR region. The termination triplet is indicated by ◆, the conserved polyadenylation signal is denoted by *PA* and primer annealing sites for EMROG31, EMROG36, EMRYE32 (Table 2) are *underlined*. The position of consensus primer EMC3 is indicated by an *arrow*

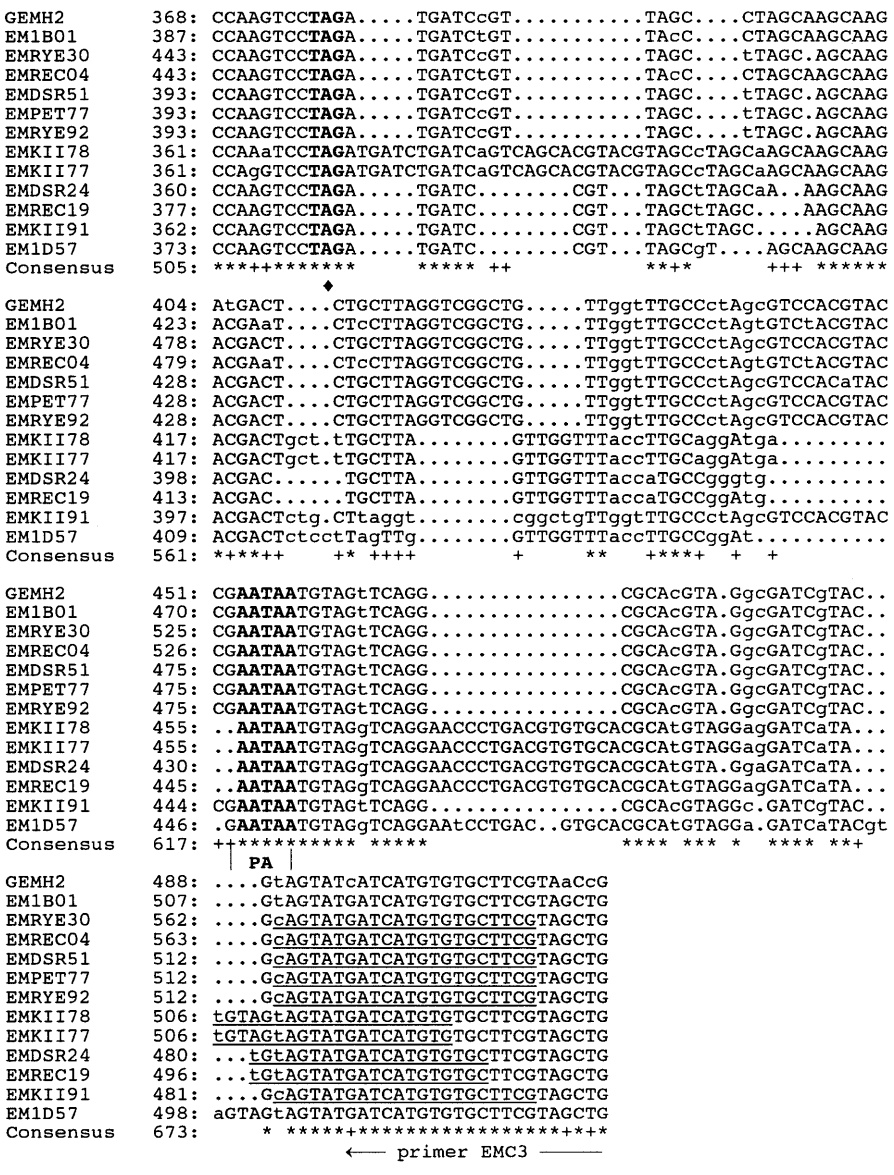
A

```
GEMH2       113: CTCGCCGAAGGTACGT.....................................
EM1B01      113: CTCGCCGAAGGTgCGTACG....GGCA..........................
EMRYE30     113: CTCGCCGAAGGTACGTACGTaCGGGCATGATgcatGCgtgagAaatGccgttgtAc
EMREC04     113: CTCGCCGAAGGTACGTACGTaCGGGCATGATgcatGCgtgagAaatGccgttgtAc
EMDSR51     113: CTCGCCGAAGGTACGTACGTaCGGGCATG.........................
EMPET77     113: CTCGCCGAAGGTACGTACGTgCGGGCATG.........................
EMRYE92     113: CTCGCCGAAGGTACGTACGTgCGGGCATG.........................
EMKII78     113: CTCGCCGAAGGTACGTA..................GCAAcgcA.......CGTCAA
EMKII77     113: CTCGCCGAAGGTACGTA..................GCAAcgcA.......CGTCAA
EMDSR24     104: CTCGCCGAAGGTACGTA..................aCAAcgcA.......CGTCgA
EMREC19     113: CTCGCCGAAGGTACGTA..................GCAAgcacgtcGagtCGTCgA
EMKII91     113: CTCGCCGAAGGTACGTA..................GCAAcgcA.......CGTCAA
EM1D57      113: CTCGCCGAAGGTgCGTAacgcaaccgAgaATCGT....................
Consensus   113: **************++++                  +
                          |→intron
GEMH2       128: .............GCATgcGtGacaGaa..........................
EM1B01      136: tgatg......ATGCATgc..GTgaGaa..........................
EMRYE30     169: taTac......ATaCAT..GcGTgaGaa..........................
EMREC04     169: taTac......ATaCAT..GcGTgaGaa..........................
EMDSR51     142: ...........ATGCAT..GcGTgaGaa..........................
EMPET77     142: ...........ATGCAT..GcGTgaGaa..........................
EMRYE92     142: ...........ATGCAT..GcGTgaGaa..........................
EMKII78     144: c.......GTTAT......GtGTtgGtgATGG..TGGTGGcGGCTTCAGCTACTCA
EMKII77     144: c.......GTTAT......GtGTtgGtgATGG..TGGTGGcGGCTTCAGCTACTCA
EMDSR24     135: ccTtatcaGTTATGtgT......tgGtgATGG..TGGTGGtGGCTTCAGCTACTCA
EMREC19     151: cc.......TTATGtgTtgGtGatgGtggTGGacTGGTGGtGGCTTCAGCTACTCA
EMKII91     144: c.......GT......TatGtGTtgGtgATGG..TGGTGGcGGCTTCAGCTACTCA
EM1D57      147: ............tgTgcGtGTtgG........TGaTGGtGGCTTCAGCTACTCA
Consensus   169:             + +++  + ++  *

GEMH2       144: ........................AtGtCGTTGTACT....ATACT........
EM1B01      156: ........................AtGtCGTcGTACT....ATACT........
EMRYE30     189: ........................AtGcCGTTGTACTATACATACTTGAGCGAT
EMREC04     189: ........................AtGcCGTTGTACTATACATACTTGAGCGAT
EMDSR51     157: ........................AtGGCGTTGTACTATACATACT........
