Theor Appl Genet (2001) 103:733–742 © Springer-Verlag 2001

A. De Bustos · P. Rubio · N. Jouve

Characterisation of two gene subunits on the 1R chromosome of rye as orthologs of each of the *Glu-1* genes of hexaploid wheat

Received: 11 December 2000 / Accepted: 17 April 2001

Abstract The visco-elastic properties of bread flour are firmly associated with the presence or absence of certain HMW subunits coded by the *Glu-1* genes. Identifying allelic specific molecular markers (AS-PCR) associated with the presence of *Glu-1* genes can serve as a valuable tool for the selection of useful genotypes. This paper reports the use of primers designed from nucleotide sequences of the *Glu-D1* gene of wheat (AS-PCR for *Glu-D1y10*) that recognise and amplify homologous sequences of the *Glu-R1* gene subunits of rye. The primers amplify the complete coding regions and provided two products of different size in rye, in wheats carrying the substitution 1R(1D) and in rye-wheat aneuploid lines carrying the long arm of chromosome 1R. The location, the molecular characterisation of these sequences and their expression during grain ripening seem to demonstrate that the amplification products correspond to structural genes encoding the high-molecular-weight (HMW) glutenins of rye. The homology of the rye gene to subunits encoding HMW glutenins in wheat was confirmed by Southern blots and sequencing. The amplification-products were cloned, sequenced and characterised, and the sequences compared with the main glutenin subunits of wheat and related species. Further, an RT-PCR experiment was performed using primers designed from the sequence of both amplified products. This assay demonstrated that both sequences are expressed in endosperm during grain ripening. The results of these analyses suggest that both gene subunits correspond to x- and y-type genes of the *Glu-R1* locus of rye.

Keywords High-molecular-weight glutenin · *Glu-R1* · Orthologous genes · *Secale cereale* · *Triticum aestivum*

Communicated by J.W. Snape

A. De Bustos · P. Rubio · N. Jouve (✉)
Department of Cell Biology and Genetics, University of Alcalá, 28871 Alcalá de Henares, Madrid, Spain
e-mail: nicolas.jouve@uah.es
Tel.: +34-91-8854750, Fax: +34-91-8854799

Introduction

The *Glu-1* genes of wheat have different alleles whose products can be separated using SDS-PAGE electrophoresis (Payne et al. 1981). It has been demonstrated that breadmaking quality is firmly associated with the presence or absence of certain HMW subunits of the *Glu-1* genes (Payne 1987; Odenbach and Mahgoub 1988). The visco-elastic properties of bread flour have been associated with the formation of disulphide bridge-linked aggregates, owing to the presence of several cysteine residues in both the N-terminal and C-terminal regions of HMW gluten peptides (Miflin et al. 1983). The influence of the size and type of glutenin polymers on baking-flour quality has been studied extensively (Huebner and Wall 1976; Dachkevitz and Autran 1989).

The chromosomal locations of the structural genes for secalins in rye were formerly investigated by comparing the endosperm protein patterns of wheat, rye, and lines of wheat with added rye chromosomes or chromosome arms (Lawrence and Shepherd 1981; Singh and Shepherd 1984). These studies showed that the structural genes for HMW secalins (designated *Sec-3* and, more appropriately, *Glu-R1*) are located on the long arm of chromosome 1R in rye. This is similar to the location of the orthologous genes for homologous prolamins of barley and wheat (Shewry et al. 1983).

The molecular characterisation of the *Glu-1* loci from genomic clones of different cultivars has been performed in wheat by different workers (Forde et al. 1985; Thompson et al. 1985; Anderson and Greene 1989; Halford et al. 1992). These studies have revealed that all glutenin genes are organised with a large central region containing repeated short motifs coding for the same amino acids, flanked by unique N- and C-terminal domains. The sequences of each orthologous locus of hexaploid wheat were analysed comparatively by Anderson and Greene (1989), who showed them to have similar structures. The usefulness of PCR-based analysis for distinguishing between wheats with different HMW glutenin subunits has also been demonstrated (D'Ovidio and

Anderson; 1994; D'Ovidio et al., 1994, 1995; Varghese et al. 1996; De Bustos et al. 2000, 2001). More recently, Van Campehout et al. (2000) designed primer sets from wheat consensus sequences in order to specifically amplify each of the coding regions of the orthologous loci *Em* in common wheat. The primers amplified not only the wheat genes but also those of other Triticeae species, including rye.

This paper reports the use of primers designed from nucleotide sequences of the *Glu-D1* gene that amplify orthologous sequences in the rye genome. The primers were designed making use of the sequences of the allele *Glu-D1y10*, and recognised upstream and downstream flanking positions of the coding regions of the gene. This work also describes the amplification, isolation, chromosomal assignation, cloning, sequencing and characterisation of the x and y gene subunits of the *Glu-R1* locus of rye, and their specific expression during endosperm ripening.

Material and methods

Plant material

Analyses were carried out on the following materials: *Triticum aestivum* L.: '7841' [2n=20″+1R(1D)″] and '48–136' (2n=20″+1BL-1RS″) are homogeneous selected lines that belong to a working collection of wheats derived from crosses involving the cultivars 'Anza' (2n=21″), 'Anza'/S-149″ [2n=20″+1R(1D)″], 'Amigo' (2n=20″+1AL-1RS″) and 'Glennson 81' (2n=20″+1BL-1RS″), and the hexaploid triticale 'GA-250', kindly supplied by Dr J.P. Gustafson. These lines were formerly used in experiments examining rye gene expression in a wheat background (Bittel et al. 1991, 1992; Rubio et al. 1999). The cultivars 'Chinese Spring'; the addition lines of 'Imperial' rye chromosomes into 'Chinese Spring', 'Pané-247', 'Hope', 'Ablaca' and 'Bezostaya' were also used as testers in PCR and SDS-PAGE experiments.

Additionally, samples of the following species were also used as testers in different experiments: *Secale cereale* L. cv 'Imperial' and 'Petkus'; *xTriticosecale* Wittmack cv 'GA-250'; *Triticum monococcum* L., *Triticum turgidum* L.: cv 'Jerez-137'; *Aegilops squarrosa* L.

HMW glutenin analysis

In order to analyse endosperm proteins without interfering with germination, a piece of the endosperm was removed, making sure it contained no part of the embryo. This was ground and the proteins extracted as described by Singh and Shepherd (1991). Prolamins were analysed using vertical polyacrylamide-gel electrophoresis (7.5%) in the presence of SDS (SDS-PAGE).

DNA purification and heterologous amplification using primers designed for AS-PCR markers in wheat

Genomic DNA from wheat was extracted from leaves of single plants about 4–6 weeks old as described by Sharp et al. (1988). PCR markers that amplify the complete coding sequence of the specific alleles of the HMW glutenin genes of wheat have recently been reported (De Bustos et al. 2000, 2001). A set of AS-PCR molecular markers that use primers designed from nucleotide sequences of the G*lu-D1y10* gene amplified two products when using genomic DNA from the line '7841' [2n=20″+1R(1D)″] (which lacks chromosome 1D). The sequences of the forward and reverse primers are as follows: forward: 5′ CTAACTCGCCGTG-CACA 3′; reverse: 5′ AGCTAAGGTGCATGCATG 3′.

PCR amplifications were carried out using a Perkin Elmer 480 thermocycler. Reactions were performed in a final volume of 25 μl using 2 units of *Taq* DNA polymerase (Sigma) and 10×*Taq* PCR buffer (Sigma). The quantities of other components in the reaction mixture were 100–200 ng of DNA, 200 μ M of each dNTP, and 0.8 μ M of each primer. The cycling conditions were: one cycle of 2 min at 94°C, 40 cycles of 1 min at 94°C, 40 s at 66°C, 3 min at 72°C and a final incubation cycle of 10 min at 72°C.

Amplification products were then analysed by electrophoresis in agarose gels (4.5 V/cm for 2–3 h in a TAE buffer) and by staining with ethidium bromide.

DNA cloning and sequencing

Both bands amplified by PCR using the primers G*lu-D1y10* on template DNA from '7841' were purified from the agarose gels using the QIAquick Gel Extraction Kit (QIAGEN). Each fragment was ligated into a Bluescript phagemid vector and used to transform competent cells of *Escherichia coli*. Three clones per fragment were selected for subcloning and sequencing. Each clone was further subcloned for nucleotide sequencing using the Erase-a-base kit (Promega). For this purpose, each clone was previously digested with *Kpn*I (to protect the primer binding site at the 3′ end) and *Xho*I (to leave a 5′ end adjacent to the insert from which deletions were made), and incubated at 35°C with Exonuclease III, withdrawing 1-μl aliquots at 1-min intervals. This was followed by S1 nuclease digestion. A series of different subclones were obtained from each original clone and used for sequencing in an ABI Prism 377 sequencer (Applied Biosystem). The consensus sequence was deduced using the Sequence Navigator programme (Applied Biosystem). Nucleotide sequences were aligned using the Clustal W 1.5 program (Thompson et al. 1994). Phylogenetic trees were built using the distance method. The Phylip program (version 3.5) (Felsenstein 1993) was used to estimate the distance matrices of the DNADIST program using the two-parameter method of Kimura (1980).

