Theor Appl Genet (2000) 100:789–793

Mustapha Benmoussa · Louis-P. Vézina · Michel Pagé
Serge Yelle · Serge Laberge

Genetic polymorphism in low-molecular-weight glutenin genes from *Triticum aestivum*, variety Chinese Spring

Received: 7 May 1999 / Accepted: 17 June 1999

Abstract Low-molecular-weight (LMW) glutenin subunits consist mainly of two domains, one at the N-terminus which contains repeats of short amino-acid motifs, and a non-repetitive one rich in cysteine, at the C-terminal region. In previous reports, polyacrylamide-gel electrophoresis has been used to show that large size variation exists among LMW and HMW glutenin subunits, and it has been suggested that deletions and insertions within the repetitive region are responsible for these variations in length. In this study, PCR-amplification of genomic DNA (*Triticum aestivum* variety Chinese Spring) was used to isolate three full-length LMW glutenin genes: LMWG-MB1, LMWG-MB2 and LMWG-MB3. The deduced amino-acid sequences show a high similarity between these ORFs, and with those of other LMW glutenin genes. Comparisons indicate that LMWG-MB1 has probably lost a 12-bp fragment through deletion and that LMWG-MB1 and LMWG-MB2 have an insertion of 81 bp within the repetitive domain. The current study has shown direct evidence that insertions and/or deletions provide a mechanistic explanation for the allelic variation, and the resultant evolution, of prolamin genes. Single-base substitutions at identical sites generate stop codons in both LMWG-MB2 and LMWG-MB3 indicating that these clones are pseudogenes.

Key words Glutenin · *Triticum aestivum* · Wheat storage proteins

Communicated by G. Wenzel

M. Benmoussa · S. Yelle
Centre de recherche en horticulture, Département de phytologie, FSAA, Université Laval, Québec G1K 7P4, Canada

L.-P. Vézina, (✉) · S. Laberge
Agriculture et Agroalimentaire Canada, 2560 boul. Hochelage, Ste-Foy Québec G1V 2J3, Canada
e-mail: laberges@em.agr.ca
Tel.: +1-(418) 657-7980, Fax: +1-(418) 648-2402

M. Pagé
Laboratoire de cancérologie, Département de biochimie, Faculté de médecine, Université Laval, Québec G1K 7P4, Canada

Introduction

The endosperm is the main tissue in the wheat seed. It contains the majority of the seed-storage proteins, which are in the form of prolamins. These proteins are synthesized during seed maturation and act as a source of amino acids for the germinating embryo and subsequent seedling growth. The prolamins have been classified into two major groups, glutenins and gliadins, based on their solubility in aqueous solvents (Osborne 1924). In the endosperm cells, glutenins aggregate through the formation of intra- and inter-molecular bonds (disulfide and other non-covalent bonds) to form large insoluble complexes. The reduction of these complexes and their separation on SDS-PAGE indicate that glutenins are composed of two groups of high- and low-molecular-weight subunits respectively, which are referred to as HMW and LMW subunits (Payne 1987).

These storage proteins play a key role in the physical behaviour of flour during bread-making. Several studies (Gupta et al. 1991; Khelifi and Branlard 1992) have indicated that predictive models for dough properties should include the relative composition in both low- and high-molecular-weight glutenin subunits. However, genetic correlations between dough quality and LMW glutenin abundance have been very difficult to establish in hexaploid wheat (Gupta et al. 1989).

Polyacrylamide-gel electrophoresis (SDS-PAGE) has shown a wide polymorphism in each class of wheat storage protein genes (Ellen et al. 1992; Sissons et al. 1998). The molecular weight of HMW subunits ranges between 70 and 120 kDa. The low-molecular-weight LMW glutenin subunits are divided into two groups (B and C) with an apparent molecular weight in the range of 75–40 and 36–30 kDa, respectively. The gliadins can be classified on the basis of electrophoretic mobility at low pH into alpha-, beta-, gamma- and omega-gliadin families. Similarities between these prolamins suggest that they have evolved through mutational events from an ancestral gene (Shewry et al. 1995). Punctual mutations are responsible for some variations in amino sequence with both repetitive and non-

repetitive domains (Anderson and Greene 1997). However, small insertions and deletions are also often present in the repetitive domain. Unequal crossing-over or slippage during replication are probably responsible for mutational events in the evolution of prolamins (Shewry et al. 1989). Direct evidence to explain large amino-acid insertions or deletions requires the characterization of a large number of storage proteins. Gupta and Shepherd (1990) reported that two-dimensional gel electrophoresis reveals the presence of about 40 different subunits of LMW glutenins among 222 cultivars of *Triticum aestivum*.

