Theor Appl Genet (1997) 95: 1119–1126 © Springer-Verlag 1997

R. D'Ovidio · M. Simeone · S. Masci · E. Porceddu

# Molecular characterization of a LMW-GS gene located on chromosome 1B and the development of primers specific for the *Glu-B3* complex locus in durum wheat

Received: 30 May 1997 / Accepted: 29 July 1997

**Abstract** Low-molecular-weight glutenin subunits (LMW-GS) represent a specific class of wheat storage proteins encoded at the *Glu-3* loci. Particularly interesting are the LMW-GS encoded at the *Glu-B3* locus because they have been shown to play an important role in determining the pasta-making properties of durum wheat. Genes encoding LMW-GS have been characterized but only a few of them have been assigned to specific loci. Notably, no complete LMW-GS gene encoded at the *Glu-B3* locus has yet been described. The present paper reports the isolation and characterization of a *lmw-gs* gene located at the *Glu-B3* locus. The clone involved, designated pLDNLMW1B, contains the entire coding region and 524 bp of the 5′ upstream region. A nucleotide comparison between the pLDNLMW1B clone and other LMW-GS genes showed the presence of some peculiar structural characteristics, such as short insertions in the promoter region, the presence of a cysteine codon in the repetitive domain, and a more regular structure of this region, which could be important for its tissue-specific expression and for the functional properties of the encoded protein, respectively.

**Key words** *Triticum durum* · LMW glutenin gene · PCR · *Glu-B3* locus

## Introduction

Storage proteins are wheat grain components endowing dough with viscoelastic properties and, consequently, the quality characteristics of its end products. The largest part of these proteins is represented by the prolamins, so termed because of their high proline and glutamine content. Wheat prolamins have been classified into two groups, the gliadins and glutenins, according to their solubility in aqueous/alcohol solutions (Osborn 1924). The gliadins are monomeric proteins having only intramolecular disulphide bonds, whereas glutenins are polymers of protein subunits, termed glutenin subunits, linked together by intermolecular disulphide bonds (Kasarda 1989). On the basis of their mobilities in SDS-PAGE under reducing conditions glutenin subunits are classified into two main groups: high-molecular-weight glutenin subunits (HMW-GS) and low-molecular-weight glutenin subunits (LMW-GS). The structure of HMW-GS and their encoding genes have been well characterized and nucleotide sequences of the complete set of HMW genes present in the bread-wheat cultivar Cheyenne have been reported (for a review see Shewry et al. 1992).

LMW-GS are encoded by genes at the orthologous *Glu-3* loci and are represented by several different components. The presence of some of them has been correlated with differences in the gluten properties of durum-wheat cultivars. In particular, pasta-making quality has been associated with the presence of specific LMW-GS encoded at the *Glu-B3* locus (Pogna et al. 1990). Despite their importance on the qualitative properties of durum wheat, only a limited number of LMW-GS genes have so far been characterized.

The lack of information on the structure of all the members of the LMW-GS gene family prevents any clear definition of their role on final gluten properties. The only extensive analysis aimed to characterize the LMW-GS gene family in a single cultivar has been carried out on the bread-wheat cultivar Cheyenne where a total of six genes, five encoded at the *Glu-D3* and one at the *Glu-A3* locus, have been sequenced (O. D. Anderson, personal communication). However, no LMW-GS genes at the *Glu-B3* locus have so far

Communicated by G. E. Hart

R. D'Ovidio (✉) · M. Simeone · S. Masci · E. Porceddu
Dipartimento di Agrobiologia e Agrochimica, Università della Tuscia, Via S. Camillo de Lellis, 01100 Viterbo, Italy
Fax: + 39 761 357242
E-mail: dovidio@unitus.it

been cloned, except for an amplification product corresponding to a partial coding region isolated from the bread-wheat cultivar Chinese Spring (Volkaert, GenBank Nucleotide Sequences Data Base, accession number: X84960).

In order to obtain an insight into the LMW-GS gene family, and particularly on members related to the qualitative characteristics of durum wheat, a research project aimed at characterizing the complete LMW-GS set of genes in a durum-wheat cultivar was started (D'Ovidio et al. 1996). The present paper reports the first complete sequence of a LMW-GS gene encoded at the *Glu-B3* locus.

## Materials and methods

### Plants

Analyses were carried out on the *Triticum durum* cultivar Langdon. Chromosome assignment was performed by utilizing D-genome chromosome substitution lines of durum-wheat cultivar Langdon (Joppa and Williams 1988) and nulli-tetrasomic lines of the bread-wheat cultivar Chinese Spring (Sears 1966). Primer specificity was evaluated on several cereal species and wheat cultivars including: *Hordeum vulgare*, *Oryza sativa*, *Zea mays*, *Secale cereale*, *Triticum urartu*, *Aegilops longissima*, *Ae. squarrosa*, *T. durum* cultivars Valnova, Creso, Aldura, *T. aestivum* cultivars Cheyenne, Salmone, Newton.

### DNA extraction

Genomic DNA was isolated from 5 g of leaves from single plants as reported in D'Ovidio et al. (1992 a).

### PCR amplification

PCR analyses were performed in a final reaction volume of 100 μl by using 100–300 ng of genomic DNA, 2.5 units of *Taq* DNA polymerase (Boehringer), 1 × *Taq* PCR buffer (Boehringer), 250 ng of each of the two primers, and 200 μM of each deoxyribonucleotide. Amplification conditions were for 30 cycles at 94°C for 1 min, 60°C (with primers **a-b**) or 62°C (with primers **c-b**) for 1 min, and 72°C for 1 min. A final step at 72°C for 7 min was also performed. Oligonucleotides used as primers were synthesized on the basis of LMW-GS genes previously published and have the following sequences: **a** 5′ CGAGCATATCCTAACAGCCCA 3′, **b** 5′ GTAGGCACCAACTCCGGTGC 3′; **c** 5′ TCCTGAGAAGTGCATGACATG 3′. Aliquots (10 μl) of the amplification products were fractionated on a 1.5% agarose gel in 1 × TBE buffer following standard procedures (Sambrook et al. 1989).