EMPET77     157: ........................AtGGCGTTGTACTATACATACT........
EMRYE92     157: ........................AtGGCGTTGTACTATACATACT........
EMKII78     185: TGTCg..............TGTCcAcGG.........................
EMKII77     185: TGTCg..............TGTCcAcGG.........................
EMDSR24     183: TGTCaA.............TGTCtAcGG.........................
EMREC19     200: TGTCaA.............TGTCtAcGG.........................
EMKII91     185: TGTCaA.............TGTCtAcGG.........................
EM1D57      181: TtTCgtcatgtatatgtgttgTGTacAcGG.........................
Consensus   225:                          * *+

GEMH2       162: ..........cGAGCGATCTGTGGCTT.CcaatACtGAtTGCTAgtATT.......
EM1B01      174: ..........TGAGCGATCTGTGaCTT.CcgatgCtGAtTGCTAgtATTGcTGTgT
EMRYE30     219: CTGTGGCTTCTGAGCGATCTGTGGCTTcCgaTAACtGACTGCTACCATTGaTGTAT
EMREC04     219: CTGTGGCTTCTGAGCGATCTGTGGCTTcCgaTAACtGACTGCTACCATTGaTGTAT
EMDSR51     179: ..........TGAGCGATCTGTGGCTTcCgaTAACaGACTGCTACCATTGccGTAT
EMPET77     179: ..........TGAGCGATCTGTGGCTTgCagTAACaGACTGCTACCATTGcTGTAT
EMRYE92     179: ..........TGAGCGATCTGTGGCTTgCagTAACaGACTGCTACCATTGcTGTAT
EMKII78     199: ..........................................................
EMKII77     199: ..........................................................
EMDSR24     198: ..........................................................
EMREC19     215: ..........................................................
EMKII91     200: ..........................................................
EM1D57      211: ..........................................................
Consensus   281:

GEMH2       200: GCTGtATtTgtGCAGGGCGCAGCCGtGGCGGGCAGACtCGCAgGGAGCAGATGGGg
EM1B01      219: GCTGtATtTgtGCAGGGCGCAGCCGtGGCGGGCAGACtCGCAgGGAGCAGATGGGg
EMRYE30     275: GCTGCATtTgtGCAGGGCGCAGCCGCGGtGGGCAGACaCGCAgGGAGCAGATGGGt
EMREC04     275: GCTGCATtTgtGCAGGGCGCAGCCGCGGtGGGCAGACaCGCAgGGAGCAGATGGGt
EMDSR51     225: GCTGCATtTgtGCAGGGCGCAGCCGCGGCGGGCAGACaCGCAgGGAGCAGATGGGt
EMPET77     225: GCTGCATtTgtGCAGGGCGgAGCCGCGGCGGGCAGACaCGCAgGGAGCAGATGGGt
EMRYE92     225: GCTGCATtTgtGCAGGGCGgAGCCGCGGCGGGCAGACaCGCAgGGAGCAGATGGGt
EMKII78     199: ......TgTtgGCAGGGCGCAGCCGCGGCGGGCAGACgCGCAaGGAGCAGATGGGg
EMKII77     199: ......TgTtgGCAGGGCGCAGCCGCGGCGGGCAGACgCGCAaGGAGCAGATGGGg
EMDSR24     198: ......TgTtgGCAGGGCGCAGCCGCGGCGGGCAGACgCGCAaGGAGCAGATGGGg
EMREC19     215: ......TgTtgGCAGGGCGCAGCCGCGGCGGGCAGACgCGCAaGGAGCAGATGGGg
EMKII91     200: ......TgTtgGCAGGGCGCAGCCGCGGCGGGCAGACgCGCAaGGAGCAGATGGGg
EM1D57      211: ......TgTtgGCAGGGCGCAGCCGCGGCGGGCAGACgCGCAaGGAGCAGATGGGg
Consensus   337:       * *  ********+******+**+******** **** ************
                    intron ←|
```