Genes encoding members of three wheat prolamin families have been isolated from both genomic and cDNA libraries (Colot et al. 1989; Anderson and Green 1989; 1999; Cassidy and Dvorak 1991), and the molecular genetics of HMW glutenin genes have been studied extensively (Shewry et al. 1982; Payne and Lawrence 1983). However, the LMW glutenin genes from groups B and C are still poorly characterized. In the current study we have isolated three new genes with ORFs homologous to available glutenin gene sequences using PCR-amplification of wheat genomic DNA. This study demonstrates that consensus sequences deduced from the 5′- and 3′-termini of glutenin genes can be used as probes to amplify new full-length encoding regions from genomic DNA. It also shows that PCR-amplification can be used to gain information on the allelic variation and evolution of the LMW gene family and the generation of prolamin pseudogenes in the wheat genome.

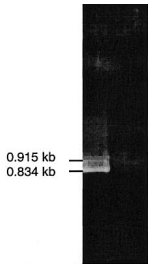

Fig. 1 Agarose gel (1%) of a PCR product from genomic DNA of *T. aestivum* var. Chinese Spring

Materials and methods

DNA extraction and PCR-amplification

Genomic DNA was isolated from young leaves of *T. aestivum* var. Chinese Spring following the procedure described by Rogers and Bendich (1988). Since the signal sequence and the carboxy terminus regions of LMW glutenins and γ-gliadins appear to be conserved (Colot et al. 1989; Cassidy and Dvorak 1991), they were used as primer sequences for PCR-amplification. Base composition was selected from the 5′ and 3′-coding regions of the LMWG-1D1 gene previously isolated by Colot et al. (1989). The sequences of the two primers were as follows: P1 (5′ATGAA GACCTTCCTCGTCTTTGC3′) spanning nucleotides +46 to +68 and encoding for the first amino acids of the signal sequence, and P2 (5′TCAGTAGGCACCAACTCCGGTGC3′) spanning nucleotides +901 to +924 and encoding for the reverse strand of the last amino acids of the protein. Three nanograms of each primer and 100 ng of genomic DNA were used for PCR-amplification in 50 µl of the following reaction mixture: 50 mM Tris-HCl, 1.75 mM MgCl2, 16 nM $(NH4)2SO_4$, pH 9.2, 0.2 mM of each dNTP and 25 units of the enzyme mix ExpandTM for long-template PCR systems (Boehringer Manheim, Germany). Conditions for PCRs were 95°C for 4 min, 35 cycles at (1) 94°C for 45 s, (2) 50°C for 45 s, (3) 68°C for 45 s and one final cycle at 94°C for 1 min and 68°C for 7 min. PCR-products were separated by electrophoresis on 1% agarose gels. Bands (0.83 kb and 0.91 kb) were excised from agarose gels and purified by the Prep-A-Gene method (Bio-Rad). The fragments were cloned into pGEM-T (Promega, USA).