### Digoxigenin-labelling and hybridization experiments

Southern-blot analyses were carried out following standard procedures (Sambrook et al. 1989) and 300 ng of the LMW-GS insert contained in the pLMW21 clone (D'Ovidio et al. 1992b) were labelled with digoxigenin by nick translation using the 'Nick Translation Kit' (Boeheringer) and following the manufacturer's procedures.

### Cloning, nucleotide sequencing and computer analysis

The 1B LMW-GS amplification product was cloned into the modified *Eco*RV site of the pGEM-T plasmid vector (Promega) and subjected to nucleotide-sequence analysis using the chain-terminator method (Sanger et al. 1977). The PC/GENE computer program (IntelliGenetics, Inc., U.S.A.) was used to analyze the sequence data.

## Results

### PCR analysis and cloning

In order to amplify the 5′ flanking and complete coding regions of LMW-GS genes, several sets of primers were developed on the basis of published sequences (Colot et al., 1989; Cassidy and Dvorak, 1991; D'Ovidio et al. 1992b). PCR analysis on Langdon genomic DNA, performed by using the different sets of primers, produced different amplification patterns, the most interesting of which was obtained with primers **a** and **b** that showed four different fragments whose size ranged between 1350 bp and 1600 bp (Fig. 1). Southern-blot analysis, carried out using the pLMW21 clone (D'Ovidio et al. 1992 b) as a probe, confirmed that all four products corresponded to LMW-GS DNA fragments. PCR analyes with primers **a** and **b**, performed on genomic DNA isolated from D-genome chromosome substitution lines of the durum-wheat cultivar Langdon (Joppa and Williams 1988), indicated that three of these amplification products, whose sizes were about 1350 bp, 1400 bp and 1600 bp, could be assigned to chromosome 1A (Fig. 1, lane 3), and the fourth one, of about 1570 bp, to chromosome 1B (Fig. 1, lane 2).

The PCR product from chromosome 1B was purified and cloned into the pGEM-T vector. The transformation result produced a few recombinant colonies,

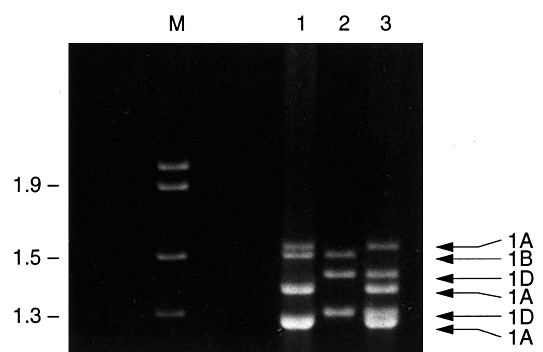


**Fig. 1** A 1.5% agarose gel of PCR products corresponding to LMW-GS genes obtained using primers **a** and **b**. Aliquots (10 μl) of amplification product obtained from *T. durum* cv Langdon (*1*), a Langdon 1D (1A) substitution line (*2*) and Langdon 1D (1B) substitution line (*3*). On the right side the chromosomal assignment is reported. *M* λDNA digested with *Hin*dIII used as a molecular-weight marker; the figure shows only part of the complete restriction pattern

probably as a consequence of difficulties in the insertion of the PCR fragment into the plasmid vector. The recombinant colonies contained the expected PCR fragment and one of them, designated pLDNLMW1B, was sequenced.

## Sequencing and comparison analyses

The nucleotide sequence of the pLDNLMW1B clone is 1574 bp long and contains an open reading frame that starts at nucleotide 525 and encodes a polypeptide of 330 amino acids (Fig. 2). The nucleotide sequence of pLDNLMW1B showed the same general structure of as previously reported LMW-GS genes with differences, however, in the structure of the repetitive domain and in the position of one cysteine codon. The repetitive domain is 393 bp long and is composed of 18 repeats ranging from 18 to 27 bp. The consensus sequence of the shorter repeat is an 18-bp fragment coding for a hexapeptide whose sequence is PPFSQQ (Fig. 3). The deduced protein contains the typical 20