Southern blotting and hybridisation

Genomic DNA from the wheats '7841' and 'Chinese Spring', the rye-wheat addition lines, *T. monococcum, T. turgidum, Ae. Squarrosa*, and the rye lines 'Imperial' and 'Petkus' was digested with *Eco*RI restriction endonuclease and size-fractionated in 1% agarose gels (Maniatis et al. 1982). Fragments were blotted onto nylon membranes (Roche). In initial experiments, the clone UAHP7841–1 was labelled with digoxigenin (Roche) and used as a probe to hybridise to the genomic DNA. Further, the amplification products obtained by using the primer set with genomic DNA from the same samples were removed from the agarose gels and blotted onto a nylon membrane. The labelled clones UAHP7841–1 and UAHP7841–2 were used as probes.

RNA extraction and reverse transcriptase polymerase chain reaction (RT-PCR)

Crude extracts from immature seeds harvested at 10 days after anthesis were prepared following De Vries et al. (1988). Poly(A)+ RNA was purified by the use of DNase. Three-hundred nanograms of poly(A)+ RNA were used to obtain cDNA in an amplification reaction mixture following the recommendations of the SuperScript One-Step RT-PCR System by Gibco-BRL. PCR conditions were: one step at 45°C for 30 min, followed by a cycle at 95°C for 2 min, 35 cycles of 1 min at 94°C, 1 min at 62°C, and 1 min at 72°C. As a control, additional reactions were performed using *Taq* polymerase to detect the presence of contaminant DNA. Specific primers were designed to amplify selected regions of 330 bp and 220 bp of the presumptive mRNAs encoded by '7841–1' and '7841–2' respectively (see Figs. 5 and 6). Samples of 0.8 μM of

Table 1 Amino-acid composition of the mature protein coded by the *Glu-R1x* and *Glu-R1y*

Amino acids	*GluR1-x*	*GluR1-y*
Alanine	18	26
Arginine	11	14
Asparagine	7	1
Aspartate	6	5
Cysteine	4	8
Glutamine	250	233
Glycine	154	131
Histidine	5	14
Isoleucine	5	8
Leucine	28	24
Lysine	8	8
Methionine	1	4
Phenylalanine	4	2
Proline	108	79
Serine	39	40
Treonine	30	26
Tryptophan	4	4
Tyrosine	49	33
Valine	10	13
Total	745	692
Mol. Wt	81, 339.97	76, 122.45
Isoelectric point	5.97	6.26

the following degenerated primers were used: forward: 5′ CCA-ACTTCTC(GC)GCAGCAAC 3′; reverse: 5′ CCGC(AG)CTAAC-ATGG(CT)ATG 3′. The PCR products were separated on 1.8% agarose gels.

Results

Heterologous amplification

In a previous paper the authors reported the use of primers designed to amplify the complete coding regions of a series of different alleles (AS-PCR) of the x and y subunits of genes *Glu-A1* and *Glu-D1* (De Bustos et al. 2000, 2001). The study of sequences external to the coding region of these genes helped resolve difficulties in designing primers with the capacity to distinguish genes and alleles of the highly homologous members of the *Glu-1* complex loci of wheat. This difficulty is greater for the coding sequence than for the flanking regions due to the existence of repetitive domains in all HMW glutenin genes, and because the N-terminal and C-terminal regions are highly conserved.

Electrophoretic analyses of the PCR reactions performed on the genomic DNA from the line ‘7841’ using the *Glu-D1y10* set of primers, resulted in a clear production of two DNA bands designated ‘7841–1’ and ‘7841–2’ (2,490 bp and 2,331 bp respectively) (Fig. 1). The same primers amplify a lower molecular weight product of 2,135 bp in the line ‘48–136’ and the wheat cultivars ‘Hope’, ‘Ablaca’ and ‘Bezostaya’ (which have the *GluD1-y10* allele), but form no visible product with wheat cultivars that lack this allele (data not shown). Primers designed to amplify the alternative allele *Glu-D1y12* showed a single product of 2,190 bp in

Fig. 1 Separation of the amplification products on agarose gels using the designed *Glu-Dy10* primers. From left to right are included the molecular-weight markers and the products of amplification in the wheat lines ‘48–136’ and ‘7841’

‘Pané-247’ and ‘Chinese Spring’, but no visible amplification products with the line ‘7841’ or the cultivars Hope’, ‘Ablaca’ and ‘Bezostaya’ (results reported in De Bustos et al. 2000).

Products of the same size as ‘7841–1’ and ‘7841–2’ were generated in different rye cultivars. This analysis demonstrates that all rye lines assayed display the prominent band ‘7841–1’. However, size polymorphism was shown by ‘7841–1’ in different rye cultivars, and polymorphism for the presence-absence of ‘7841–2’ was seen in ‘Petkus’ As the line ‘7841’ has chromosome 1D substituted by 1R, and the primers were designed to amplify the y subunit of gene *Glu-D1* located in the long arm of 1D, it was presumed that ‘7841–1’ and ‘7841–2’ might represent sequences located in 1R, and that these sequences might be related to the *Glu-1* gene family. This encouraged the undertaking of experiments that might confirm this idea. The first approach consisted of direct sequencing starting from the ends of the amplification products. For this, each single band was sliced from the agarose gels, extracted, purified and directly sequenced in both directions using the designed primers. In all cases, direct sequencing of about 350 bases from the ends of each fragment gave a sequence highly homologous to that reported by Anderson and Greene (1989) for HMW glutenin subunits coded for by *Glu-1* of wheat. This seems to confirm that the amplified products are members of the *Glu-1* family.

Cloning and probing the amplified sequences

The next step was to clone the ‘7841–1’ and ‘7841–2’ amplification products. After using the primers with template DNA from ‘7841’, the DNA fragments were purified from the agarose gels. Both sequences were ligated into the Bluescript phagemid vector and used to transform competent *E. coli* cells.

The plasmids with the insert were labelled and assayed by Southern blotting with genomic DNA of

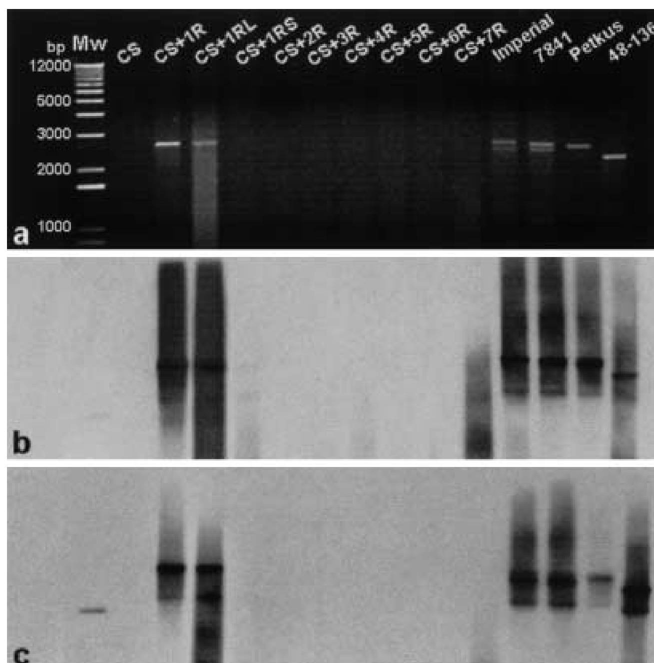

Fig. 2 a Separation of amplification products on agarose gel (1.8%) using the designed *Glu-D-y10* primers. From left to right are included the PCR products of disomis and ditelosomic addition lines of rye chromosomes in 'Chinese Spring', the rye cultivars 'Imperial' and 'Petkus' and the wheat lines '7841' and '48–136' The band present in '48–136' belongs to the amplification of *Glu-D-y10* allele that is present in this wheat. The products present in '7841', 'Imperial', 'Petkus', and the addition lines of either 1R or 1RL correspond to the glutenin genes located at 1RL. **b** and **c** show the DNA products of the PCR experiment after blotting on nylon membrane and alternatively hybridised with the plasmids carrying the inserts '7841–1' **(a)** and '7841–2' **(b)**

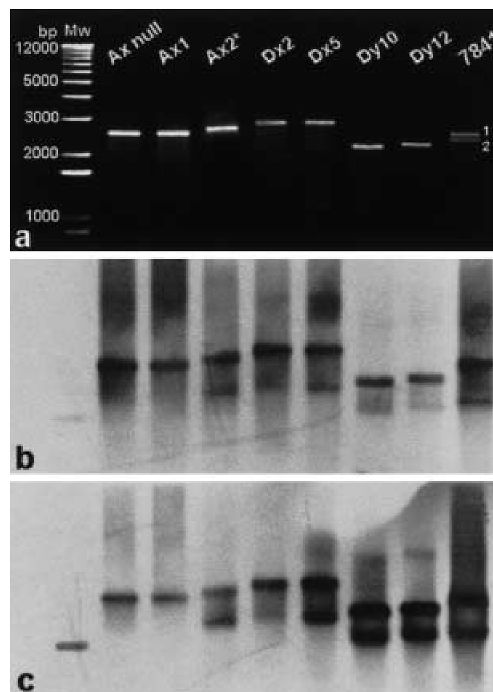

Fig. 4 a Separation of specific amplification products on agarose gels (1.8%) using sets of designed primers (AS-PCR) in genomic DNA from the following wheats: *Glu-A1xNull* in 'Pané 247', *Glu-A1×1* in 'Hope', *Glu-A1×2** in 'Ablaca', *Glu–D1×2* and *Glu–D1y12* in 'Chinese Spring', *Glu–D1×5* and *Glu–D1y10* in 'Hope'. The genomic DNA from the line '7841' was amplified with the designed primers *Glu–D1y10.* **b** and **c** show the DNA products of the PCR experiment after blotting on nylon membrane and alternatively hybridised with the plasmids carrying the inserts '7841–1' **(b)** and '7841–2' **(c)** labelled with digoxigenin. '7841–1' hybridised preferentially with x-type wheat sequences, and '7841–2' was associated preferentially with y-type sequences