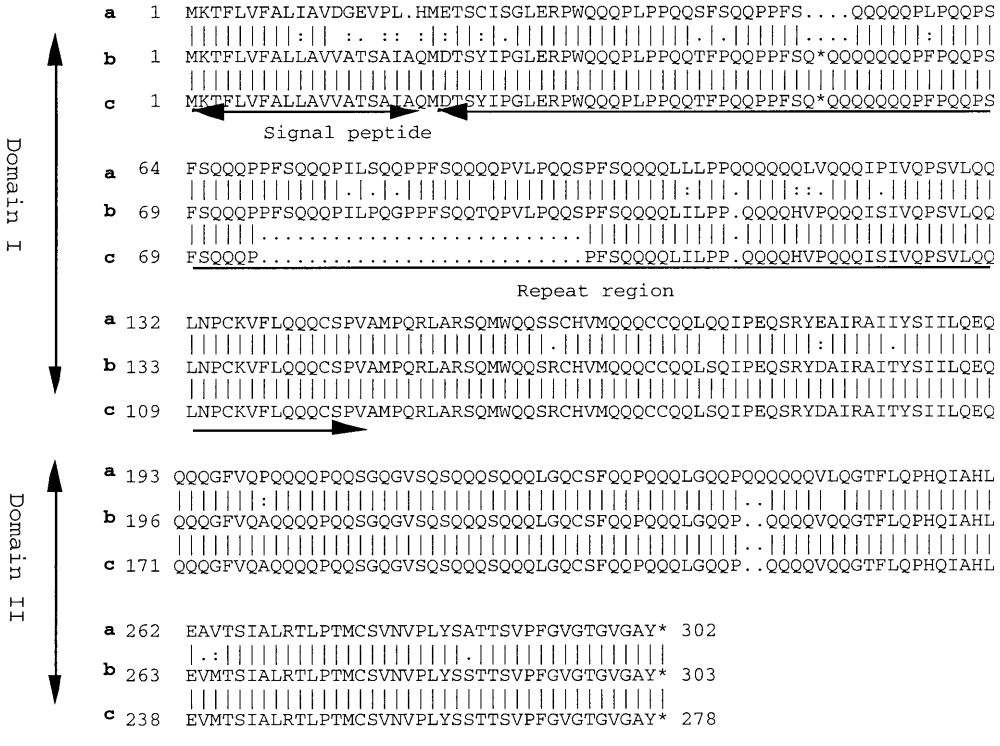

Fig. 2 Deduced amino-acid sequences of (**a**) LMWG-MB1, (**b**) LMWG-MB2 and (**c**) LMWG-MB3

Northern Hybridization

Northern-blot analysis was done using total RNA (15 μg) from seeds of *T. aestivum* var. Chinese Spring, harvested 1, 2 and 3 weeks after pollination, and from mature leaves, as described by de Vries et al. (1988). Total RNA samples were fractionated on a 1% formamide/formaldehyde agarose gel and transferred to a nylon membrane (Hybond, Amersham). The filter was pre-hybridized for 3 h and hybridization was carried overnight at 65°C in 2 × SSC, 0.25% (w/v) low-fat powder milk with a ^{32}P dCTP-labelled probe prepared from purified LMWG-MB1. The filter was washed twice at 65°C in 2× SSC for 15 min and autoradiographed at –80°C during 2 h.

Southern hybridization

PCR products were obtained by amplification at different annealing temperatures, 65°C, 55°C and 45°C. After agarose-gel size-fractionation, the PCR products were transferred to a nylon membrane (Hybond, Amersham, England) and hybridized with the LMWG-MB1 probe as described for Northern hybridizations.

Results

PCR-amplification of genomic DNA resulted in two major DNA fragments of 0.91 kb and 0.83 kb (Fig. 1). Cloning and sequencing of these fragments led to the isolation of three LMW glutenin genes, designated LMWG-MB1, LMWG-MB2 and LMWG-MB3. The LMWG-MB1, LMWG-MB2 and LMWG-MB3 ORFs are 912, 915 and 834 nucleotides long respectively, and can be translated into proteins of 302, 303 and 278 amino acids, including a putative signal peptide for each ORF (Fig. 2).

Nucleic-acid and peptide alignments, and the calculation of similarity scores were performed with version 8 of the UWGCG (University of Wisconsin Genetics Computer Group) programs (Devereux et al. 1984). The amino-acid sequences of the coding regions of LMWG-MB1, LMWG-MB2 and LMWG-MB3 were compared to the corresponding regions of published glutenin genes. As expected for LMW sequences, LMWG-MB1, 2 and 3 are composed of two major domains (I and II). The N-terminal sequence (domain I) is rich in proline and glutamine but poor in cysteine and characterized by tandem amino-acid repeats (PFSQQQ and FPQQP). The C-terminal sequence (domain II) contains a non-repetitive amino-acid sequence which is poor in proline and rich in cysteine. Comparison of amino-acid sequences from the three clones showed length variations within domain I. A small deletion of one tetrapeptide at amino-acid 81 differentiates clones LMWG-MB2 and LMWG-MB3 from LMWG-MB1. A single insertion of 27 amino acids at amino-acid 73 in the repetitive domain differentiates LMWG-MB2 from LMWG-MB3 (Fig. 2). The amino-acid sequences of LMWG-MB2 and LMWG-MB3 have an identical domain II. Nine single-base substitutions and one deletion of two amino acids are the only difference between LMWG-MB1 and LMWG-MB2 at the C-terminus.

