Theor Appl Genet (1993) 85:616–624

Analysis of a genomic DNA segment carrying the wheat high-molecular-weight (HMW) glutenin Bx17 subunit and its use as an RFLP marker

P. Reddy and R. Appels

Wheat Research Unit, Division of Plant Industry, CSIRO, P.O. Box 7, North Ryde, NSW 2113, Australia

Received March 2, 1992; Accepted September 19, 1992
Communicated by J. S. Beckmann

Summary. A genomic fragment containing the Bx17 high-molecular-weight (HMW) glutenin gene was isolated from a wheat genomic library. The fragment contains a coding region of 2.82 kb with 1.98-kb downstream and 12.8-kb upstream flanking regions. The fragment was sequenced and compared with previously published glutenin genes from chromosomes 1A, 1B and 1D using a computer alignment package. The Bx17 gene shows marked similarity to the Bx7 gene sequence. A phenetic tree derived from the alignments is presented. Also shown are restriction fragment length polymorphisms (RFLPs) at the glutenin loci in a set of Australian and international wheat varieties using different regions of the glutenin clone as probes. The RFLPs correlated well with the protein composition in all cultivars analysed.

Key words: Wheat – Glutenin – Dough – Evolution – RFLPs

Introduction

Genes coding for the high-molecular-weight (HMW) glutenins are expressed in a highly regulated fashion in the developing endosperm tissue of wheat seeds. Hexaploid wheats contain up to six copies of these genes which are located in pairs, made up of an x-type and a y-type subunit gene, on the long arm of chromosomes 1A, 1B and 1D. Several studies have shown that one or more of the HMW glutenin genes are not expressed in wheat cultivars at the mRNA (Forde et al. 1985) and protein level (Lawrence and Shepherd 1980; Galili and Feldman 1983). These studies have also shown that the y-subunit gene from the *Glu-A1* locus is silent in hexaploid wheat.

The HMW glutenin genes that are expressed exhibit extensive allelic variation in the protein products that they code. Variation in the allelic composition of European wheats is known to correlate significantly with the variation in bread-making quality (Payne et al. 1981; Moonen et al. 1983). Since glutenins have been implicated in forming highly complex secondary interactions with other gluten protein components, it is likely that variation in HMW glutenin subunits affects the properties of the aggregates containing them. Amino-acid sequences deduced from genomic clones for subunits Ax2*, Bx7, By9, Dx5, Dy10, Dx2 and Dy12 (Sugiyama et al. 1985; Thompson et al. 1985; Halford et al. 1987; Anderson and Greene 1989) have shown that the glutenins are made up of a large central domain, composed of repeated amino-acid sequences, which is flanked by unique N- and C-terminus regions. Computer predictions of secondary structures based on these sequences, supported by circular dichroism spectroscopy and scanning tunnelling microscopy, suggest that the central repeated domain adopts a beta-spiral structure (Tatham et al. 1985; Miles et al. 1991) that confers elasticity to the glutenin molecule.

The nucleotide sequences of subunits Ax2*, Bx7 and Dx5 from cultivar Cheyenne were compared by Anderson and Greene (1989). They found that while the Ax2* and Dx5 sequences could be aligned to each other, the Bx7 sequence contained repeats of a different pattern that prevented alignment with the other two genes. We have isolated and sequenced a Bx17 subunit from wheat and aligned its sequence with previously published HMW glutenin genes using a computer-assisted sequence alignment algorithm. In addition, we have obtained the nucleotide sequence information of regions extending 1.98 kb downstream and 1.91 kb upstream from the gene. Sequences from the coding and flanking regions have been used as 'markers' to study restriction fragment length polymorphisms (RFLPs) in a set of 59 Australian and 15 international wheat cultivars, and these RFLPs are compared with the protein composition of the cultivars.