similarity to the barley product) and *Th. bessarabicum* (89% similarity to the wheat GEMH2 gene, Futers et al. 1993). The size polymorphism displayed by the EMC5/EMC3 amplification product in oats and the most prominent band in sorghum (Fig. 1) was due both to the presence of an intron of, respectively, about130 bp and 150 bp, and to variation (four and three repeats) in the length of the central hydrophilic motif. The weakly amplified product of about 750 bp in wheat and rye (Fig. 1) was shown to be amplified by primer EMC3 alone and had sequence homology with retrotransposon-like elements (Van Campenhout and Volckaert 1999). The EMC5/EMCO31 amplification products consisted of a mixture of Em gene sequences, except for sorghum, rice (corresponding perfectly to the rice *Emp1* gene, Litts et al. 1992) and *Dasypyrum villosum* (96% similarity to the

Fig. 3 B Legend see page 331

B



wheat TAEMG gene, Litts et al. 1991). As for the other set, the size polymorphism by EMC5/EMCO31-directed amplification (Fig. 1) was the result of variation in the number of hydrophilic repeats (two in maize, three in sorghum and *E. giganteus*), and differences in intron sizes.

Isolation and characterisation of *Em-R1* DNA sequences

Em-R1 sequences were generated from EMC5/EMC3-directed amplicons of template from five different rye sources. The sequencing of 34 Em gene-containing clones revealed a total of ten *Em-R1* variants (Table 1). Each includes the whole coding sequence of 282 bp or 273 bp (there is a 9 bp deletion in the first exon in EMDSR24), interrupted at a conserved position by an intron of variable length (ranging from 85 bp to 167 bp), and a part of the 3' UTR (ranging from 137 bp to 167 bp). A multiple alignment of the ten sequences with seven known wheat *Em-1* sequences (Litts et al. 1991; Futers et al. 1993; Van Campenhout and Volckaert 1997) revealed inter-sequence relationships, which are presented in a dendrogram form (Fig. 2). This analysis demonstrates that EMRYE30, EMREC04, EMDSR51, EMPET77 and EMRYE92 belong to one group (group I) that is related to a wheat Em subfamily composed of sequences GEMH2 and EM1B01, whereas EMKII78, EMKII77, EMDSR24, EMREC19 and EMKII91 are grouped (group II) together with EM1D57 from wheat. The group-I sequences can be split into subgroups Ia (EMRYE30 and EMREC04) and Ib (EMDSR51, EMPET77, EMRYE92), primarily on the basis of a 50-bp difference in intron length (Table 1), the intron being composed of two tandem repeats (one of 32 bp and one of 18 bp) in both EMRYE30 and EMREC04 (Fig. 3A). The subdivi-