In all LMWG-MB clones isolated in this study, the position of the cysteine residues in domain II are identical to that of the glutenin sequences published previously (Cassidy and Dvorak 1991). However the nucleotide substitution in clones LMWG-MB2 and LMWG-MB3 causes a change from cysteine to tyrosine at position 25 in domain I (Fig. 3), and consequently the number of cysteine residues is reduced to seven in LMWG-MB2 and LMWG-MB3.

```
LMWG-MB1     MKTFLVFALIAVDGEVPL.HMETSCISGLE
LMWG-MB2     MKTFLVFALLAVVATSAIAQMDTSYIPGLE
LMWG-MB3     MKTFLVFALLAVVATSAIAQMDTSYIPGLE
WHTGLI       MKTFLVFALIAVVATSAIAQMETSCISGLE
TDGLUT     MKTFLVFALLAVVATSTIAQMETSCIPGLE
TAGLUA           FALIAVVATSTIAQMETSCIPGLE
TDLMWG       MKTFLVFALLAVVATSAIAQMDTSCIPGLE
TAGLU1       MKTFLVFALLAVAATSAIAQMETRCIPGLE
PTDUCD1      MKTFLVFALLAVVATSTIAQMETSCIPGLE
```

Fig. 3 The deduced signal peptide and N-terminus amino-acid sequences of LMW glutenin and gliadin genes

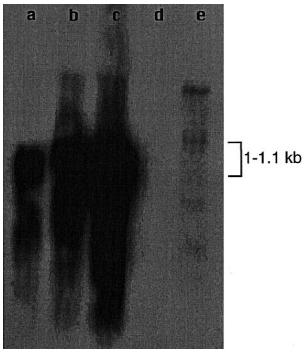

Fig. 4 Northern-blot analysis of total RNA (15 μg) from seeds *T. aestivum* var. Chinese Spring, harvested 1, 2 and 3 weeks after pollination (*lanes a, b and c*), and from the leaf (*lane d*). A molecular-weight marker is in *lane e*. Fifteen micrograms of total RNA were fractioned on a 1% formamide/formaldehyde agarose gel and transferred to a nylon membrane (Hybond, Amersham). The filter was pre-hybridized for 3 h in 2 × SSC, 0.25% (w/v) low-fat powder milk. Hybridization was carried out overnight at 65°C in 2 × SSC, 0.25% (w/v) low-fat-powder milk with a ^{32}P dCTP-labelled probe prepared from purified LMWG-MB1. The filter was washed at 65°C twice in 2 × SSC for 15 min and autoradiographed at –80°C for 2 h

Northern analysis was used to show the specificity of expression of LMWG-MB glutenin genes in seed and leaf tissues of *T. aestivum*. Total RNA was isolated from wheat seeds harvested at different stages of development (1, 2 and 3 weeks of post-anthesis) and from wheat leaves. The transcript level of LMWG-MB genes was analysed by hybridization with LMWG-MB1 as a probe (Fig. 4). LMWG mRNA was detected 1 week after anthesis, and transcript level increased during seed development. No transcript was detected in leaf extracts (Fig. 4).

Since amplification with LMW primers produced fragments with a wide size variation, the identity of the amplification products was verified by hybridization using the LMWG-MB1 glutenin gene as a probe (Fig. 5).

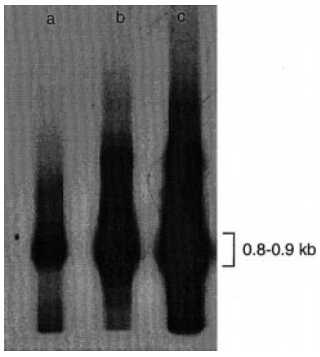

Fig. 5 Southern hybridization of PCR products at different annealing temperatures (**a**) 65°C, (**b**) 55°C and (**c**) 45°C. After agarose-gel fractionation, the PCR products were transferred to a nylon membrane (Hybond, Amersham) and hybridized with the LMWG-MB1 probe

Southern hybridization showed an extensive size polymorphism of fragments ranging from 0.6 kb to 1.7 kb (Fig. 5). The size of the most-abundant band was 1 kb, which corresponds to the size of previously reported LMW genes as well as that of LMWG-MB1,2,3.