Correspondence to: R. Appels

Materials and methods

Wheat material for isolation of total genomic DNA

Wheat leaf DNA was extracted from 2-month old plants by incubating 1 g of finely ground leaf material in 5 ml of extraction buffer (0.05 M Tris-HCl, pH 7.0; 0.1 M NaCl; 0.3 M EDTA; 0.3 mg proteinase-K and 0.6 ml of 5% SDS) at 37 °C for 2 h. The DNA was precipitated in ethanol and treated with RNase (50 µg/ml) before phenol/choroform extraction and final ethanol precipitation.

The wheat lines included the null subunit lines that are deficient for various combinations of the HMW glutenins and were supplied by Dr. G. J. Lawrence. The null lines L86-69, L86-110, L86-111 and L86-114 were derived from crosses between mutant lines of cultivars Olympic and Gabo (see Lawrence et al. 1988). L86-69 contains the subunits 1, 17+18 and 5+10; L86-110 contains only subunit 1, L86-111 only subunits 17+18, and L86-114 only subunits 5+10. In addition, DNA from the following Australian and international cultivars, kindly supplied by Dr. C. E. May, was analysed:

Australian cultivars – Banks, Bencubbin, Bodallin, Canna, Cocamba, Condor, Cook, Corella, Cranbook, Dagger, Diaz, Egret, Eradu, Eureka, Falcon, Gabo, Gamenya, Gluclub, Gutha, Halberd, Hartog, Harrier, Isis, Kalkee, Kiata, Kite, Koda, Kulin, Lowan, Madden, Miling, Millewa, Minto, Moray, Olympic, Osprey, Oxley, Quadrat, Quarrion, Robin, Rosella, Schomberg, Shortim, Skua, Songlen, Spear, Sundor, Sunbird, Suneca, Sunstar, Sunkota, Sunco, Teal, Timson, Tincurrin, Vasco, Warigal, Wren and Wyuna.

International varieties – Mexico-1856, Israel M68, Chinese Spring, Maris Huntsman, Maris Ranger, Kapstein, Neepawa, Norin 10, Pictic 68, Chile 1B, Mexico 120, Sonora 64A, Sonora 64B and La Prevision.

Gene library construction and screening

Genomic libraries were constructed by ligation of DNA from the euploid and three double-null subunit lines to lambdaGem-11. Wheat DNA was partially digested with *Sau*3A restriction endonuclease and the DNA size-fractionated on a 10–40% glycerol gradient (Maniatis et al. 1982). DNA fractions were ligated into *Bam*HI-digested lambdaGem-11 (Promega Corp., Australia) and transformed in NW2 (Woodcock et al. 1988), a recA$^-$, methylation-tolerant, *rg1B*$^-$ derivative of *E. coli* DS410. Plaques were transferred onto nitrocellulose filters and probed with nick-translated ^{32}P-labelled p1By-9 DNA (kindly supplied by Dr. P. Shewry). Single positive plaques were selected after four rounds of plaque purification and one of these clones, λGlu69 from the line L86-69, was used for sequence characterisation.

The *Sac*I sites located within the λGlu69 clone yielded fragments with lengths ranging from 300 bp to 11 kb which were subcloned into the *Sac*I site of phagemid Bluscript SK$^+$ (Stratagene) for nucleotide sequencing. Nested deletions were generated in these *Sac*I subclones using the Erase-a-base kit (Promega Corp.) by incubating DNA at 30 °C with Exonuclease III and withdrawing 1 µg aliquots at 1 min intervals followed by S1 nuclease digestion. The recessed ends were filled in with Klenow polymerase and ligated at 26 °C for 4 h. The clones showing deletions of 300–500 bp size were sequenced using the 24-mer M13 universal primer on the ABI automatic sequencing machine (Applied Biosystems, USA). The template for sequencing in both orientations was the ssDNA rescued using M13K07 infection of the Bluescript phagemid. The *Sac*I sites in the sequence were "crossed" by recovering the appropriate DNA fragment using the polymerase chain reaction (Cetus *Tag*1 DNA polymerase).