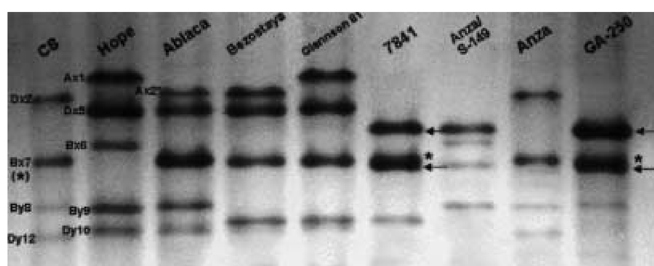

Fig. 3 HMW separation by SDS-PAGE. From letf to right: patterns of several wheats used as testers for different glutenin combinations, the line '7841', the wheat lines 'Glennson 81', 'Anza/S-149' and 'Anza' and the triticale 'GA-250'. This triticale line is the donor of the chromosome 1R present in '7841'. The *arrows* on the right indicate the rye *Secale* lines corresponding to the glutenin gene located in 1RL. The main glutenin subunits have been indicated. The *asterisks* indicate the glutenin coded by *Glu-B1×7*

T. aestivum '7841' and 'Chinese Spring', the rye-wheat addition lines, *T. monococcum, T. turgidum* and *Ae. squarrosa*, and the rye cultivars 'Imperial' and 'Petkus'. *Eco*RI restrictions of total DNA digested from all the mentioned materials were separated on agarose gels, transferred to a membrane, and then probed with '7841–1' and '7841–2'. The probes hybridised to different fragments of all the genomic DNA samples used, owing to the partial homology of the cloned sequences to the sequences of the glutenin gene family present in all the plant genomes.

In order to assign the amplified sequences to specific chromosomes, genomic DNA from 'Chinese Spring', an aneuploid stock of addition lines of 'Imperial' in a 'Chinese Spring' background, as well as the rye testers were amplified using the designed *Glu-D-y10* primers. The rye cultivars and addition lines in which the long arm 1RL was present amplified one or both products '7841–1' and '7841–2', showing polymorphism for the presence-absence of '7841–2' (Fig. 2a). The products of amplification were transferred to a membrane and alternatively probed with the plasmids carrying the inserts '7841–1' and '7841–2' (Fig. 2b and c). As expected, hybridisation of the labelled probes identified both amplified products of the rye cultivars and the aneuploid lines of 'Chinese Spring' with the chromosome arm 1RL, but were absent in Southern profiles of 'Chinese Spring' and all the other aneuploid lines.

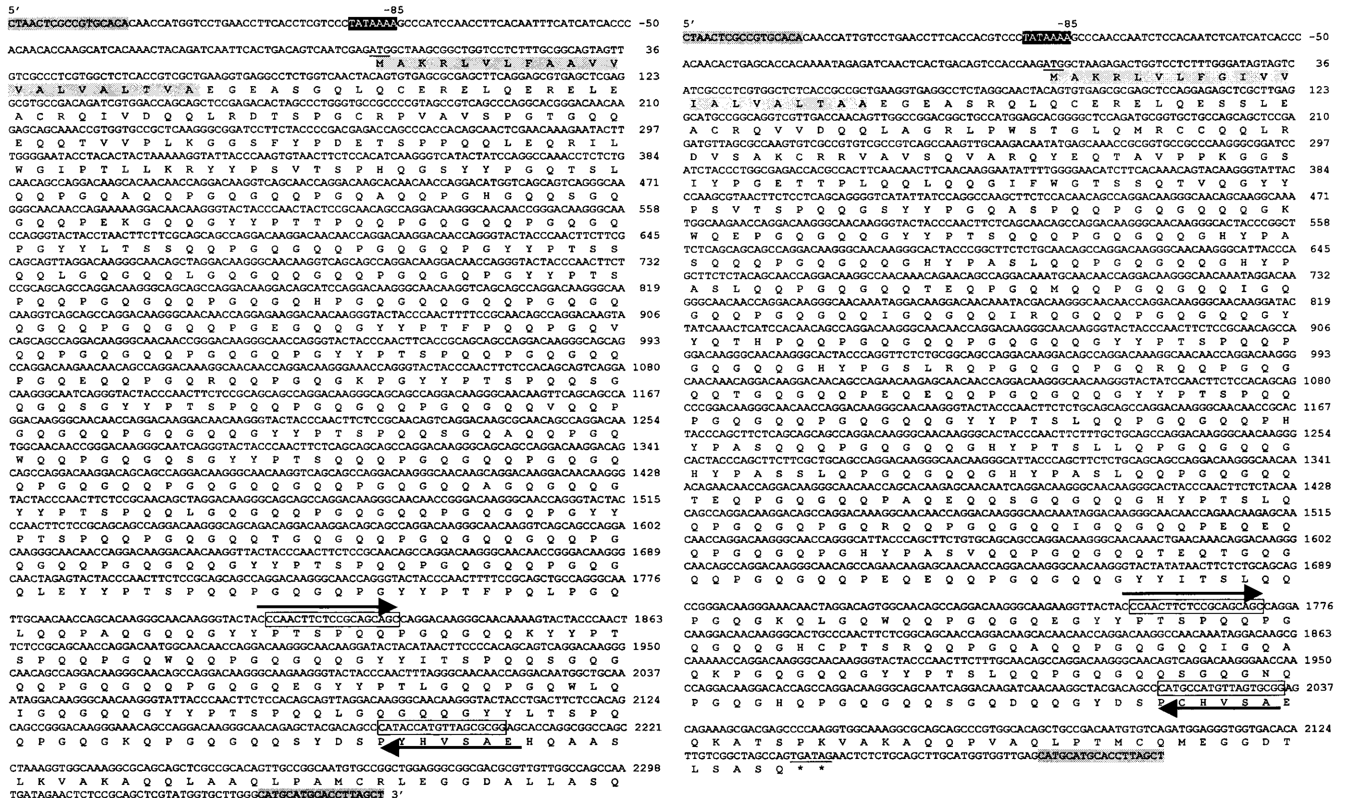

Fig. 5 The nucleotide sequences of the coding regions and deduced amino acids of the rye glutenins coded by *Glu-R1x* (left) and *Glu-R1y* (right). The dark grey boxes indicate the sites of the primers used to amplify both sequences. The *black box* signals the position of the TATA box. The *clear grey box* correspondes to the signal peptide. *Arrows* at the end of the sequences indicate the position of the designed primers used to perform the RT-PCR experiment to detect expression of both genes in the endosperm. The sequences have been deposited in the EMBL GenBank and DDBJ under accession numbers AF216868 and AF216869

Glutenin nature of the '7841–1' and '7841–2' sequences

Electrophoretic analyses of HMW glutenins were performed using the selected line '7841', their original parentals and other wheat varieties, to identify the gene markers coding for the HMW glutenins. The wheat '7841' showed a distinctive profile different to that from the remaining wheats of known HMW composition. Moreover, '7841' presented the same profile as 'Anza/S-149' and triticale 'GA-250' which is the putative donor of the 1R chromosome present in this line (Fig. 3).

The next approach consisted of demonstrating the relationship of the amplified products '7841–1' and '7841–2' with specific sequences coding for the HMW glutenins of wheat. The plasmids containing each fragment were used as probes in Southern experiments to test homology with the PCR products obtained using specific primers previously reported as AS-PCR (De Bustos et al. 2000, 2001) which amplify the alleles *xNull*, *x1* and *x2** (*Glu–A1*), *x2, x5, y10* and *y12* (*Glu–D1*). Samples of genomic DNA of different wheats with different allelic combinations were amplified with each specific set of primers (AS-PCR). The amplification products were separated on agarose gels and transferred to a membrane for hybridisation with the labelled probes '7841–1' and '7841–2' (Fig. 4a). The long sequence '7841–1' hybridised preferentially with x-type wheat sequences, while the short sequence '7841–2' was associated preferentially with y-type sequences (Fig. 4b and c).

Sequence analysis

Three clones for each amplification product of '7841–1' and '7841–2' were used in subcloning and sequencing. The consensus sequence of the three isolates for each amplification product of '7841–1' and '7841–2' was considered. The complete nucleotide sequence and the deduced amino-acid sequence of the coding region of each subunit are shown in Fig. 5, and have been deposited in the EMBL GenBank and DDBJ under accession numbers AF216868 and AF216869. Comparison with amino-acid sequences in GenBank revealed that the proteins coded by '7841–1' and '7841–2' are homologous to a group of glutenin genes of wheat and related species.