Table 2 Primer sets developed and analysed in this study, the targeted sequence and the deduced length of the PCR product

Primer set	Primer sequence	Target	Length of PCR product
EMGPI	EMROG52: TCGCCGAAGGTACGTACGT EMROG-31: CGAAGCACACATGATCATACTG	Group I (EMRYE30, EMREC04, EMDSR51, EMPRET77, EMRYE92)	471 bp EMRYE30 472 bp EMREC04 421 bp Group Ib
EMGPIIa	EMROG51:ACCTCGCCGAAGGTACGTAG EMROG36: CACATGATCATACTACTACA	Group IIa (EMKII77, EMKII78)	415 bp
EMGPIIb	EMRYE52: CTCGCCGAAGGTACGTAACA EMRYE32: GCACACATGATCATACTACA	EMDSR24	396 bp
EMGPIIc	EMROG51: see above EMRYE32: see above	EMREC19	405 bp
EMGPIId	EMROG51: see above EMROG31: see above	EMKII91	393 bp
EMINGPI	EMROG52: see above EMRYE31: CCCTGCACAAATGCAGCAT	Group I	178 bp Group Ia 128 bp Group Ib
EMCG52	EMROG52: see above EMC3: see above	Group I	477 bp EMRYE30 478 bp EMREC04 427 bp Group Ib
EMCG51	EMROG51: see above EMC3: see above	EMKII77, EMKII78, EMREC19, EMKII91	415 bp EMREC19 428 bp EMKII77, EMKII78 399 bp EMKII91
EMCR52	EMRYE52: see above EMC3: see above	EMDSR24	406 bp
EMCG31	EMC5: see above EMROG31: see above	Group I, EMKII91	584 bp EMRYE30 585 bp EMREC04 534 bp Group Ib 503 bp EMKII91
EMCG36	EMC5: see above EMROG36: see above	EMKII77, EMKII78	525 bp
EMCR32	EMC5: see above EMRYE32: see above	EMDSR24, EMREC19	499 bp EMDSR24 515 bp EMREC19

sion in group II (types IIa, IIb, IIc, IId; Fig. 2) is due mainly to size differences in several parts of both the introns and the 3' UTR (Table 1, Fig. 3A and B). Nucleotide differences at positions scattered throughout the sequences are responsible for a further subdivision of subgroups (Fig. 3A and B).

Development and analysis of *Em-R1*-specific PCR markers

In order to construct *Em-R1*-specific primers, we exploited sequence differences between the rye and the (known) wheat Em gene sequences. Since most of this variation is present in the introns and 3' UTRs, sequence-specific primers were designed to target sequence-specific polymorphisms in these regions. Primer design was based on the criterion that the 3'-terminal positions should be different from that present in any of the known wheat sequences (Fig. 3A and B, only related wheat sequences are shown in this comparison). In this way, six sequence-specific primers (EMROG52, EMROG51, EMRYE52, EMROG31, EMROG36 and EMRYE32) were constructed, allowing 5 primer combinations (EMGPI, EMGPIIa, EMGPIIb, EMGPIIc and EMGPIId) to direct the amplification of all the isolated *Em-R1* sequences (Table 2). These primer combinations amplified products of the expected lengths (Table 2) that were rye- (and hence *Em-R1*-) specific, as no amplification product was formed in any of the non-rye carrying wheat templates tested (Table 3). Among the rye templates tested, primer set EMGPI generated an amplification length polymorphism visualised by agarose gel electrophoresis that discriminated 'Petkus', DS2×RXL and 'Imperial' (with a PCR product of approx. 420 bp) from 'King II', 1R(1B), CS+1R, SB, 'Neuzucht' and the RL recombinants (with a PCR product of approx. 470 bp) (Table 3, illustrated in Fig. 4). The R recombinants lack the *Em-R1* locus. Amplification with primer set EMINGPI (Table 2), which targets the introns of these sequences, confirmed these results (Table 3). The *Em-R1* target sequence of primer sets EMGPI, EMGPIIb and EMGPIIc was shown to be present in every rye-template tested, while primer sets EMGPIIa and EMGPIId amplified less frequently occurring *Em-R1* alleles (Table 3).