Computer-assisted sequence alignment

Nucleotide and amino-acid sequences from various regions of the glutenin genes were aligned phenetically by the Sequence Alignment package developed by Smith (1987). Alignment scores were derived from similarity measures of pairwise comparisons and the scores utilized in generating phenetic trees as described previously (Reddy and Appels 1989).

Results

Characterisation of a HMW glutenin gene clone

A 17.6-kb DNA fragment derived from the genomic library of the L86-69 line was characterized by nucleotide sequencing and restriction analysis. This fragment contains a 2.26-kb coding region bounded by a 12.8-kb upstream and 1.98-kb downstream flanking sequence (Fig. 1). The amino-acid sequence of the glutenin polypepide was deduced from the nucleotide sequence of the cloned gene. It translated into a 755 amino-acid long precursor protein comprising a signal peptide of 21 amino-acids, an N-terminus of 81 residues and a C-terminus of 42 residues, with an estimated molecular weight of 80,750 Da.

The upstream promoter region contains the typical eukaryotic TATA sequence at −91 and the CCAT sequence at the −119 position from the start of translation (Fig. 1). The open reading frame of λGlu69 ends at a double stop codon 'TGATAG' that is followed by a putative polyadenylation site 'AATAAA' 51 bases downstream from the stop codons. A similar sequence (AATAAT) is present 41 bases downstream from the polyA site which might be responsible for adenylating any read-through transcripts. Multiple polyadenylation signals have also been detected in other seed-storage protein genes of rice, barley and wheat (Okita et al. 1985; Pitts et al. 1988; Masumura et al. 1990).

Analysis of sequences flanking the coding region in λGlu69 revealed duplication and inversion events. A tandem duplication of 53 bases is located 569 bases upstream of the gene (Fig. 1). Nested within this duplication is a 24-bp sequence which bears sequence similarity to the −300 conserved element (Kreis et al. 1985). The duplication of the −300 element appears to be unique to the Bx subunit genes since it is also present in the Bx7 gene (Anderson and Greene 1989) but not the By subunit or any other HMW glutenin gene sequenced so far (Forde et al. 1985; Halford et al. 1987). This suggests that either a duplication of this sequence has occurred in the Bx subunit genes after divergence of the three genomes in wheat or else that the other subunit genes have lost the sequence by deletion.

-1795
GAGCTCTCCCATCCAATTGAAATATTCAATTTTAGATTTGGTTCACAATTTTGATCGGGTCGATGGTATCAATCCCGCGCATCCTCTCACCATAAAAA

AAAGAGTCAATATATGATATTGTAGCATTAATACATTACATGTAGCCACCGACATGGAAAACGTCCATGCATAAGTATGGTTGATTAATGAAAAACTC
-1599
ATAATCACACTGCATTTTTATAAATAAAAAGAAAATACACTTCATCATCTTCCTCACACACCAAGCCAAATACTCCATCTGTTTCTTTGACCAAGTTC

ACAACCAAAAACATCAATATCCATAATACTAAATAAATAAAATATGAAAATTCGTTTCATGCTTAATCCAATGATATCGACTTGATATTATGGATATT
-1403
TATGTATTTCTCTACAAATTTGATCAAACTTTAAAAGGTTTAACTTCTCAAAAAAAAACTAATACACCTCATATTTTAAAACAGAGAGAGTAGTTTTT

TTTTACCCAAACCCCCAACTGCCCAGCAAAGCGCGCTCAACTCTTCCTAGTCTAAATAACTAGCATCCACTAACACATTTCTCCCGACATGCAAGCAC
-1207
GTCACCTATGAAAATGCCCACCTCAATATGCAACCATGCATAGAAGAAAGCTCACCTCAGCATGCAAATCATGCAGCAAATTTCCATTTTACTTGGCT