The coding region of both subunits lacks introns and show the same general structure as other *Glu-1* sequences previously investigated (Anderson and Greene 1989). Both sequences contain TATAAAA between 85

```
GLU-Dy10       MAKRLVLFAAVVIALVALTTAEGEASR----------QLQCERELQESS---LEACRQVVDQQLAGRLPWSTGLQMRCCQQLRDVSAKCRSVAVSQVARQYEQTVVPP-KGGSFYPGETTPLQQLQQGIFWGTS
Ae-squarro     MAKRLVLFAAVVIALVALTTAEGEASR----------QLQCERELQESS---LEACRQVVDQQLAGRLPWSTGLQMRCCQQLRDVSAKCRSVAISQVARQYEQTVVPP-KGGSFYPGETTPLQQLQQGIFWGTS
GLU-Dy12       MAKRLVLFAAVVIALVALTTAEGEASR----------QLQCERELQESS---LEACRQVVDQQLAGRLPWSTGLQMRCCQQLRDVSAKCRSVAVSQVARQYEQTVVPP-KGGSFYPGETTPLQQLQQGIFWGTS
GLU-By9        MAKRLVLFATVVIALVALTTAEGEASR----------QLQCERELQESS---LEACRQVVDQQLAGRLPWSTGLQMRCCQQLRDVSAKCRPVAVSQVVRQYEQTVVPP-KGGSFYPGETTPLQQLQQVIFWGTS
GluR1y-7841-2  MAKRLVLFGIVVIALVALTAAEGEASR----------QLQCERELQESS---LEACRQVVDQQLAGRLPWSTGLQMRCCQQLRDVSAKCRRVAVSQVARQYEQTAVPP-KGGSIYPGETTPLQQLQQGIFWGTS
GLU-Ax1        MTKRLVLFAAVVVALVALTAAEGEASG----------QLQCERELQEHS---LKACRQVVDQ------------------QLRDVSPECQPVGGGPVARQYEQQVVVPPKGGSFYPGETTPPQQLQQSILWGIP
GLU-Ax2*       MTKRLVLFAAVVVALVALTAAEGEASG----------QLQCERELQEHS---LKACRQVVDQ------------------QLRDVSPECQPVGGGPVARQYEQQVVVPPKGGSFYPGETTPPQQLQQSILWGIP
GLU-Dx2        MAKRLVLFVAVVVALVALTVAEGEASE----------QLQCERELQELQERELKACQQVMDQ------------------QLRDISPECHPVVVSPVAGQYEQQIVVP-KGGSFYPGETTPPQQLQQRIFWGIP
GLU-Dx5        MAKRLVLFVAVVVALVALTVAEGEASE----------QLQCERELQELQERELKACQQVMDQ------------------QLRDISPECHPVVVSPVAGQYEQQIVVPPKGGSFYPGETTPPQQLQQRIFWGIP
GLU-Bx17       MTKRLVLFAAVVVALVALTAAEGEASG----------QLQCEHEL--------EACQQVVDQ------------------QLRDVSPGCRPITVSPGTRQYEQQPVVASKAGSFYPSETTPSQQLQQMIFWGIP
GLU-Bx7        MAKRLVLFAAVVVALVALTAAEGEASG----------QLQCEHEL--------EACQQVVDQ------------------QLRDVSPGCRPITVSPGTRQYEQQPVVPSKAGSFYPSETTPSQQLQQMIFWGIP
GluR1x-7841-1  MAKRLVLFAAVVVALVALTVAEGEASG----------QLQCERELQERE---LEACRQIVDQ------------------QLRDTSPGCRPVAVSPGTGQQEQQTVVFLKGGSFPDETSPPQQLEQRILWGIP
H-vulgare      MAKRLVLFVAVIVALVALTTAEREINGNNIFLDSRSRQLQCERELQESS---LEACRRVVDQQLVGQLPWSTGLQMQCCQQLRDVSPECRPVALSQVVRQYEQQTEVPSKGGSFYPGGTAPP--LQQGGWWGTS

SQTVQGYYPGVTSPRQ---GS--YYPGQASPQQPG------QGQQPGKWQ---------EPGQGQQ---------------WYYPTSLQQ---------------------PGQG----QQ--------------------IGKG
SQTVQGYYPGVTSPRQ---GS--YYPGQASPQQPG------QGQQPGKWQ---------EPGQGQQ---------------WYYPTSLQQ---------------------PGQG----QQ--------------------IGKG
SQTVQGYYPSVTSPRQ---GS--YYPGQASPQQPG------QGQQPGKWQ---------EPGQGQQ---------------WYYPTSLQQ---------------------PGQG----QQ--------------------IGKG
SQTVQGYYPSVSSPQQ---GP--YYPGQASPQQPG------QGQQPGKWQ---------ELGQGQQGYYPTSLHQSGQGQQGYYPSSLQQ---------------------PGQG----QQ--------------------IGQG
SQTVQGYYPSVTSPHQ---GS--YYPGQASPQQPG------QGQQGKWQ----------EPGQGQQ---------------GYYPTSQQQ---------------------PGQG----QQGHYPAS----------Q-QQPGQG
-ALLRRYYLSVTSPQQ---VS--YYPGQASSQRPG------QGQQPGQGQ------QEYYLTSPQQSG---QWQQPGQGQAGYYPTSPQQSGQEQPGYYPTSPWQPEQLQQPTQGQQRQQPGQGQQL------RQGQQGQQSGQG
-ALLRRYYLSVTSPQQ---VS--YYPGQASSQRPG------QGQQ------------EYYLTSPQQSG---QWQQPGQGQQSGYYPTSPQQSGQKQPGYYPTSPWQPEQLQQPTQGQQRQQPGQGQQL------RQGQQGQQSGQG
-ALLKRYYPSVTSPQQ---VS--YYPGQASPQRPG------QGQQPGQGQQGQGQQPGYYPTSPQQPG---QWQQPGQGQPGYYPTSPQQ-----PGQ----------LQQPAQG---QQPGQGQ------------GQRQPGQG
-ALLKRYYPSVTCPQQ---VS--YYPGQASPQRPG------QGQQPGQGQ---------QGYYPTSPQQPG---QWEEPQGQQGYYPTSPQQ-----PGQ----------LQQPAQG---QQPGQGQ------------GQQQPGQG
-ALLRRYYPSVTSSQQ---GS--YYPGQASPQQSG------QGQQPGQEQ---------Q-PGQGQQD------QQPGQGRQGGYYPTSPQQ-----PGQG----------QQLGQG----QPGYYPTSQQPGQKQQAGQGQQSGQG
-ALLRRYYPSVTSSQQ---GS--YYPGQASPQQSG------QGQQPGQEQ---------Q-PGQGQQD------QQPGQGRQGGYYPTSPQQ-----PGQG----------QQLGQG----QPGYYPTSQQPGQKQQAGQGQQSGQG
-TLLKRYYPSVTSPHQ---GS--YYPGQTSLQQPGQAQQPGQGQQPGQAQ---------Q-PGHGQQSG---QGQQPEKGQQGYYPTTPQQ-----PGQG------------QQPGQG---QPGYYLTS-----SQQPGQGQQPGQG
---VKWYYPDQTSSQQSWQGQQGYHQSVTSSQQPG-----QQGQ-GSYPGSTFP---QQPGQGQQ--------PGQRQPWSYPSATFP------------------------QQPGQG------------------------QG

QQGYYPTSLQQ--PGQGQ--------------QGYYPTSLQHTGQ---------RQQPVGQG--------------------QPEQGQQ---------PGQWQQGYYP--TSPQQLGQGQQ---PRQW--QQSGQGQQGHYPT
QQGYYPTSLQQ--PGQGQ--------------QGYYPTSLQHTGQ---------RQQPVGQG--------------------QPEQGQQ---------PGQWQQGYYP--TSPQQLGQGQQ---PGQW--QQSGQGQQGHYPT
KQGYYPTSLQQ--PGQGQQI---------QGQQGYYPTSPQHTGQ---------RQQPVGQGQQI---GQGQ--------------QPEQGQQ---------PGQWQQGYYP--TSPQQLGQGQQ---PGQW--QQSGQGQQGHYPT
QQGYYPTSLQQ--PGQGQQI---------GQGQQGYYPTSPQHPGQ---------RQQPGQGQQI---GQGQ-------LGQGRQIGQGQQ---------SGQGQQGYYP--TSPQQLGQGQQ---PGQW--QQSGQGQQGYYPT
QQGHYPASLQQ--PGQQ---------------QGHYPASLQQPGQGQQTEQPGQMQQPGQGQQI---GQGQQPGQGQQIGQGQQIRQGQQ---------PGQGQQGYYQ--THPQQPGQGQQ---P-----GQGQQGYYPT
QPRYYPTSSQQ--PGQLQQLAQGQQGQQPERGQQGQQSGQGQQLGQGQQ------GQQPGQGKQQS---GQGQQGYYPI---SPQQLGQGQQ----------SGQGQLGYYP--TSPQQSGQGQSGYYPTSA--QQPGQLQQSTQEQ
QPRYYPTSSQQ--PGQLQQLAQGQQGQQPERGQQGQQSGQGQQLGQGQQ------GQQPGQGKQQS---GQGQQGYYPI---SPQQLGQGQQ----------SGQGQLGYYP--TSPQQSGQGQSGYYPTSA--QQPGQLQQSTQEQ
QPGYYPTSSQLQ-PGQLQQPAQGQQGQQPGQGQQPGQAQQGQQPGQGQQPGQGQQ------GQQPGQGQQP---GQGQQG---------QQLGQGQQGYYPTSLQQSGQGQPGYYP--TSLQQLGQGQSGYYPTSP--QQPGQGQQPG---
QPGYYPTSSQLQ-PGQLQQPAQGQQGQQPGQAQQGQQPGQGQQPGQGQQ------GQQPGQGQQP---GQGQQG---------QQLGQGQQGYYPTSLQQSGQGQPGYYP--TSLQQLGQGQSGYYPTSP--QQPGQGQQPG---
QQGYYPTSPQQ--SGQGQ---------QPGQGQPGYYPTSPQQSGQWQQP---GQGQQPGQGQQS---GQGQQG---------QQPGQGQR---------PGQGQQGYYP--ISPQQPGQGQQ------------GQ--QQSGQGQPGYYPT
QQGYYPTSPQQ--SGQGQ---------QPGQGQPGYYPTSPQQSGQWQQP---GQGQQPGQGQQS---GQGQQG---------QQPGQGQR---------PGQGQQGYYP--ISPQQPG------------------QG--QQSGQGQPGYYPT
QPGYYPTSSQQ--LGQGQQLGQGQQGQQPGQGQQPGQGQQPGYYPTSPQQPGQP---GQQHPGQGQQPGQGQQPGQGQQ---------QQPGQGQQ---------PGEGQQGYYP--TFPQQPGQVQQ---PGQG--QQPGQGQQGYYPT
QQGYYPGATSLLQPGQGQ------------QG-PYQSATSPQQPGQ------------GQGQQE----TYPIATSP------HQPGQWQQ---------PGQGQQGFYPSVTSPQQSGQGQQG-YPSTTSPQQSGQGQQLG---

SLQQPGQGQQGHYLASQ-QQPGQGQ--------------QGHYPASQ--QQPGQGQQGHYP--ASQQQPGQGQQ------GHYPASQQ-----EPGQGQQG-------------------QIPASQQ---QPGQ
SLQQPGQGQQGHYLASQ-QQPGQGQ--------------QGHYPASQ--QQPGQGQQGHYP--ASQQQPGQGQQ------GHYPASQQ-----EPGQGQQG-------------------QIPASQQ---QPGQ
SLQQPGQGQQGHYLASQ-QQPAQGQ--------------QGHYPASQ--QQPGQGQQGHYP--ASQQQPGQGQQ------GHYPASQQ-----EPGQGQQG-------------------QIPASQQ---QPGQ
SQQQPGQGQQGQYPASQ-QQPGQGQ--------------QGQYPASQ--QQPGQGQQGQYP--ASQQQPGQGQQ------GHYLASQQ-----QPGQGQQR-------------------HYPASLQ---QPGQ
S---P---QQPGQG---QQPGQEQ--------------Q---PGQR--QQPGQGKPGYYP--TSPQQPGQGQQ------GQGYYPTSPQ------QPGQGQQQ------------------GQGQQ----VQ--QPGQGQ--QPGQ
QLGQEQQDQQSGQGRQ-QQSGQRQ--------------QDQQSGQG--QQPGQRQPGYYS--TSPQQLGQG------QPRYYPTSPQ------QPGQEQQPRQLQQPEQGQGQQPEQGQGQQPGQGEQGQQPGQGQQGQPGQ
QLGQEQQDQQSGQGRQ-QQSGQRQ--------------QDQQSGQG--QQPGQRQPGYYS--TSPQQLGQG------QPRYYPTSPQ------QPGQEQQPRQLQQPEQGQGQQPEQGQGQQRQEQGQQPGQGQQGQPGQ
QLQQPAQGQQPEQGQQG-QQPGQGQ--------------QGQQPGQG--QQPGQGQPGYYP--TSPQQSGQG------QPGYYPTSSQ-----QPTQSQQP-----------GQ----GQQGQQVGQGQQAQQPGQGQ---QPGQ
QLQQPAQGQQPGQGQQG-QQPGQGQ--------------QGQQPGQG--QQPGQGQPGYYP--TSPQQSGQG------QPGYYPTSSQ-----QPTQSQQP-----------GQ----GQQGQQVGQGQQAQQPGQGQ---QPGQ
SLRQPGQGQWQQPGQG----QQPGQGQ--------------QGQQPGQG--QQSGQGQQGYYP--TSLQQPGQGQQLGQGQPGYYPTSQQSEQGQQPGQPKQP----------------GQGQQGYYPTSPQQSGQGQ---QLGQ
SLRQPGQGQWQQPGQG----QQPGQGQ--------------QGQQPGQG--QQSGQGQQGYYP--TSLQQPGQGQQLGQGQPGYYPTSQQSEQGQQPGQGKQP----------------GQGQQGYYPTSPQQSGQGQ---QLGQ