The sequence-specific primers were also tested in combination with the appropriate consensus primer (sets EMCG52, EMCG51, EMCR52, EMCG31, EMCG36 and EMCR32; Table 2). Except for EMCR32, this resulted in the loss of rye-specificity (Table 3), presumably because of the presence of as yet unidentified Em gene members in wheat. Thus, when sequence-specific primers EMROG52 and EMROG31 were used in conjunction (set EMGPI), a polymorphic *Em-R1*-specific product was formed (discussed above). However, in the combination EMROG52/EMC3 (set EMCG52), a comparison between the amplification profiles of CS and an aneuploid stock in a CS background lacking chromosome 1D (e.g. nullisomic 1D tetrasomic1B) shows that a chromosome 1D-specific product of approx. 430 bp was ampli-

Table 3 Amplification analysis of the developed primer sets (see Table 2)

Primer set	1R(1B)	CS+1R	SB, Neuzucht	Petkus	KingII	DS2×RXL	Imperial	R1 to R8	RL2, RL3	RL6	Wheat[a]
EMGPI[b]	+[l]	+[l]	+[l]	+[s]	+[l]	+[s]	+[s]	–	+[l]	+[l]	–
EMGPIIa	–	–	–	–	+	–	–	–	–	–	–
EMGPIIb	+	+	+	+	+	+	+	–	+	+	–
EMGPIIc	+	+	+	+	+	+	+	–	+	+	–
EMGPIId	+	+	+	–	+	–	–	–	+	+	–
EMINGPI[b]	+[l]	+[l]	+[l]	+[s]	+[l]	+[s]	+[s]	–	+[l]	+[l]	–
EMCG52[b]	+[d]	+[d]	+[d]	+[s]	+[l]	+[s]	+[s]	+[s]	+[l]	+[d]	+[s] (1D-specific)
EMCG51	+	+	+	+	+	+	+	+	+	+	+ (weaker)
EMCR52	+	+	+	+	+	+	+	+	+	+	+
EMCG31[b]	+[d]	+[d]	+[d]	+[s]	+[d]	+[s]	+[s]	+[s]	+[d]	+[d]	+[s]
EMCG36	–	–	–	–	+	–	–	+	+	+	+
EMCR32	+	+	+	+	+	+	+	–	+	+	–

[a] Tested wheat varieties were: 'Cadenza', 'Cappelle-Desprez', 'Hobbit', 'Chinese Spring' and its group-1 nulli-tetrasomic lines
[b] Size polymorphism on agarose gel electrophoresis: s, small; l, large; d, double amplification bands

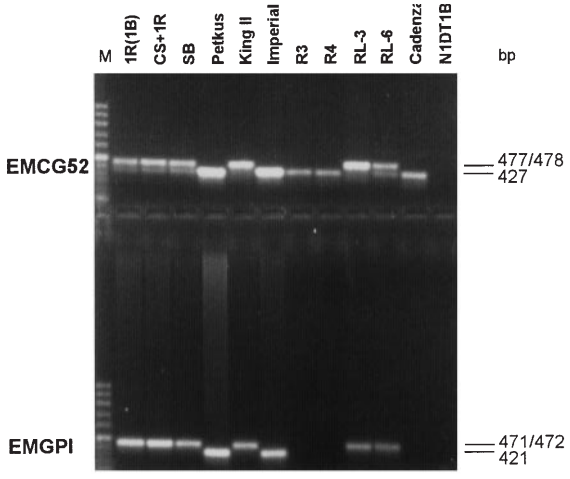


Fig. 4 PCR with primer set EMGPI (EMROG52+EMROG31) amplifies a rye-specific *Em-R1* product that shows amplification length polymorphism. Set EMCG52 (EMROG52+consensus primer EMC3) amplifies a 1D-specific product from wheat in addition to the rye-specific polymorphic *Em-R1* product. The size of the amplification products is indicated, as are the genotypes. *M* Size marker (Biozym Low Ladder)

fied in addition to the rye sequence (illustrated in Fig. 4). Differences in the amplification pattern generated by set EMCG52 and EMCG31 allow differentiation between the different rye and rye-carrying genotypes (Table 3).