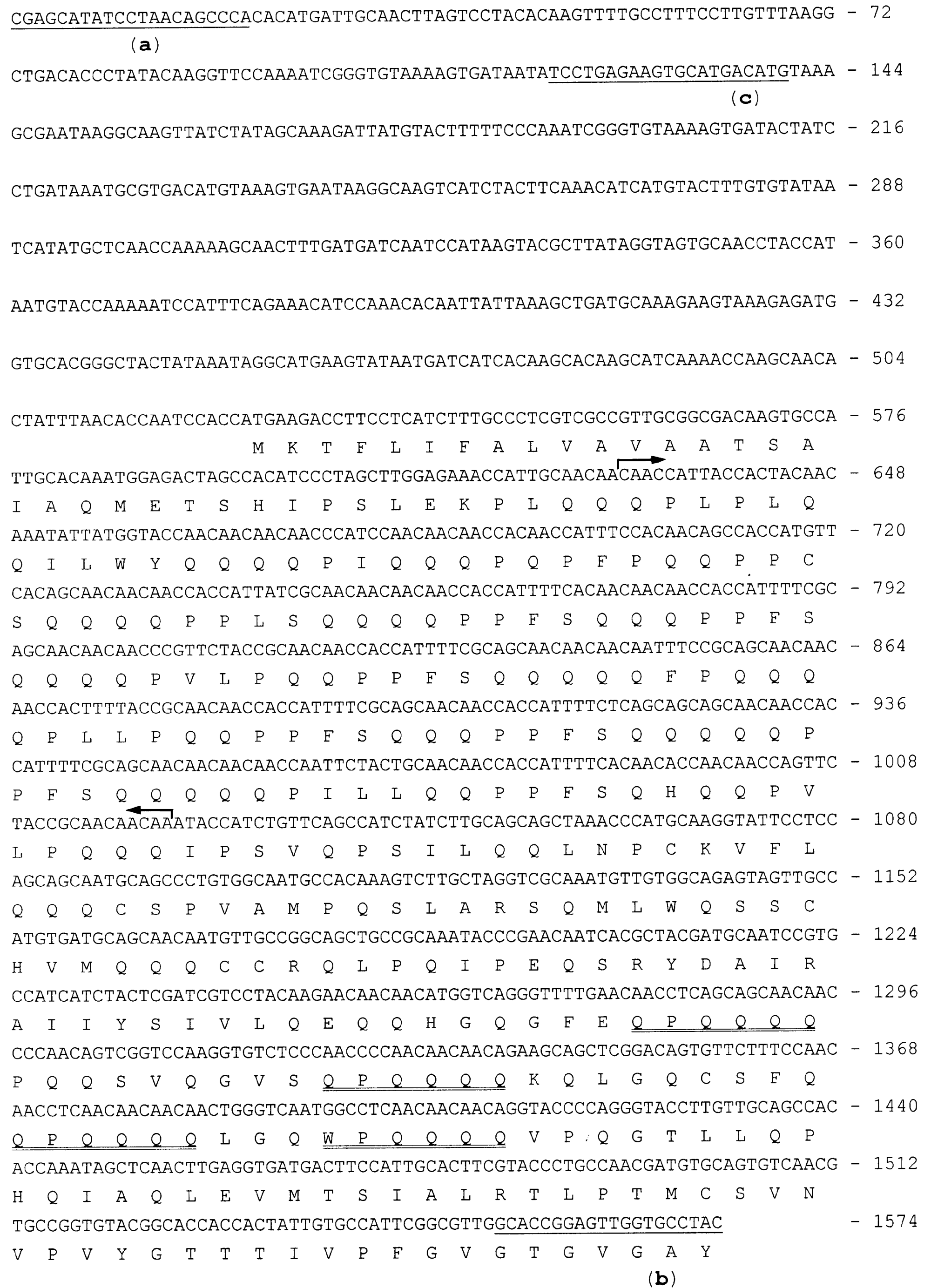


**Fig. 2** Nucleotide and deduced amino-acid sequences of the pLDNLMW1B clone. The primers used in amplification reactions are *underlined* and identified by a *bold letter in brackets*. Repeated motifs in the C-terminal region are *double underlined*. *Arrows* indicate the position and length of the repetitive domain. EMBL Data library accession number Y14104

amino-acid signal peptide followed by an N-terminal region composed of 15 amino acids, a repetitive domain of 131 amino acids, and a C-terminal domain of 164 amino acids. The polypeptide possesses a calculated molecular mass of 37 743, an isoelectric point of 7.55, and a glutamine and proline content of 34% and 15%, respectively. The hydropathy profile (data not shown) revealed the hydrophilic character of the repetitive domain and the general hydrophobic character of the N-terminal and C-terminal regions. The deduced LMW-GS contains eight cysteine residues, seven of which are located in the C-terminal domain and one in the upstream region of the repetitive domain (Fig. 2).

Comparison of the nucleotide sequence of pLDNLMW1B with reported sequences of LMW-GS genes indicated a high degree of homology between them, ranging from 65% to 80% (Table 1 and Fig. 4). The C-terminal domain was strongly conserved, whereas the N-terminal and repetitive domains contain several differences. Alignment of the deduced amino-acid sequences also showed that the homology is dispersed along the entire sequence, with a greater homology in the C-terminal region. Worthy of note is the position of cysteine residues that are conserved, with the exception of the first and seventh ones. In fact, a number of LMW-GS genes possess a cysteine residue in the short N-terminal region, whereas the pLDNLMW1B clone and the incomplete LMW-GS genes from the hexaploid wheat cultivar cultivar Chinese Spring (Volkaert, EMBL accession numbers X84960 and X84961) show the first cysteine in the repetitive domain. The seventh cysteine residue can be present in two slightly different positions at the C-terminus, about 15 amino-acid residues apart from each other (Fig. 4).

To further analyze the extent of similarity between the pLDNLMW1B clone and other LMW-GS genes, the 5′ flanking region of pLDNLMW1B was compared with corresponding regions of other LMW-GS genes. This analysis was particularly interesting because the length of the 5′ upstream region of the pLDNLMW1B clone contained the nucleotide sequence necessary to drive endosperm-specific expression, as previously demonstrated in a LMW-GS gene from the *Glu-D3* locus (Colot et al. 1987). A nucleotide comparison was carried out with the upstream regions of the LP1211 (Pitts et al. 1988) and LMW-1D1 (Colot et al. 1989) clones, which represent the only LMW-GS gene clones for which the 5′ flanking regions have been reported so far. The result of the analysis showed a high degree of homology between all three 5′ upstream regions (86–89%); their homology is dispersed along the entire 5′ region and is interrupted by the presence of an insertion of about 100 bp (from base 172 to 274) in the pLDNLMW1B clone (Fig. 5).