SPQQPGQGQQGHYPGSL-RQPGQGQPGQRQQPGQGQQTGQGQQPEQE--QQPGQGQQGYYP--TSPQQPGQGQQPGQGQQGYYPTSLQ-----QPGQGQQP-------------------------HYPASQQ---QPGQ
QGQQPGQGQQGYPSATFPQQPGQWQ--------------QGSYPSTTSPQQSGQGQQGYNPSGTSTQQPGQVQQLGQGQQGYYPIATSP---QQPGQGQQLG-----------------QGQQPG---HGQQLVGQ---QQGQ

GQQGHYP--ASLQQPG------------QGQQGHYPTSL--QQLGQGQQTGQPGQ------KQQPGQGQQTGQ-GQ------QP------------EQEQQ---------PGQGQ-QGYYPTSLQQ------PGQGQQQGQGQ--
GQQGHYP--ASLQQPG------------QGQQGHYPTSL--QQLGQGQQTGQPGQ------KQQPGQGQQTGQ-GQ------QP------------EQEQQ---------PGQGQ-QGYYPTSLQQ------PGQGQQQGQGQ--
GQQGHYP--ASLQQPG------------Q--QGHYPTSL--QQLGQGQQIGQPGQ------KQQPGQGQQTGQ-GQ------QP------------EQEQQ---------PGQGQ-QGYYPTSLQQ------PGQGQQQGQGQ--
GQQGHYT--ASLQQPG------------QGQQGHYPASL--QQVGQGQQIGQLGQ------RQQPGQGQQTRQ-GQQLEQGQQPGQGQQTRQGQQLEQGQQ---------PGQGQ-QGYYPTSPQQSGQGQQPGQSQQPGQGQ--
GQQGHYP--TSLLQPG------------QGQQGHYPASS--LQPGQGQQGHYPAS------LQQPGQGQQTEQPGQ----GQQP------------AQEQQ---------SGQGQ-QGHYPTSLQQPGQGQ-PGQRQQPGQGQQI
GQPGYYP--TSPQQSG------------QGQPGYYPTSP--QQSGQ--------------LQQPAQGQQPGQ-EQQ---GQQP------------GQGQQGQQPGQGQQPGQGQ-PGYYPTSPQQSGQEQQLEQWQQSGQGQ--
GQPGYYP--TSPQQSG------------QGQPGYYPTSP--QQSGQ--------------LQQPAQGQQPGQ-EQQ---GQQP------------GQGQQ---------PGQGQ-PGYYPTSPQQSGQEQQLEQWQQSGQGQ--
GQPGYYP--TSPLQSG------------QGQPGYYLTSP--QQSGQGQQPGQLQQSAQGQKGQQPGQGQQPGQ-GQQ---GQQP------------GQGQQGQQ------PGQGQ-PGYYPTSPQQSGQGQQPGQWQQPGQGQ--
GQPGYYP--TSPQQSG------------QGQPGYYLTSP--QQSGQGQQPGQLQQSAQGQKGQQPGQGQQPGQ-GQQ---GQQP------------GQGQQGQQ------PGQGQ-PGYYPTSPQQSGQGQQPGQWQQPGQGQ--
GQPGYYP--TSPQQSGQGQQSG------QGQQGYYPTSP--QQSGQGQQPGQGQSGYFPTSRQQSGQGQQPGQ-GQQ---SGQG------------QQGQQ---------PGQGQ-QAYYPTSSQQSRQRQQAGQWQRPGQGQ--
GQPGYYP--TSPQQSGQGQQSG------QGQQGYYPTSP--QQSGQGQQPGQGQSGYFPTSRQQSGQGQQPGQ-GQQ---SGQG------------QQGQQ---------PGQGQ-QAYYPTSSQQSRQRQQAGQWQRPGQGQ--
GQQGYYP--TSPQQSGQAQQPGQWQQPGQGQSGYYPTSQ--QQPGQGQQPGQGQ----------QPGQGQQPGQ-GQQ---GQQP------------GQGQQ---------AGQGQ-QGYYPTSPQQLGQGQQPGQGQQPGQGQ--
GQQGHYPSMTSPHQTG------------QGQKGYYPSAISPQQSGQGQQGYQPSG------ASSQGSVQGACQHSTS--SPQQQ------------AQGCQASS------PKQGLGSLYYPSGAYTQ---QKPGQGYNP--G---

-QGYYPTSLQQPGQGQQGHYPASLQQPGQGQ----------PGQRQQ----PGQGQH--------------------------------------------------------------------------PEQGKQ-------
-QGYYPTSLQQPGQGQQGHYPASLQQPGQGQ----------PGQRQQ----PGQGQH--------------------------------------------------------------------------PEQGKQ-------
-QGYYPTSLQQPGQGQQGHYPASLQQPGQGQ--------GQPGQRQQ----PGQGQH--------------------------------------------------------------------------PEQGQQ-------
-QGYYSSSLQQPGQGLQGHYPASLQQPGQGH----------PGQRQQ----PGQGQQ--------------------------------------------------------------------------PEQGQQ-------
GQGQQPEQEQQPGQGQPGHYPASVQQPGQGQQ----TE--QTGQGQQ----PGQGQQ--------------------------------------------------------------------------PEQEQQ-------
-PGHYPTSPLQPGQGQPGYYPTSPQQIGQGQ---------QPGQLQQ----PTQGQQGQ------------------------------------QPGQGQQ---------GQQPGQG---------QQGQQPGQGQQ-------
-PGHYPTSPLQPGQGQPGYYPTSPQQIGQGQ---------QPGQLQQ----PTQGQQGQ------------------------------------QPGQGQQ---------GQQPGEG---------QQGQQPGQGQQ-------
-PGYYPTSPLQPGQGQPGYDPTSPQQPGQGQ---------QPGQLQQ----PAQGQQGQQLAQGQQGQQPAQVQQGQQPAQGQQGQQLGQGQQGQQPGQGQQ---PAQGQQGQQPGQG---------QQGQQPGQGQQ-------
-PGYYPTSPLQPGQGQPGYDPTSPQQPGQGQ---------QPGQLQQ----PAQGQQGQQLAQGQQGQQPAQVQQGQQPAQGQQGQQLGQGQQGQQPGQGQQGQQPAQGQQGQQPGQGQHGQQPGQGQQGQQPGQGQQ-------
-PGYYPTSPQQPGQEQ---QSGQAQQSGQWQLVYYPTSPQQPGQLQQ----PAQGQQPAQ-----------------------------------GQQSAQEQQ------PGQAQQSGQW---------QLVYYPTSPQQ-------
-PGYYPTSPQQPGQEQ---QSGQAQQSGQWQLVYYPTSPQQPGQLQQ----PAQGQQPAQ-----------------------------------GQQSAQEQQ------PGQAQQSGQW---------QLVYYPTSPQQPGQLQQP
-PGYYPTSPQQPGQGQ---QTGQGQQPGQGQ------QG-QQPGQGQQ----PGQGQQ-------------------------------------GYYPTSPQQ------PGQGQQPGQG---------QLEYYPTSPQQ-------
--GTSPLHQQGGGFGG----GLTTEQPQGGK---------QPFHCQQTTVSPHQGQQ-----------------------------------------------------------------------TTVSPHQGQQ------

-----------------------------PGQGQQGYYPTSPQQPGQGQQLG------------QGQQ---------------GYYPTSPQQPGQGQQPG------QGQQGHCPTSPQQSGQAQQPG--------QG-------Q
-----------------------------PGQGQQGYYPTSPQQPGQGQQLG------------QGQQ---------------GYYPTSPQQPGQGQQPG------QGQHGHCPTSPQQSGQAQQPG--------QG-------Q
-----------------------------PGQGQQGYYPTSPQQPGQGQQLG------------QGQQ---------------GYYPTSPQQPGQGQQPG------QGQQGHCPMSPQQTGQAQQLG--------QG-------Q
-----------------------------PGQGQQGYYPTSPQQPGQGKQLG------------QGQQ---------------GYYPTSPQQPGQGQQPG------QGQQGHCPTSPQQTGQAQQPG--------QG-------Q
-----------------------------PGQGQQGYYITSLQQPGQGKQLG------------QWQQP----------GQGQEGYYPTSPQQPGQGQQ-------------GHCPTSRQQPGQAQQPG--------QG-------Q
-----------------------------PGQGQPGYYPTSLQQSGQGQQPG------QWQQPGQGLPGYYPTSSLQPEQGQQGYYPTSQQQPGQGPQPGQWQQSGQGQQGYYPTSPQQSGQGQQPG------------------
-----------------------------PGQGQPGYYPTSLQQSGQGQQPG------QWQQPGQGQPGYYPTSSLQPEQGQQGYYPTSQQQPGQGPQPGQWQQSGQGQQGYYPTSPQQSGQGQQPG------------------
-----------------------------PGQGQPWYYPTSPQESGQGQQPGQWQQPGQWQQPGQGQPGYYLTSPLQLGQGQQGYYPTSLQQPGQGQQPGQWQQSGQGQHGYYPTSPQLSGQGQRPG------------------
-----------------------------PGQGQPWYYPTSPQESGQGQQPG------QWQQPGQGQPGYYLTFSVAARTGQQGYYPTSLQQPGQGQQPGQWQQSGQGQHWYYPTSPKLSGQGQRPG------------------
-----------------------------SGQGQQGYYPTSPQQSGQGQQPG-------Q-----GQQ---------PRQGQQGYYPISPQQSGQGQQPG------QGQQGYYPTSPQQSGQGQQPG--------HE-------Q
AQGQQGYYPTSPQQSGQGQQGYYPTSPQQSGQGQQGYYPTSPQQSGQGQQPG-------Q-----GQQ---------PRQGQQGYYPISPQQSGQGQQPG------QGQQGYYPTSPQQSGQGQQPG--------HE-------Q
-----------------------------PGQGQPGYYPTFPQLPGQLQQPA-------Q-----GQQGYYPTSPQQPGQGQQKYYPTSPQQPGQWQQPG------QGQQGYYITSPQQSGQGQQPGQGQQPGQGQEGYYPTLGQ
-------------------------TTVSPHQGQQ--TTVSPHQ-GQQTTVS-----------PHQGQQ--------------TTVSPHQGQQTTVSPHPG--------Q--QTTVSPHQ---GQQTT------------------

QIGQVQQPGQGQQ-GYYPTSVQQPGQGQQ------SGQGQQ---SGQGHQPGQGQ---------------QSGQEQQGYDSPYHVSAEQQAASPMVAKAQQPATQLPTVCRMEGGDALSASQ
QIGQVQQPGQGQQ-GYYPTSLQQPGQGQQ------SGQGQQ---SGQGHQPGQGQ---------------QSGQEQQGYDSPYHVSAEQQAASPMVAKAQQPATQLPTVCRMEGGDALSASQ
QIGQVQQPGQGQQ-GYYPTSLQQPGQGQQ------SGQGQQ---SGQGHQPGQGQ---------------QSGQEKQGYDSPYHVSAEQQAASPMVAKAQQPATQLPTVCRMEGGDALSASQ
QIGQVQQPGQGQQ-GYYPISLQQSGQGQQ------SGQGQQ---SGQGHQLGQGQ---------------QSGQEQQGYDNPYHVNTEQQTASPKVAKVQQPATQLPIMCRMEGGDALSASQ
QIGQAQKPGQGQQ-GYYPTSLQQPGQGQQ------SGQGNQ---PGQGHQPGQGQ---------------QSGQDQQGYDSPCHVSAEQKATSPKVAKAQQPVAQLPTMCQMEGGDTLSASQ
---QWLQPGQWLQSGYYLTSPQQLGQGQQPRQWLQPRQGQQGYYPTSPQQSGQGQQLGQGQQGYYPTSPQQSGQGQQGYDSPYHVSAEHQAASLKVAKAQQLAAQLPAMCRLEGGDALLASQ
---QWLQPGQWLQSGYYLTSPQQLGQGQQPRQWLQPRQGQQGYYPTSPQQSGQGQQLGQGQQGYYPTSPQQSGQGQQGYDSPYHVSAEHQAASLKVAKAQQLAAQLPAMCRLEGGDALLASQ
---QWLQPGQGQQ-GYYPTSPQQSGQGQQLGQWLQPGQGQQGYYPTSLQQTGQGQ---------------QSGQGQQGYYSSYHVSVEHQAASLKVAKAQQLAAQLPAMCRLEGGDALSASQ
---QWLQPGQGQQ-GYYPTSPQQPPQGQQLGQWLQPGQGQQGYYPTSLQQTGQGQ---------------QSGQGQQGYYSSYHVSVEHQAASLKVAKAQQLAAQLPAMCRLEGGDALSASQ
QPGQWLQPGQGQQ-GYYPTSSQQSGQGHQ------SGQGQQGYYPTSLWQPGQG--------------------QQGYASPYHVSAEYQAARLKVAKAQQLAAQLPAMCRLEGSDALSTRQ
QPGQWLQPGQGQQ-GYYPTSSQQSGQGHQ------SGQGQQGYYPTSLWQPGQG--------------------QQGYASPYHVSAEYQAARLKVAKAQQLAAQLPAMCRLEGSDALSTRQ
QPGQWLQIGQGQQ-GYYPTSPQQLGQ------------GQQGYYLTSPQQPGQGK---------------QPGQGQQSYDSPYHVSAEHQAASLKVAKAQQLAAQLPAMCRLEGGDALLASQ
----VS-PHQGQQTTVSPHQGQQPGE-QP-----CGFPGQQ---TTVSLHHGQQS------------------NELYYGSPYHVSVEQPSASLKVAKAQQLAAQLPAMCRLEGGGGLLASQ
```