Em-R1-specific PCR markers to discriminate wheat-rye recombinants

Of the 12 primer combinations (Table 3) 8 were able to discriminate among the wheat-rye recombinants involving chromosome-arm 1RL. These lines were characterised for a number of RFLP and PCR loci by Rogowsky et al. (1993), and the analysis showed that the wheat-rye recombinants RL-2, RL-3 and RL-6 have proximal rye chromatin, while R1, R2, R3, R4, R5, R7 and R8 have distal rye chromatin. Sets EMGPI, EMGPIIb, EMGPIIc, EMGPIId, EMINGPI and EMCR32, which are rye-specific, discriminated the RL from the R recombinants by presence (in RL)/absence (in R) polymorphism (Table 3, illustrated for EMGPI in Fig. 4), while EMCG31 and EMCG52 distinguished the RL from the R recombinants by the difference in the amplification lengths of the wheat and rye products (Table 3, illustrated for EMCG52 in Fig. 4). In addition, this latter set did not amplify the 1D-specific product in RL-2 and RL-3 (Fig. 4). As these lines carry *Em-R1* but lack *Em-D1*, the recombinant chromosome in RL-2 and RL-3 probably represents a homoeologous exchange between 1RL and 1DL.

Discussion

The strategy: from consensus primers to species-specific PCR markers

We have previously shown that wheat *Pur-1* consensus primers can be used to develop a PCR assay for the rye orthologue (Van Campenhout et al. 1998). The present work demonstrates that while this approach is valid across a wide spectrum of grass species and genera, it can not be generalised across the major monocot/dicot divide, presumably because of large-scale sequence divergence, which has been sufficient to have resulted in differences in codon usage between the two major plant phyla. The use of heterologous probes in comparative RFLP mapping of the grass species has revealed a remarkable level of synteny between their genomes (Devos and Gale 1997). Analogously, it is likely that consensus primers, in combination with DNA sequence analysis, can take comparative genetics to a further level, giving insight into the conservation of DNA sequences between genes of similar function. It is significant that the cross-species transferability of the primer set directed to the protein coding region (EMC5/EMCO31) reached further than that of the primer combination with the reverse primer located in the 3' UTR (EMC5/EMC3). This is consistent with the behaviour of cDNA-based STS

primers across conifer genera (Perry and Bousquet 1998). While the placement of a primer in a non-coding region reduces the species range over which primers can be transferred, it increases the probability of generating informative polymorphism by ensuring the inclusion of more variable DNA in the PCR product. Sequence differentiation between wheat and rye was indeed predominantly present in the non-coding regions of the amplicons. Intron and 3' UTR lengths are under less selective constraint than the coding region, and 4 of the 12 rye assays readily gave scorable length polymorphisms of this type. Thus, in terms of a general strategy, attention should be focussed on intron- and UTR-containing sequences. Although currently the extent of such sequences for wheat in public DNA databases is rather limited, the substantial level of DNA sequence conservation among the grasses in the coding regions (as inferred from the generality of the Em primers and the cross-hybridisation of orthologous cDNAs), makes it feasible that consensus primers should be derivable from sequences of those cereal species (especially rice) that are becoming the target of a substantial DNA sequencing effort.

Heterogeneity between and within rye varieties

In the investigation reported here the consensus primer approach was, for the first time, elaborated on a polymorphic crop. Detailed analysis resulted in the characterisation of ten different Em gene sequences from five different sources of rye, and the subsequent tests with the derived sequence-specific primer sets enabled the detection of allelic variation. In some cases, the extent of allelic DNA sequence variation in rye was such that the specific primer set constructed for a particular sequence from one variety was genotype-specific and, consequently was not generally transportable to rye in general. Examples of this are sequences EMKII77 and EMKII78, whose specific primer set (EMGPIIa) only amplified template from 'KingII', while set EMGPIId, directed to sequence EMKII91, failed to amplify template from either 'Petkus', DS2×RXL or Imperial (Table 3). Moreover, the amplification difference between SB and Neuzucht on one hand and the 'Petkus' individual on the other with, for example, sets EMGPI, EMGPIId and EMINGPI (Table 3, Fig. 3) reflects heterogeneity in the open-pollinated 'Petkus' population from which these 1R(1B) substitutions were derived. The differences in allelic constitution (as shown by sets EMGPI, EMGPIId, EMINGPI, EMCG52 and EMCG31; Table 3, Fig. 3) between the sample of 'Imperial' and the addition, substitution and recombinant lines (in which the rye chromatin originates from a single isolation of CS addition line 1R 'Imperial') demonstrates that 'Imperial' rye, despite its derivation by enforced inbreeding, is also heterogeneous. This heterogeneity within 'Imperial' has been noted by Shepherd (1971), who observed a similar inconsistency between the seed storage protein profiles of the CS/'Imperial' amphiploid (from which the addition lines were generated) and those of 'Imperial' rye itself. This multiple allelism exhibited in rye suggests that allele-displaying PCR assays can realistically be developed in rye (and, by implication, in any species with high levels of polymorphism). While this polymorphism is an advantage for the genome analysis of rye, it also represents a drawback for the development of introgression-specific markers. Indeed, for the latter application, sequences must be isolated from the genetic background in which they will be employed so that upon applying the consensus primer strategy on the addition lines (from which the recombinats are generated) the rye sequence has to be identified among interfering wheat sequences (unless the original rye donor individual is known).