**Fig. 3** Alignment of the amino-acid repeats present in the repetitive domain of the pLDNLMW1B clone. The figure shows the regular distribution of the repeats within the repetitive domain which lies between amino acids 36 and 166. *Cons* = consensus sequence

```
1     QPLPLQ
2     QILWYQ    QQQ
3     PIQQQP
4     QPFPQQ
5     PPCSQQ    QQ
6     PPLSQQ    QQ
7     PPFSQQ    Q
8     PPFSQQ    QQ
9     PVLPQQ
10    PPFSQQ    Q
11    QQFPQQ    QQ
12    PLLPQQ
13    PPFSQQ    Q
14    PPFSQQ    QQQ
15    PPFSQQ    QQQ
16    PILLQQ
17    PPFSQH    QQ
18    PVLPQQ    Q

Cons  PPFSQQ
```

**Table 1** Nucleotide sequences of LMW-GS genes so far reported. C, entire coding region; P, partial coding region; Prom, clone containing the promoter region

| Clone | EMBL accession number | Genotype | Locus | Homology with the promoter region of the PLDNLMW1B clone | Homology with coding region of the PLDLNLM1B clone | Type of clone | Reference |
|---|---|---|---|---|---|---|---|
| LP1211 | X07747 | *T. aestivum* | Unknown | 86.6% | 62.3% | Genomic (C) (Prom) | Pitts et al. 1988 |
| pTAG544 | J01309 | *T. aestivum* | Unknown | | 84.5% | cDNA (P) | Bartels and Thompson 1983 |
| LMWG-1D1 | X13306 | *T. aestivum* | *Glu-D3* | 89.0% | 79.8% | Genomic (C) (Prom) | Colot et al. 1989 |
| pB48 | M11335 | *T. aestivum* | Unknown | | 71.7% | (P) | Okita 1984 |
| pB11-33 | M11077 | *T. aestivum* | Unknown | | 81.3% | (C) | Okita et al. 1985 |
| pLMW21 | X62588 | *T. durum* | Unknown | | 75.8% | Genomic(C) | D'Ovidio et al. 1992 a |
| pTdUCD1 | X51759 | *T. durum* | Unknown | | 77.0% | cDNA (C) | Cassidy and Dvorak 1991 |
| | X84960 | *T. aestivum* | *Glu-B3* | | 98.9% | Genomic, (P) | Volckaert G., GenBank |
| | X84961 | *T. aestivum* | *Glu-D3* | | 69.2% | Genomic (P) | Volckaert G., GenBank |