and 91 bases upstream of the ATG starting codon. This agrees, both in position and sequence, with the expectations of a 'TATA' box in prolamin genes. The nucleotide sequence of the '7841–1' fragment is 2,490-bp long. It contains an open reading frame of 2,304-bp that starts at nucleotide 139 and encodes a foreseeable polypeptide of 766 amino acids. The fragment '7841–2' is 2331 bp long and contains an open reading frame of 2145 bp that starts at nucleotide 139, and encodes a peptide of 713 amino acids. These results suggest that '7841–1' sequences correspond to an x-glutenin and '7841–2' to a y-glutenin.

The coding sequence of the mature polypeptide is similar to that of other HMW glutenins of wheat in that it can be divided into a number of distinct segments on the basis of amino-acid composition. The sequences of '7841–1' and '7841–2' also resemble the *Glu-1* genes in that they have a tripartite structure, consisting of a non-repetitive N-terminal region, an extensive repetitive central region, and a non-repetitive C-terminal region. The coding regions of '7841–1' and '7841–2' have an N-terminal non-repetitive domain of 87 and 105 residues respectively. Both sequences start with a 21-residue signal peptide (1...63 bp from the initiator codon) at the N-terminal end of the encoded protein (MAKRLVLFAAVVVALVALTVA and MAKRLVLAGIVVIALVALTA A). This peptide is almost identical to the signal peptide found in all the corresponding gene subunits encoding glutenins in wheat. The mature protein consists of 745 and 692 residues for '7841–1' and '7841–2' respectively. The N-terminal non-repetitive domain is followed by a repetitive domain of 634 residues in '7841–1', and 557 in '7841–2' (Table 1). The sequences encoded by '7841–1' and '7841–2' reveal the presence of tandem and interspersed repeats based on different type of tripeptides (YYP, PGQ, GQQ, PQQ), hexapeptides (PGQGQQ, GYYPAS, GYYPTS), a nonapeptide (PGQGQQPGQ) and a dodecapeptide (PGGPGQGQQPGQ) motif. The structure of this region and the type of repeats are similar to those found in the proteins encoded by the *Glu-1* genes of wheat. The repeats make up 76% of the total mature proteins. Finally, there is a second, non-repetitive domain at the C-terminal end that has 45 residues in '7841–1' and 51 in '7841–2'. The C-terminal sequence ASQ, found in both rye subunits, is the same as the C-terminal regions found in almost all reported HMW glutenins.