Selection markers for recombinants

Combinations between sequence-specific and consensus primers often amplified an unexpected product in wheat, suggesting a close similarity in the organisation of the Em gene family in wheat and rye. Hence, besides allelic variation, and especially for gene families, unknown wheat gene members can complicate the development of rye-specific PCR markers that are to be used in a wheat background. Nevertheless, 8 primer sets were appropriate for the analysis of current 1RL recombinants. These would be suitable as a tool to select further recombinants more effectively than has previously been possible. Indeed, the *Em-R1*-primers are the most proximal STS-PCR markers available for 1RL, mapping closer to the centromere than the *Glu-R1* storage protein gene (Van Deynze et al. 1995) that was used as proximal marker by Koebner and Shepherd (1985). Although the rye chromosome arm does not carry any known genes of agronomic interest, the development of a fuller set of recombinants would provide an excellent mapping tool by which to assign markers to specific rye and wheat chromosome regions and, by inference, to the equivalent regions in related grass genomes. Proximal rye/distal wheat recombinants would be of particular benefit, as these would complement the *Ae. cylindrica*-derived terminal wheat chromosome deletion lines (Endo 1988). Finally, the strategy demonstrated in this paper can serve as a model of what is becoming possible in the light of our present understanding of syntenic relationships across wide botanical frontiers. For example, a functional rice sequence mapped in rice can be converted with some confidence to generate a PCR marker for the equivalent region of an *Aegilops* chromosome that is the target for introgression into wheat.

Acknowledgements We thank Dr. K.W. Shepherd, Dr. P. Langridge and Mr. G. De Wever for gifts of plant material. This work was financed in part by a European Commission Research Training Grant Biotechnology "category 20 Fellowship – BIOCT965039". S. Van Campenhout acknowledges the "Onderzoeksfonds KULeuven" for a postdoctoral fellowship.

References

Blake TK, Kadyrzhanova D, Shepherd KW, Islam AKMR, Langridge PL, McDonald CL, Erpelding J, Larson S, Blake NK, Talbert LE (1996) STS-PCR markers appropriate for wheat-barley introgression. Theor Appl Genet 93: 826–832

Burkhamer RL, Lanning SP, Martens RJ, Martin JM, Talbert LE (1998) Predicting progeny variance from parental divergence in hard red spring wheat. Crop Sci 38:243–248

Calvo ES, Wurtle ES, Shoemaker RC (1997) Cloning, mapping, and analyses of expression of the *Em*-like gene family in soybean [*Glycine max* (L). Merr.]. Theor Appl Genet 94: 957–967

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol 35:3–15

Driscoll CJ, Sears ER (1971) Individual addition of the chromosomes of 'Imperial' rye to wheat. Agron Abstr 1971: 6

Dubcosky J, Luo M-C, Dvorák J (1995) Linkage relationships among stress-induced genes in wheat. Theor Appl Genet 91: 795–801

Endo TR (1988) Induction of chromosomal structural changes by a chromosome of *Aegilops cylindrica* L. in common wheat. J Hered 79: 366–370

Erpelding JE, Blake NK, Blake TK, Talbert LE (1996) Transfer of sequence tagged site PCR markers between wheat and barley. Genome 39: 802–810

Espelund M, Saeboe-Larssen S, Hughes DW, Galau GA, Larsen F, Jacobsen KS (1992) Late embryogenesis-abundant genes encoding proteins with different numbers of hydrophilic repeats are regulated differentially by abscisic acid and osmotic stress. Plant J 2: 241–252

Futers TS, Onde S, Turet M, Cuming AC (1993) Sequence analysis of two tandemly linked Em genes from wheat. Plant Mol Biol 23: 1067–1072