Discussion

The aim of this study was to gain information on the LMW glutenin (LMWG) gene family in *T. aestivum*. DNA was isolated from plants of the variety Chinese Spring, which is well recognized as a source of genetic variability within the *Triticum* families. Amplification of genomic DNA with a single set of primers deduced from one fully sequenced LMWG clone (Colot et al. 1989) yielded a wide variety of amplicons. Southern hybridization with a LMWG gene fragment as a probe demonstrated that these amplicons all have homologies with an LMWG gene. Isolation of DNA from a restricted region of the agarose gel, and cloning of the amplicons yielded three new LMWG clones, which have been designated LMWG-MB1, LMWG-MB2 and LMWG-MB3.

The open reading frame of these clones are 912-, 915- and 834-bp long respectively. The deduced amino-acid sequences of these clones are all homologous to the LMWG consensus reported previously, although this consensus was established on a very limited number of available sequences (Colot et al. 1989; Cassidy and Dvorak 1991). LMWG-MB1 contains all the consensus cysteine residues, including Cys at position 25 of the N-terminus which is involved in intermolecular bonding between the glutenin subunits (Kasarda 1989; Shewry and Tatham 1997). However, LMWG-MB2 and LMWG-MB3 both lack this important Cys residue, suggesting that the ability to form intermolecular bonds was lost in these clones.

One other important feature common to LMWG-MB2 and LMWG-MB3 is the presence of a premature stop codong (TAA) in the repetitive domain at position 53. The repetitive domain of prolamins is characterized by the abundance of Gln residues and, consequently, single-base mutations in this region are bound to produce a high frequency of stop codons since a C/T transition at the first nucleotide of both Gln codons (CAA and CAG) yields TAA and TAG. Moreover, it has been hypothesized that since methylation at position 5 of cytidine is a common event, and since methylated cytidine can be incorrectly replicated into thymidine, C/T transitions predominate over other single-base transition (Gojobori and Grauer 1992). In the α-gliadin gene family, another type of prolamin with Gln-rich repeated motifs, 50% of the genes isolated contain stop codons in the repetitive region (Anderson and Greene 1997).

However, LMWG-MB2 and LMWG-MB3 are not identical; a 81-bp insertion, which is also found in LMWG-MB1 is present in the repetitive domain of LMWG-MB2 but not in LMWG-MB3. This insertion contains three repeats of the motif PFSQQQ which is also found as a repeat in other parts of all three clones. Apart from this insertion, LMWG-MB2 and LMWG-MB3 are 100% identical. LMWG-MB1 and LMWG-MB2 also have slight differences in their amino-acid composition, the major one being the presence of a tetrapeptide in the repetitive region. Repeat sequences have been identified as hot spots for recombination, and it has been hypothesized that the repetitive structure of LMWGs might provide a basis for rapid evolution and divergence by duplication and/or deletion (Shewry et al. 1989). The current study tends to confirm this hypothesis since, of all variations identified between the three clones, almost all are found in the repetitive domains.

Comparison of the amino-acid sequences suggests that prolamins have evolved from a single ancestral protein (Shewry et al. 1995). The non-repetitive domain contains three conserved regions (A, B and C) implying that they originate from the triplication of a short ancestral domain. This domain is highly conserved not only among wheat seed-storage proteins (LMW glutenin and gliadin) but also within γ-secalins and β-hordein families (Kreis et al. 1985; Colot et al. 1989). This study indicates that the length polymorphism of LMWG gene is mainly due to size variation of the repetitive domain. Direct evidence has shown that insertions and/or deletions provide a mechanistic explanation for the allelic variation, and hence the resultant evolution, of prolamin genes.