Figure 6 shows the alignment of the amino acid sequences encoded by the '7841–1' and '7841–2' rye sub-

◀ **Fig. 6** Comparison of the deduced amino acid sequences for '7841–1' and '7841–2' and nine glutenin sequences from *Glu-A1, Glu-B1, Glu-D1* of wheat, *Ae. squarrosa* and *H. vulgare* (EMBL accessions D82941, M22209, X03041, X03042, X03346, X12928, X12929, X13927, X13928, X61009.1, X61026) that have been aligned for maximal homology using the Clustal W 1.5 program. The *grey boxes* at the N-terminal region indicate the sequence of the signal peptide. The cysteine residues have also been boxed. Target sequences near the C-terminal region of '7841–1' and '7841–2' depict amino acids encompassed by forward and reverse primers used in the survey of the mRNA synthesis during grain ripening (regions between arrows in Fig. 5)

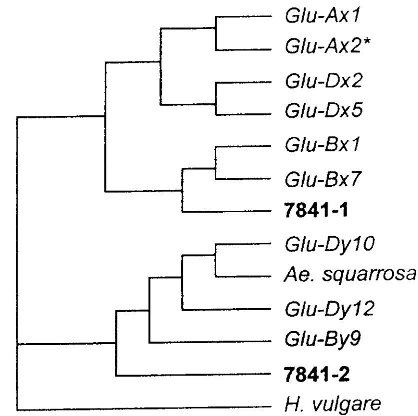

Fig. 7 Phylogenetic tree of amino acid sequences obtained from comparisons and alignment of the 11 Glu-1 sequences using the Phylip program (version 3.5)

units with those of nine previously published HMW glutenin genes of wheat, *Aegilops squarrosa* and *Hordeum vulgare*. The amino-acid sequence of the mature proteins of subunits '7841–1' and '7841–2' were compared. The analysis is presented as a dendrogram (Fig. 7) showing that '7841–1' and '7841–2' are resectively related to those of the *x* and *y* of other HMW genes.

The peptide encoded by '7841–1' contains more essential amino acids than the HMW glutenins encoded by *Glu-A1x, Glu-B1x* and *Glu-D1x* of wheat (data available in the NCBI Genbank). The peptide encoded by '7841–2' contains similar levels of essential amino acids to that of the y-type subunits of the genes *Glu-B1* and *Glu-D1* of wheat. Both peptides have at least the same number of cysteine residues as the equivalent peptides of wheat. Thus, '7841–1' has three cysteine residues in the N-terminal non-repetitive domain, and one in the C-terminal domain. Interestingly, '7841–2' has eight cysteine residues, five in the N-terminal domain, one near the end of the repetitive region, and two in the C-terminal non-repetitive domain.

RT-PCR to assess the expression of mRNA corresponding to the sequences '7841–1' and '7841–2' in endosperm

A final experiment was performed to assess the expression of the x- and y-type glutenin subunits encoded by '7841–1' and '7841–2' in the line '7841'. Different sets of primers were designed using information on the nucleotide sequences of '7841–1' and '7841–2' to amplify the mRNA presumably expressed during grain ripening. The RT-PCR reactions were performed using a common set of primers designed to amplify DNA fragments of different size for the x- and y- subunits located near the end of the non-repetitive C-terminal coding region (arrowed in Fig. 5). Electrophoretic analyses of the RT-PCR reactions from the common wheat line '7841' and testers

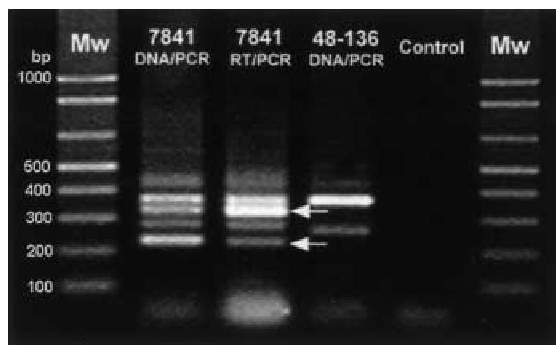

Fig. 8 RT-PCR analysis of *Glu-R1x* and *Glu-R1y* in '7841' and '48–136'. RT-PCR reactions were carried out on reverse transcribed mRNA samples from grain endosperm during ripening (10 days after anthesis), and a control sample using *Taq* polymerase to detect the presence of contaminant DNA

gave the expected results (Fig. 8). The RT-PCR including the extracts of m-RNA of '7841' in the presence of reverse transcriptase and *Taq* polymerase showed bands of cDNA of about 330 bp and 220 bp in agarose gels. The size of these products corresponded to those of the selected regions of the '7841–1' and 7841–2' nucleotide sequences. No amplification products were observed in control PCR reactions after treating the endosperm extracts with DNase in the absence of reverse transcriptase and the presence of *Taq* polymerase. A control reaction was also performed using genomic DNA from line '7841' in the presence of *Taq* polymerase. This reaction gave amplification products of the same size as those obtained from cDNA. Other RT-PCRs using m-RNA extracts from lines lacking the 1RL chromosome arm showed no amplification products of the expected sizes in the corresponding reactions (data not shown).

Discussion

DNA sequences of the *Glu-1* genes have indicated that the non-coding 5′- and 3′-flanks of glutenin genes are highly conserved up to 280 bp upstream of the TATA box and 140 bp downstream from the polyadenylation signal (Forde et al. 1985; Sugiyama et al. 1985). The primers that amplify the allele *Glu-D1y10* in wheat, and also in the rye locus described in this paper, corresponded to these regions. Primer designs and cycling conditions are critical for directing the amplification of specific sequences. Those variations residing at the 3′-terminal position of the primer are critical for annealing and amplifying the sequences present in any and all analogous gene subunits. In fact, the forward primer designed to amplify the wheat gene shows complete homology with the corresponding region of both –x and –y rye glutenin subunits, but has some minor variations with respect to all the remaining genes and alleles of the complex *Glu-1* family of wheat. The main differences are point changes (deletions, insertions and/or substitutions of at least two bases).

The alignment of the sequences '7841–1' and '7841–2' with 11 known wheat *Glu-1* sequences revealed inter-sequence relationships. This analysis demonstrates that '7841–1' belongs to the x-type group of gene subunits, whereas '7841–2' belongs to the y-type. Accordingly, the designations *Glu-R1x* and *Glu-R1y* for '7841–1' and '7841–2' are proposed. Further, the *Glu-1 x*-type subfamily splits into two subgroups (*Glu-A1x* + *Glu-D1x*, and *Glu-B1x*), with '7841–1' proximal to *Glu-B1x,* and '7841–2' proximal to *y* subunits of *Glu-B1, Glu-D1* and *Ae. squarrosa*. The *Glu-R1* subunits seems to group more-closely to the corresponding *Glu-B1* subunits than to subunits of the orthologous genes of the A and D genomes. The differences between the homoeologous genomes of wheat, as based on the biochemical and structural characteristics of the glutenins, have been previously discussed by Anderson and Greene (1989) These authors observed that the central repetitive region of the gene subunit encoded by *Glu-B1* is significantly different to the homoeologous regions of the A and D genomes, which may indicate a more distant ancestry. By comparing sequences of homoeologous (=orthologous) glutenin genes from all four genomes, the present results suggest that R is more-closely related to the B genome than to A or D.