Gaubier P, Raynal M, Hull G, Huestis GM, Grellet F, Arenas C, Pagès M, Delseny M (1993) Two different *Em*-like genes are expressed in *Arabidopsis thaliana* seeds during maturation. Mol Gen Genet 238: 409–418

Koebner RMD, Shepherd KW (1985) Induction of recombination between rye chromosome 1RL and wheat chromosomes. Theor Appl Genet 71: 208–215

Koebner RMD, Shepherd KW (1986) Controlled introgression to wheat of genes from rye chromosome arm 1RS by induction of allosyndesis. 1. Isolation of recombinants. Theor Appl Genet 73:197–208

Litts JC, Colwell GW, Chakerian RL, Quatrano RS (1991) Sequence analysis of a functional member of the Em gene family from wheat. J DNA Sequence 1: 263–274

Litts JC, Erdman MB, Huang N, Karrer EE, Noueiry A, Quatrano RS, Rodriguez RL (1992) Nucleotide sequence of the rice (*Oryza sativa*) Em protein gene (*Emp1*). Plant Mol Biol 19: 335–337

McCubbin WD, Kay CM, Lane BG (1985). Hydrodynamic and optical properties of the wheat germ Em protein. Can J Biochem Cell Biol 63: 803–811

Morris PC, Kumar A, Bowles DJ, Cuming AC (1990) Osmotic stress and abscisic acid induce expression of the wheat Em genes. Eur J Biochem 190: 625–630

Olson M, Hood L, Cantor C, Botstein D (1989) A common language for physical mapping of the human genome. Science 245: 1434–1435

Peakall R, Gilmore S, Keys W, Morgante 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

Perry DJ, Bousquet J (1998) Sequence-tagged-site (STS) markers of arbitrary genes: the utility of black spruce-derived STS primers in other conifers. Theor Appl Genet 97:735–743

Rogowsky PM, Sorrels ME, Shepherd KW, Langridge P (1993) Characterisation of wheat-rye recombinants with RFLP and PCR probes. Theor Appl Genet 85: 1023–1028

Sharp PJ, Kreis M, Shewry PR, Gale MD (1988) Location of β-amylase sequences in wheat and its relatives. Theor Appl Genet 75: 286–290

Shepherd KW (1971) Genetic control of rye endosperm proteins. Experientia 27: 98–99

Shepherd KW (1973) Homoeology of wheat and alien chromosomes controlling endosperm protein phenotypes. In: Sears ER, Sears LMS (eds) Proc 4th Int Wheat Genet Symp., University of Missouri, Columbia, Mo., pp 745–760

Talbert LE, Blake NK, Chee PW, Blake TK, Magyar GM (1994) Evaluation of "sequence-tagged-site" PCR products as molecular markers in wheat. Theor Appl Genet 87:789–794

Van Campenhout S, Volckaert G (1997) PCR-based isolation and chromosome assignment of members of the Em gene family from wheat. J DNA Sequence 7: 289–300

Van Campenhout S, Volckaert G (1999) The fortuitous cloning of retroelement-like sequences from wheat and rye as by-products of a specific polymerase chain reaction. J DNA Sequence (in press)

Van Campenhout S, Vander Stappen J, Sági L, Volckaert G (1995) Locus-specific primers for LMW glutenin genes on each of the group 1 chromosomes of hexaploid wheat. Theor Appl Genet 91: 313–319

Van Campenhout S, Sági L, Vander Stappen J, Volckaert G (1998) Characterisation of type-I thionin loci from the A, B, D and R genomes of wheat and rye. Theor Appl Genet 96: 80–86

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

Verhasselt P, Van Campenhout S, Voet M, Wellens B, Volckaert G (1993) A general method for routine sequencing of cloned DNA fragments using commercial dye primers and its application in sequence analysis of *Toxoplasma gondii* RH genome fragments. J DNA Sequence 4: 71–77

Volckaert G (1987) A systematic approach to chemical DNA sequencing by subcloning in pGV451 and derived vectors. Methods Enzymol 155: 321–250

Whitton J, Rieseberg LH, Ungerer MC (1997) Microsatellite loci are not conserved across the Asteraceae. Mol Biol Evol 14: 204–209

Zeller FJ (1973) 1B/1R wheat-rye substitutions and translocations. In: Sears ER, Sears LMS (eds) Proc 4th Int Wheat Genet Symp. University of Missouri, Columbia Mo., pp 209–221