**Fig. 4** Comparison of the deduced amino-acid sequences of all LMW-GS genes reported so far. Conserved positions in all sequences are indicated by *asterisks*. *Dots* and *arrow-heads* indicate conserved and non-conserved positions of cysteine residues, respectively

```
pLDNLMW1B   MKTFLIFALVAVAATSAIAQMETSHIPSLEKPLQQQPLPLQQILWYQQQQPIQQQ    55
X84960      ------------------------HIPSLEKPLQQQPLPLQQILWYQQQQPIQQQ    31
X84961      ------------------------HIPSLEKPSQQQPLPLQQILWYHQQQPIQQQ    31
LP1211      MKTFLVFALLALAAASAVAQISQQ---QQAPPFSQQQQP---PFSQQQQPPFSQQ    49
pTAG544     -----------------------------------------------------QQ     2
LMWG-1D1    MKTFLVFALLAVAATSAIAQMETRCIPGLERPWQQQPLP---P----QQT-FPQQ    47
pB11-33     MKTFLVFALIAVVATSAIAQMETSCISGLERPWQQQPLP---P----QQS-FSQQ    47
pTdUCD1     MKTFLVFALLAVVATSTIAQMETSCIPGLERPWQEQPLP---P----QHTLFPQQ    48
pLMW21      MKTFLVFALLAVVATSAIAQMDTSCIPGLERPWQQQPLP---P----QQT-FPQQ    47
pB48        ---------------------------------------------------PQQ     3
                                                                **

pLDNLMW1B   PQ-PFPQQ---PPCSQQQQPPLSQQQQPPFSQQQPPFSQQQQPVLPQQ--PPFSQ   104
X84960      PQ-PFPQQ---PPCSQQQQPPLSQQQQPPFSQQQPPFSQQQQPILPQQ--PPFSQ    80
X84961      PQ-PFPQQ---PPCSQQQQPPLSQQQQPPFSQQQPPFSQQELPILPQQ--PPFSQ    80
LP1211      QQSPFSQQQQQPPFAQQQQPPFSQQP-PISQQQQPPFSQQQQPQFSQQQQPPYSQ   103
pTAG544     P--PFSQQQQ--------------------------------PVLPQ--QPPFSQ    21
LMWG-1D1    P--LFSQQQQQQLFPQQ--PSFSQQQ-PPFWQQQPPFSQQ-PILPQ--QPPFSQ    94
pB11-33     P--PFSQQQQQPL----------PQQ-PSFSQQQPPFSQQ-PILSQ--QPPFSQ    86
pTdUCD1     Q--PFPQQQQ----------------PPFSQQQPSFLQQQ-PILPQ--LP-FSQ    80
pLMW21      P--PFSQQQQQQ----------------PFPQQ-PSFSQQ-PILPQ--GPPFPQ    80
pB48        P-----------FPLQ--P----QQ-SFLWQSQQPFLQQ-----PQ--QPS-PQ    31
                                                          *   *   *

pLDNLMW1B   QQQQ---------FPQQQQPLLPQQPPFSQQQPPFSQQQQQP------PFSQQQQ   144
X84960      QQQQ---------FPQQQQPLLPQQPPFSQQQPPFSQQQQQP------PFSQQQQ   120
X84961      QQQPQFSQQQQ-PFPQQQQPLLLQQPPFSQQRPPFSQQQQQPVLPQQPPFSQQQQ   134
LP1211      QQQPPYSQQQQPPFSQQQQPPFSQQ----QQQPPFTQQQQQQ--QQQQPFTQQQQ   152
pTAG544     QQQPILPQ--QPPFS----------------------QQQQQ--PVL--------    42
LMWG-1D1    QQQLVLPQ--QPPFSQQQQPVLPPQ------QSPFPQQQQH--QQL--------   131
pB11-33     QQQPVLPQ--QSPFSQQQQLVLPPQ------Q--------QQ--QQL--------   115
pTdUCD1     QQQPVLPQ--QSPFSQQQ-LVLPPQ------Q--------QY--QQV--------   108
pLMW21      QTQPVLPQ--QSPFSQQQQLILPPQ------Q--------Q---QQL--------   108
pB48        PQQVVQII--SPATPTTIPSAGKPT------SAPFPQQQQH--QQL--------    68
              *                                       *

pLDNLMW1B   ----Q-PILLQQ--PPFSQHQQPVL-PQQQIPSVQPSILQQLNPCKVFLQQQCSP   191
X84960      ----Q-PILLQQ--PPFSQHQQPVL-PQQQIPSVQPSILQQLNPCKVFLQQQCSP   167
X84961      ----QQPILPQQ--PPFSLHQQPVL-PQQQIPYVQPSILQQLNPCKVFLQQQCSP   182
LP1211      PPFSQQPPISQQQQPPFLQQQRPPFSRQQQIPVIHPSVLQQLNPCKVFLQQQCIP   207
pTAG544     --------------------------PQQQILFVHPSILQQLNPCKVFLQQQCSP    71
LMWG-1D1    --------------------------VQQQIPVVQPSILQQLNPCKVFLQQQCSP   160
pB11-33     --------------------------VQQQIPIVQPSVLQQLNPCKVFLQQQCSP   144
pTdUCD1     --------------------------LQQQIPIVQPSVLQQLNPCKVFLQQQCNP   137
pLMW21      --------------------------PQQQISIVQPSVLQQLNPCKVFLQQQCSP   137
pB48        --------------------------AQQQIPVVQPSILQQLNPCKVFLQQQCSP    97
                                       ****    ** *************** *

pLDNLMW1B   VAMPQSLARSQMLWQSSCHVMQQQCCRQLPQIPEQSRYDAIRAIIYSIVLQEQQH   246
X84960      VAMPQSLARSQMLWQSSCHVMQQQCCRQLPQIPEQSRYDAIRAIIYSIVLQEQQH   222
X84961      VAMPQSLARSQMLWQSSCHVMQQQCCQQLPRIPEQSRYDAIRAIIYSIVLQEQQH   237
LP1211      VAMQRCLARSQMLQQSICHVMQQQCCQQLRQIPEQSRHESIRAIIYSIILQQQQQ   262
pTAG544     VAMPQSLARSQMLQQSSCHVMQQQCCQQLLQIPQQSRYEAIRAIIYSIILQEQQQ   126
LMWG-1D1    VAMPQRLARSQMLQQSSCHVMQQQCCQQLPQIPQQSRYEAIRAIIYSIILQEQQQ   215
pB11-33     VAMPQRLARSQMWQQSSCHVMQQQCCQQLQQIPEQSRYEAIRAIIYSIILQEQQQ   199
pTdUCD1     VAMPQRLARSQMLQQSSCHVMQQQCCQQLPQIPEQSRYDVIRAITYSIILQEQQQ   192
pLMW21      VAIPQRLARSQMWQQSSCHVMQQQCCQQLSQIPEQSRYDAIRAITYSIILQEQQQ   192
pB48        VAMPQRLARSQMLQQSSCHVMQQQCCQQLPQIPQQSRYQAIRAIIYSIILQEQQQ   152
            **    ******  ** ********* **  ** ***    **** *** ** **

pLDNLMW1B   GQGFEQPQQQQPQQSVQGVSQPQQQ-QKQLGQC---SFQQPQQQQLGQWPQQ---   294
X84960      GQGLNQPQQQQPQQSVQGVSQPQQQ-QKQLGQC---SFQQPQQ------------   261
X84961      GQGFNQPQQQQPQQSVQGVSQPQQQ-QKQLGQC---SFQQPQQ------------   276
LP1211      ------QQQQQQQQGQSIIQYQQQQPQQLGQCVSQPLQQLQQQ-LGQQPQQQQL   310
pTAG544     VQGSIQTQQQQPQELGQCVSQPQQQ--------------QSQQQ-LGQQPQQ---   163
LMWG-1D1    VQGSIQSQQQQPQQLGQCVSQPQQQSQQQLGQ------------------QPQQ   251
pB11-33     --GFVQPQQQQPQQSGQGVSQSQQQSQQQLGQC---SFQQPQQQ-LGQQPQQQQQ   248
pTdUCD1     --GFVQAQQQQPQQLGQGVSQSQQQSQQQLGQC---SFQQPQQQ-LGQQPQQQQ-   240
pLMW21      --G--QSQQQQPQQSGQGVSQSQQQSQQQLGQC---SFQQPQQQ-LGQQPQQQQ-   238
pB48        VQGSIQSQQQQPQQLGQCVSQPQQQSQQQLGQ------------------QPQQ   188
                    **** *    *    * ***

pLDNLMW1B   QQVPQGTLLQPHQIAQLEVMTSIALRTLPTMCSVNVPVYGTTTIVPFGVGTGVGAY-   350
X84960      ---------------------------------------------------------   261
X84961      ---------------------------------------------------------   276
LP1211      ----------AHQIAQLEVMTSIALRTLPTMCNVNVPLYETTTSVPLGVGIGVGVY-   356
pTAG544     QQLAQGTFLQPQQVAQLEVMTSIALRTLPTMCRVNVPLSRTTTSVPFGVGAGVGAY-   219
LMWG-1D1    QQLAQGTFLQPHQIAQLEVMTSIALRILPTMCSVNVPLYRTTTSVPFGVGTGVGAY-   307
pB11-33     QQVLQGTFLQPHQIAHLEAVTSIALRTLPTMCSVNVPLYSATTSVPFGVGTGVGAY-   304
pTdUCD1     --VLQGTFLQPHQIAHLEVMTSIALRTLPTMCSVNVPLYSSTTSVPFSVGTGVGAYL   295
pLMW21      --VQQGTFLQPHQIAHLEVMTSIALRTLPTMCSVNVPLYSSTTSVPFGV--------   285
pB48        QQLAQGTFLQPHQIAQLEVMTSIALRILPTMCSVNVPLYRTTTSVPFGVGTGVGAY-   244
```