It has been demonstrated that the HMW glutenins have an important role in determining the visco-elastic properties of wheat gluten (Tatham et al. 1984). Two structural features are responsible for the differences in the contribution of glutenins to breadmaking quality: secondary structure and the number of terminally located cysteine residues (Tatham et al. 1984; Halford et al. 1987). The secondary structure predicted for the mature proteins encoded by the x- and y- subunits of rye are notably similar to those of the main glutenins of wheat. In fact, the predicted structure shows the turns of the helix to depend on specific, short sequences of amino acids that appear conserved in both species.

The comparative analysis of the x- and y-type subunits of the *Glu-R1* gene here described demonstrates there to be the same number of cysteine residues in *Glu-R1x* as in the homologous gene subunits of wheat (four residues in *Glu-R1x, Glu-A1×1, Glu-A1×2**, *Glu-B1×7, Glu-B1×17* and *Glu-D1×2*). However, up to three extra cysteine residues have been observed in *Glu-R1y* with respect to the related alleles of wheat (four residues in *Glu-D1y12;* seven residues in *Glu-D1y10*, and eight residues in *Glu-R1y*). This property, plus the increased charge density shown by both of the sequences could be of interest in the formation of a complex glutenin matrix. Reconstitution studies, in which purified HMW rye glutenins were added to wheat doughs, have demonstrated a negative effect of these proteins on dough strength and this effect has been considered as due to a higher frequency of cysteine residues (Kipp et al. 1996; Graybosch et al. 1999). This could interrupt the correct formation of the large glutenin polymers necessary for strong doughs. Previous studies on the quality effects of rye chromosome 1R transferred to wheat have demonstrated that

genes on 1RL reduced grain weight, grain hardiness, mixograph time, mixograph tolerance and SDS sedimentation volumes (Graybosch et al. 1999). However, 1RL had no effect on flour yield or grain and flour protein concentrations, and the negative effects might also be related with unidentified genes on 1RL. It has been previously evidenced that in attempting to optimise grain quality for breeding purposes consideration must be given not only to the HMW-glutenin composition, but also to the presence of other genes in chromosome 1R and to many other biochemical traits which have a direct or indirect influence on technological quality (Bedo et al. 1999). Characterisation of the quality effects of HMW rye glutenins in different genotypic wheat backgrounds is now approachable using appropriate promotors in transforming plasmids. The more glutenin genes there are available to be used in experiments of transformation, the more precisely it will be possible to improve dough quality in transgenic forms, depending on the interactions of the other families of storage proteins.

Acknowledgements The authors declare that the experiments comply with the current laws of Spain and would like to thank the CICYT (*Comisión Asesora de Ciencia y Tecnología*) of Spain for financial support of this work (Grant No. AGL2000–0762-C02), the Education Council of the Autonomous Community of Madrid for the award of a postdoctoral fellowship, and Adrian Burton for linguistic assistance.

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

Bedo Z, Vida G, Láng L, Juhász A, Karsai I (1999) Breeding a wheat variety with different lines for technological quality and HMW glutenin composition. J Genet Breed 53:57–61

Bittel DC, Hueros G, Jouve N, Gustafson JP (1991) Changes in expression of seed storage protein genes affected by chromosome 1D of wheat. Genome 34:845–848

Bittel DC, Fominaya A, Jouve N, Gustafson JP (1992) Dosage response of rye genes in a wheat Background 1. Selection of lines with multiple copies of 1RS. Plant Breed 108:283–289

Dachkevitch T, Autran JC (1989) Prediction of baking quality of bread wheats in breeding programs by size-exclusion high-performance liquid chromatography. Cereal Chem 66:448–456

De Bustos A, Rubio P, Jouve N (2000) Molecular characterisation of the inactive allele of the gene *Glu-A1* and the development of a set of AS-PCR markers for HMW glutenins of wheat. Theor Appl Genet 100:1085–1094

De Bustos A, Rubio P, Soler C, García P, Jouve N (2001) Marker-assisted selection to improve HMW-glutenins in Wheat. Euphytica 119:69–73

De Vries S, Hoge H, Bisseling T (1988) Plant molecular biology manual, BG (1988) 1–13. Kluwer Academic Publishers, Dordrecht, Belgium

D'Ovidio R, Anderson OD (1994) PCR analysis to distinguish between alleles of a member of a multigene family correlated with wheat bread-making quality Theor Appl Genet 88:759–763

D'Ovidio R, Porcedu D, Lafiandra D (1994) PCR analysis of genes encoding allelic variants of high-molecular-weight glutenin subunits at the *Glu-D1* locus. Theor Appl Genet 88:175–180

D'Ovidio R, Masci S, Porcedu D (1995) Development of a set of oligonucleotides primers specific for genes at the Glu-1 complex of wheat. Theor Appl Genet 91:189–194

Felsenstein J (1993) PHYLIP (phylogeny inference package). Version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle

Forde J, Malpica JM, Halford NG, Shewry PR, Anderson OD, Greene FC, Miflin BJ (1985) The nucleotide sequence of a HMW glutenin subunit gene located on chromosome 1 A of wheat (*Triticum aestivum* L.). Nucleic Acids Res 13:6817–6832

Graybosch RA, Peterson CJ, Chung OK (1999) Quality effects of rye (*Secale cereale* L.) chromosome arm 1RL transferred to wheat (*Triticum aestivum* I) IX. Cereal Sci 29:211–216

Halford NG, Forde J, Anderson OD, Greene FC, Shewry PR (1987) The nucleotide and deduced amino-acid sequences of an HMW glutenin subunit gene from chromosome 1B of bread wheat (*Triticum aestivum* L.) and comparison with those of genes from chromosomes 1 A and 1D. Theor Appl Genet 75:117–126

Halford NG, Field JM, Blair H, Urwin P, Moore K, Robert L, Thompson R, Flavell RB, Tatham, AR, Shewry PR (1992) Analysis of the HMW glutenin subunit encoded by chromosome 1A of bread wheat: (*Triticum aestivum* L.) indicates quantitative effects on grain quality. Theor Appl Genet 83: 373–378

Huebner FR, Wall JS (1976) Fractionation and quantitative differences of glutenin from wheat varieties varying in baking quality. Cereal Chem 53:258–269

Kimura M (1980) A simple method for estimating evolutionary rates of nucleotide substitutions through comparative studies of nucleotide sequence. J Mol Evol 16:111–120

Kipp B, Belitz HD, Seilmeier W, Wieser H (1996) Comparative studies of high molecular-weight, subunits of rye and wheat. I. Isolation and biochemical characterisation and effects on gluten extensibility. J Cereal Sci 23:227–234

Lawrence GJ, Shepherd KW (1981) Inheritance of glutenin protein subunits of wheat. Mapping the nucleolus organiser region, seed protein loci and isozyme loci on chromosome 1R in rye. Theor Appl Genet 60:333–337

Maniatis BA, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Miflin RJ, Field JM, Shewry PR (1983) Cereal storage proteins and their effect on techonological properties. In: Daussant J, Mosse J, Vaughan J (eds) Seed proteins. Academic Press, London, pp 253–319

Odenbach W, Mahgoub El-S (1988) Relationships between HMW glutenin subunit composition and the sedimentation value in reciprocal sets of inbred backcross lines derived from two winter wheat crosses. In: Miller TE, Koebner RMD (eds) Proc 7th Int Wheat Genet Symp, Cambridge, pp 987–991

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, Corfield KG, Holt LM, Blackman JA (1981) Correlation between the inheritance of certain high-molecular-weight subunits of glutenin and bread-making quality in progenies of six crosses of bread wheat. J Sci Food Agric 32:51–60

Rubio P, Daza L, Jouve N (1999) Meiotic behaviour, chromosome stability and genetic analysis of the preferential transmission of 1B-1R, 1A-1R and 1R(1D) chromosomes in intervarietal hybrids of wheat. Agronomie 19:57–68

Sharp PJ, Desai S, Gale MD (1988) Isozyme variation and RFLPs at the â-amylase loci in wheat. Theor Appl Genet 76:691–699

Shewry PR, Miflin B, Kasarda DD (1983) The structural and evolutionary relationships of the prolamin storage proteins of barley, rye and wheat. Phil Trans R Soc Lond Ser B 304:697–308

Singh NK, Shepherd KW (1984) Mapping of the genes controlling high-molecular-weight glutenin subunits of rye on the long arm of chromosome 1R. Genet Res Camb 44:117–123

Singh NK, Shepherd KW (1991) A simplified SDS-PAGE procedure for separating LMW subunits of gluitenin. J Cereal Sci 14:203–208

Sugiyama T, Rafalski A, Peterson D, Soll D (1985) A wheat HMW glutenin subunit gene reveals a highly repeated structure. Nucleic Acids Res 13:8729–8737

Tatham AS, Shewry PR, Miflin BJ (1984) Wheat gluten elasticity: a similar molecular basis to elastin? FEBS Lett 177:205–208

Thompson RD, Bartels D, Harberd NP, Flavell RB (1985) Nucleotide sequence of a gene from chromosome 1D of wheat encoding a HMW glutenin subunit. Nucleic Acids Res 13:6833–6846

Thompson J D, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting matrix choice. Nucleic Acids Res 22:4673–4680

Van Campenhout S, Koebner RMD, Volckaert G (2000) The applicability of consensus PCR primers across species and genera: the use of wheat Em sequences to develop markers for orthologous in rye. Theor Appl Genet 100:328–336

Varghese JP, Struss D, Kazman ME (1996) Rapid screening of selected European winter wheat varieties and segregating populations for the *Glu-D1d* allele using PCR. Plant Breed 115:451–454