Acknowledgments Research fund from CORPAQ (Conseil des Recherche en pêche et en Agroalimentaire due Québec). Contribution no. 640

References

Anderson OD, Greene FC (1989) The characterisation and comparative analysis of high-molecular-weight glutenin genes from genomes A and B of a hexaploid bread wheat. Theor Appl Genet 77:689–700

Anderson OD, Greene FC (1997) The gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theor Appl Genet 95:59–65

Bartels D, Altosaar I, Harberd N, Barker RF, Thompson RD (1986) Molecular analysis of g-gliadin gene families at the complex Gli-1 locus of bread wheat (*T. aestivum* L.) Theor Appl Genet 72:845–853

Cassidy BG, Dvorak J (1991) Molecular characterisation of a low-molecular-weight glutenin cDNA clone from *Triticum durum*. Theor Appl Genet 81:653–660

Colot V, Bartels D, Thompson R, Flavel R (1989) Molecular characterisation of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. Mol Gen Genet 216:81–90

Devereux J, Haebecli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 1287–395

Ellen JL Lew, Kuzmicky DD, Kasarda DD (1992) Characterization of low-molecular-weight subunits by reversed-phase high-performance liquid chromatography, sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and N-terminal amino-acid sequencing. Cereal Chem 69:508–515

Gojobori T, Li WH, Gaur D (1992) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18: 360–369

Gupta RB, Shepherd KW (1990) Two-step one-dimensional SDS-PAGE analysis of LMW subunits of glutenin. 1. Variation and genetic control of the subunits in hexaploid wheat. Theor Appl Genet 80:65–74

Gupta RB, Singh NK, Shepherd KW (1989) The cumulative effect of allelic variation in LMW and HMW glutenin subunits on dough properties in the progeny of two bread wheats. Theor Appl Genet 77:57–64

Gupta RB, Bekes F, Wrigley CW (1991) Prediction of physical dough properties from glutenin subunit composition in bread wheats: correlation studies. Cereal Chem 68:328–333

Kasarda DD (1989) Glutenin structure in relation to wheat quality. In: Pomeranz Y (ed) Wheat is unique. American Association of Cereal Chemists, St. Paul, Minnesota, pp 277–302

Khelifi D, Branlard G (1992) The effects of HMW and LMW subunits of glutenin and of gliadin on the technological quality of progeny from four crosses between poor-breadmaking quality and strong wheat cultivars. J Cereal Sci 16:195–209

Kreis M, Shewry PR, Ford BG, Ford J, Miflin BJ (1985) Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. In: Miflin BJ (ed) Oxford Surveys Plant Mol Cell Biol 2. Oxford University Press, New York, pp 253–317

Osborne TB (1924) The vegetable proteins, 2nd edn. Longmans, Green and Co., London

Payne PI (1987) Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Annu Rev Plant Physiol 38:141–153

Payne PI, Lawrence GD (1983) Catalogue of alleles for the complex gene loci, *Glu-1*, *Glu-B1* and *Glu-D1* which code for high-molecular-weight subunits of glutenin in hexaploid wheat. Cereal Res Commun 11:29–35

Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues: In: Gelvi SB, Schilperoort RA (eds) Plant molecular biology manual. Kluwer Academic Publishers. Dortrecht, pp A6:1–10

Shewry PR, Tatham AS (1997) Disulphide bonds in wheat gluten proteins. J Cereal Sci 25:207–227

Shewry PR, Field JM, Lew EJ-L, Kasarda DD (1982) The purification and characterisation of two groups of storage proteins (secalins) from rye (*Secale cereale* L.). J Exp Bot 33:261–268

Shewry PR, Halford NG, Tatham AS (1989) The high-molecular-weight subunits of wheat, barley and ryeGenetics, molecular biology, chemistry and role in wheat gluten structure and functionality. In: Miflin BJ (ed) Oxford Surveys Plant Mol Cell Biol 6, Oxford University Press, New York, pp 163–219

Shewry PR, Napier JA, Tatham AS (1995) Seed storage proteins: structure and biosynthesis. Plant Cell 7:945–956

Sissons MJ, Bekes F, Skerritt JH (1998) Isolation and functionality testing of low-molecular-weight glutenin subunits. Cereal Che 75:30–36

Vries S de, Hoge H, Bisseling T (1988) Isolation of total and polysomal RNA from plant tissues: In: Gelvin SB, Schilperoort RA (eds) Plant molecular biology manual, B6:1–13. Kluwer Academic Publishers. Dortrecht