**Fig. 5** A nucleotide comparison of the 5′ flanking region of LMW-GS genes. Conserved positions in all sequences are indicated by *asterisks*. PROMLMW1B, 5′ flanking region of the pLDNLMW1B clone; PROMLMWP12, 5′ flanking region of the LP1211 clone (Pitts et al. 1988); PROMLMW1D, 5′ flanking region of the LMW-1D1 clone (Colot et al. 1989)

```
PROMLMW1B    CGAGCATATCCTAACAGCCCACACATGATTGCAA-CTTAGTCCTACACAAGTTTT    54
PROMLMWP12   CGAGCATATCTTAACAGCCCACACACGATTGCAAACTTAGTCCTACACAAGCTTT    55
PROMLMW1D    CGAGCATATCCTAACAGCCCACACATGATTGCAAACTTAGTCATACACAAGTTTT    55
             ********** ************** ******** ******* ******** ***

PROMLMW1B    GCCTTTCCTTGTTTAAGGCTGACACCCTATACAAGGTTCCAAAATCGGGTGTAAA   109
PROMLMWP12   GCCTTTC-TTGTTTACGGCTGACAACCTATACAAGGTTCCAAACTCGGTTGCAAA   109
PROMLMW1D    GCCTTTC-TTGTTTACGGCTGACAGCCTATACAAGGTTCCAAACTCGGTTGTAAA   109
             ******* ******* ******** ****************** **** ** ***

PROMLMW1B    AGTGATAATATCCTGAGAAGTGCATGACATGTAAAGCGAATAAGGCAAGTTATCT   164
PROMLMWP12   AGTGATACTATCCTGATAAGTGCGTGACATGTAAAGTTAATAAGGTGAGTCATAT   164
PROMLMW1D    AGTGATACTATCTTGATAAGTGTGTGACATGTAAAGTTAATAAGGTGAGTCATAT   164
             ******* **** *** *****  ************  *******  *** ** *

PROMLMW1B    ATAGCAAAGATTATGTACTTTTTCCCAAATCGGGTGTAAAAGTGATACTATCCTG   219
PROMLMWP12   GTACCAAA-------------------CATCGAG---------GTTTCT------   185
PROMLMW1D    ATAGCAAA-------------------TATCGGG---------GTTTCT------   185
              ** ****                    **** *         * * **

PROMLMW1B    ATAAATGCGTGACATGTAAAGTGAATAAGGCAAGTCATCTACTTCAAACATCATG   274
PROMLMWP12   ------------------------------------------------------G   186
PROMLMW1D    ------------------------------------------------------G   186
                                                                   *

PROMLMW1B    TACTTTGTGTATAATCATATGCTCAACCAAAAAGCAACTTTGATGATCAAT----   325
PROMLMWP12   TACTTTGTGTATGATCATATGCACAACTAAAAAGCAACTTTGATGAT----GAAT   237
PROMLMW1D    TACTTTGTGTGTGATCGTATGCACAACTAAAAATCAACTTTGATGATCAATATAT   241
             ********** * *** ***** **** ***** *************

PROMLMW1B    CCATAAGTACGCTTAT--AGGTAGTGCAA-CCTACCATAATGTACCAAAAA--TC   375
PROMLMWP12   CCAAAAGTACGCTTTTGTAGCTAGTGCAACCC-AACACAATGTACCAAAAAAATT   291
PROMLMW1D    CCAAAAGTACGCT--TGTAGCTAGTGCAAACCTAACCCAATGTAACAAAATAATT   294
             *** *********  *  ** ******** ** * *  ****** *****   *

PROMLMW1B    CATTTCAGAAACATCCAAACACAATTATTAAAGCTGATGCAAAGAAG-TAAAGAG   429
PROMLMWP12   CATTTCAGATGCATCCAAACAGAATTATTAAAGCCGGTGCAAAGAAGGAAAAGAG   346
PROMLMW1D    CATTTCAGATGGAGCCAAACAGAATTATTAAAGCTGATGCAAAGAAGGAAAAGAG   349
             *********   * ******* ************ * **********  ******

PROMLMW1B    ATGGTGCACGGGCTACTATAAATAGGCATGAAGTATAATGATCATCACAAGCACA   484
PROMLMWP12   GTGGTGTCCCGGCAACTATAAATAGGCATGAAGTATAAAGATCATCACAAGTACA   401
PROMLMW1D    GTGGTTCCTGGGCTACTATAAATAGGCATGAAGTATAAAGATCATCACAAGCACA   404
              ****     *** ************************ ************ ***

PROMLMW1B    AGCATCAAAACCAAGCAACACTATTTAACACCAATCCACC                  524
PROMLMWP12   AGCATCAAAGCCAAGCAACACTAGTTAACACCAATCCACA                  441
PROMLMW1D    AGCATCAGAACCAAGCAACACTAGTTAACACCAATCCACC                  444
             ******* * ************* ***************
```

Development of *Glu-B3* locus-specific primers

On the basis of the nucleotide substitutions existing between the pLDNLMW1B clone and other LMW-GS genes, a pair of primers specific for the LMW-GS gene present at the *Glu-B3* locus was developed. One oligonucleotide is a 21-mer located in the 5′ flanking region (primer '**c**'), whereas the other one is located at the end of the coding region and corresponds to the 20-mer oligonucleotide used in the previous PCR assay (primer '**b**').

Electrophoretic separation on an agarose gel of PCR products obtained from the genomic DNA of cv Langdon showed that a single amplification product of about 1450 bp was present. To identify whether the PCR product was locus specific, a PCR assay was performed on the genomic DNA of nulli-tetrasomic lines of cv Chinese Spring (Fig. 6 A, lanes 1–4) and Langdon substitution lines (Fig. 6 A, lanes 5–7). The result demonstrated the specificity of the selected primers, being the amplification products present in genotypes carrying the 1B chromosome and absent in those lacking this chromosome (Fig. 6 A).

To confirm further the specificity of the selected primers, PCR analysis was carried out on wheat cultivars, wild wheat relatives such as *T. urartu*, *Ae. longissima* , and some cereal species including rice, barley, rye and corn. The results showed that the amplification product of about 1450 bp was present only in durum (Fig. 6, lanes 8–10)- and bread-wheat cultivars (Fig. 6, lanes 11–13) and in *Ae. longissima* (Fig. 6B).

## Discussion

The molecular characterization of all members belonging to a gene family is a basic step in understanding the role that the corresponding gene products play in

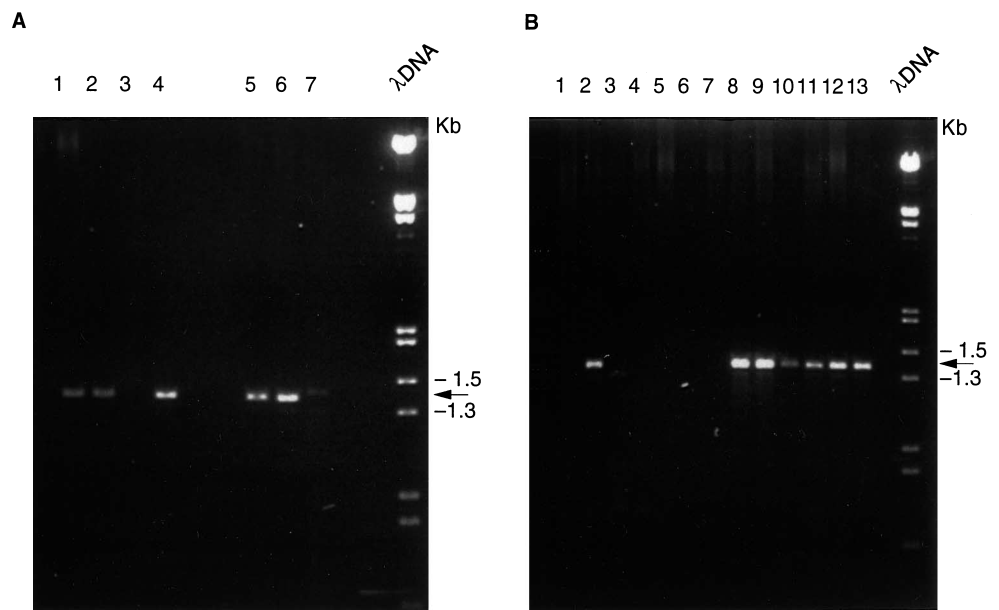


**Fig. 6A, B** A 1.5% agarose gel of amplification products obtained with primers **c** and **b**, and corresponding to LMW-GS genes at the *Glu-B3* locus. **A** Chromosomal assignment of PCR products: Lanes *1* Chinese Spring; *2* Chinese Spring N1AT1D; *3* Chinese Spring N1BT1A; *4* Chinese Spring N1DT1B; *5* Langdon; *6* Langdon 1D(1A); *7* Langdon 1D(1B). **B**; *1 T. urartu*; *2 Ae. longissima*; *3 Ae. squarrosa*; *4* Barley; *5* Rice; *6* Rye; *7* Corn; *8* Durum wheat cv Valnova; *9* Durum wheat cv Creso; *10* Durum wheat cv Aldura; *11* Bread wheat cv Cheyenne; *12* Bread wheat cv Salmone; *13* Bread wheat cv Newton

different biological and technological properties. The functional role of the different members of the LMW-GS gene family is not known; however, genetic and technological analyses have shown that specific LMW-GSs encoded at the *Glu-B3* locus play an important role in determining the viscoelastic characteristics of durum-wheat flour. In order to verify the possibility of correlating the molecular structure of LMW-GS with their functional properties, a particular effort has been made to isolate and characterize LMW-GS genes encoded at the *Glu-B3* locus. A first interesting result in this direction was obtained with the isolation of the pLDNLMW1B clone. This clone represents the first complete sequence of a LMW-GS gene present at the *Glu-B3* locus. A nucleotide comparison of this sequence clone with that of other LMW-GS genes showed the presence of differences in the promoter region and peculiar characteristics in the coding region, such as a more regular structure of the repetitive domain and the presence of a cysteine residue within this region, not found in previously reported LMW-GS genes. The repetitive domain is composed of a hexapeptide motif repeated 18 times and having the consensus sequence PPFSQQ with the possible presence of one-to-three additional glutamine residues in each repeat. The consensus sequence composed by the hexapeptide fits better for the pLNDLMW1B clone than the heptamer motif PPFSQQQ reported for other LMW-GS genes (Colot et al. 1989).

The fifth hexapeptide repeat in the pLDNLMW1B clone contains a cysteine residue in place of the phenylalanine amino-acid residue which is present in all the other LMW-GS genes reported so far; such a substitution could have been arisen from a $T \rightarrow G$ transversion event. In fact, since this cysteine is encoded by a TGT triplet, whereas the phenylalanine present in the repeating units of LMW-GS are usually encoded by TTT triplets, a possible origin of the cysteine residue is the occurrence of a $T \rightarrow G$ transversion in the second nucleotide of the triplet.

On the basis of the N-terminal amino-acid sequences of purified polypeptides, two types of LMW-GS are distinguished, the LMW-GS Ser-type and the LMW-GS Met-type, according to the first amino acid of the sequence (Lew et al. 1992). The Met-type group is mainly composed of polypeptides having the METSH- or METSC-N-terminal amino-acid sequences with the former group being both qualitatively and quantitatively better represented than the latter (Lew et al. 1992). In spite of this observation, all the deduced amino-acid sequences of *lmw-gs* genes characterized so far correspond to the METSC group, whereas the pLDNLMW1B clone here reported encodes a polypeptide with the METSH sequence.

Comparison between the deduced amino-acid sequences of *lmw-gs* reported so far, showed that the first and seventh cysteine residues can be present in alternative positions. Particularly noteworthy is the observation that these particular cysteines seem to be the only ones involved in intermolecular disulphide bonds (Keck et al. 1995).

A nucleotide comparison of the 5′ upstream region revealed a high degree of homology between the analyzed sequences and the occurrence of a large insertion in the pLDNLMW1B clone. Since the upstream region compared contains all the sequences necessary for endosperm-specific expression (Colot et al. 1987), the presence of this insertion could be of functional importance.

The characterization of the pLDNLMW1B clone allowed the development of a PCR assay specific for the LMW-GS gene encoded at the *Glu-B3* locus. PCR assays specific for this locus have already been reported but concern only the coding region and not the

promoter region (D'Ovidio 1993; Van Campenout et al. 1995). The present PCR assay allows the amplification of the 5′ flanking and complete coding regions of LMW-GS genes encoded at the *Glu-B3* locus, so providing the opportunity to isolate and analyze additional genes from this locus which so far have been poorly characterized. The usefulness of this pair of primers in amplifying LMW-GS genes from the B genome, or closely related genomes, also makes possible the characterization of specific LMW-GS genes from related wheat species for molecular phylogeny studies as well as providing a source of new alleles for breeding purposes. Moreover, the finding that these primers gave the same amplification products in *Ae. longissima* and cultivated wheats confirms further the similarity between the S-genome present in species of the *Sitopsis* Section, such as *Ae. longissima*, and the B-genome of durum and bread wheats (Kerby and Kuspira 1987). Finally, the specificity and effectiveness of this assay could also be valuable for mapping purposes in place of time-consuming hybridization experiments.

**Acknowledgments** The authors thank Dr. L. R. Joppa (UDSA, North Dakota State University, Fargo, N.D., U.S.A.) for providing Langdon substitution lines, and Prof. D. Lafiandra for critical reading of the manuscript. Research supported by the Italian 'Ministero delle Risorse Agricole, Alimentari e Forestali', National Research Project 'Plant Biotechnology'.

## References

Bartels D, Thompson RD (1983) The characterization of cDNA clones coding for wheat storage proteins. Nucleic Acids Res 11 : 2961–2977

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

Colot V, Robert LS, Kavanagh TA, Bevan MW, Thompson RD (1987) Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO J 6 : 3559–3564

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

D'Ovidio R (1993) Single-seed PCR of LMW glutenin genes to distinguish between durum wheat cultivars with good and poor technological properties Plant Mol Biol 22 : 1173–1176

D'Ovidio R, Tanzarella OA, Porceddu E (1992 a) Isolation of an alpha-type gliadin gene from *Triticum durum* Desf and genetic polymorphism at the *Gli*-2 loci. J Genet Breed 46 : 41–48

D'Ovidio R, Tanzarella OA, Porceddu E (1992 b) Nucleotide sequence of a low-molecular-weight glutenin from *Triticum durum*. Plant Mol Biol 18 : 781–784

D'Ovidio R, Simeone M, Marchitelli C, Masci S, Porceddu E (1996) Isolation and characterization of members of the LMW glutenin gene family in durum wheat. In: Wrigley CW (ed) Proc 6th Int Gluten Workshop pp 81–84

Joppa LR, Williams ND (1988) Langdon durum disomic substitution lines and aneuploid analysis in tetraploid wheat. Genome 30 : 222–228

Kasarda DD (1989) In: Pomeranz Y (ed) Wheat is unique. Am Assoc Cereal Chem, St. Paul, Minnesota, USA, pp 277–301

Keck B, Köhler P, Wieser H (1995) Disulphide bonds in wheat gluten: cystine peptides derived from gluten proteins following peptic and thermolitic digestion. Z Lebensm Unters Forsch 200 : 432–439

Kerby K, Kuspira J (1987) The phylogeny of the polyploid wheats *Triticum aestivum* (bread wheat) and *Triticum turgidum* (macaroni wheat). Genome 29 : 722–737

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

Okita TW (1984) Identification and DNA sequence analysis of a gamma-type gliadin cDNA plasmid from winter wheat. Plant Mol Biol 3 : 325–332

Okita TW, Cheesbrough V, Reeves CD (1985) Evolution and heterogeneity of $\alpha/\beta$-type and $\gamma$-type gliadin DNA sequences. J Biol Chem 260 : 8203–8213

Osborn TB (1924) The vegetable proteins. Longmans, Green and Co., London

Pitts EG, Rafalski JA and 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

Pogna NE, Autran JC, Mellini F, Lafiandra D, Feillet P (1990) Chromosome 1B-encoded gliadins and glutenin subunits in durum wheat: genetics and relationship to gluten strength. J Cereal Sci 11 : 15–34

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

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

Shewry PR, Halford NG, Tatham AS (1992) High-molecular-weight subunits of wheat glutenin. J Cereal Sci 15 : 105–120

Van Campenhout S, Vander Stappen J, Sagi 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