ATTTATGTTTGATAAATATTTCACAAATATACAATAATCAAAAACAATAAATTATATGTGTTTTTAGTTTAGTTCTCATATCCAAATATACATGTTT
-1011
CATACAACCAAATCTCATTTAAATATATTGTAAAATATTCCGGCAACAACTTGTGGGGTACATCTAGTTACAGTGGAATATTAGTGATGGCCTGACGA

AGCGATAAGGCCAACGAGAGAAGAAGTGCGTCGTCCTATGGAGGCCAGGGAAAGACAATGGACATGCCAAAGAGGTAGGGGCAGGGAAGAAACACTTG
-815
GAGATCCATAGAAGAACATAAGAGGTTAAACATCGGAGGGCATAATGGACAATTAAATCCTACATTAATTGAACTCATTTGGGAAGTAAACAAAATCC

ATATTCTGGTGTAAATCAAACTATTTGACGCGGATTTACTAAGATCCTATGTTAATTTTAGACATGACTGGCCAAAGGTTTCAGTTAGTTCATTTGTC
-619
ACGGAAAGGTGTTTTCATAAGTCCAAAACTCTACCAACTTTTTTGCACCGTCATAGATAGATAGATGCTGTGAGTCATTGGATAGATATTGTGAGTCA

GCATGGATTTGTGTTGCCTGGAAATCCAACTAAATGACAAGCAACAAAACCTGAAATGGGCTTTAGGAGAGATGGTTTATCAATTTACAGTTTCCATG
-423
CAGGCTACCTTCCACTACTCGACATGGTTAGAAGTTTTGAGTGCCGCATATTTGCGGAAGAATGGCACTACTCGACATGGTTAGAAGTTTTGAGTGCC

GCATATTTGCGGAAGAATGGCTAACAGATACATATCTGCCAAACCCCAAGAAGGATAATCACTCCTCTTAGATAAAAAGAACAGACCAAGGTACAAAC
-227
ATCCACACTTCTGCAAACAATACACCAGAACGAGGATTAAGCCCATTACGTGGCTTTAGCAGACCGTCCAAAAATCTGTTTTCGCAAGCACCAATTGC

TCCTTACTTATCCAGCTTCTTTTGTGTTGGCAAACTGCCCTTTTCCCAACCGATTTGTTCTTCTCACGCTTTCTTCATAGGCTAAACTAACCTCGGCG
-31
TGCACACAACCATGTCCTGAAGCCTTCACCTCGTCCCTATAAAAGCCCATCCAACCTTCACAATCTCATCATCACCCACAACACCGAGCACCCCAATC
68
TACAGATCAATTCACTGACAGTTCACTGAGATGACTAAGCGCCTGGTCCTCTTTGCGGCAGTAGTCGTCGCCCTCGTGGCTCTCACCGCCGCTGAAGG
MetThrLysArgLeuValLeuPheAlaAlaValValValAlaLeuValAlaLeuThrAlaAlaGluGl
166
TGAGGCCTCTGGACAACTACAATGTGAGCACGAGCTCGAGGCATGCCAACAGGTGGTGGACCAGCAACTCCGAGACGTTAGCCCCGGGTGCCGCCCCA
yGluAlaSerGlyGlnLeuGlnCysGluHisGluLeuGluAlaCysGlnGlnValValAspGlnGlnLeuArgAspValSerProGlyCysArgProI

TCACCGTCAGCCCGGGCACGAGACAATACGAGCAGCAACCTGTGGTTGCGTCCAAGGCCGGATCCTTCTACCCCAGCGAGACTACGCCTTCGCAGCAA
leThrValSerProGlyThrArgGlnTyrGluGlnGlnProValValAlaSerLysAlaGlySerPheTyrProSerGluThrThrProSerGlnGln
362
CTCCAACAAATGATATTTTGGGGAATACCTGCACTACTAAGAAGGTATTACCCAAGTGTAACTTCTTCGCAGCAGGGGTCATACTATCCAGGCCAAGC
LeuGlnGlnMetIlePheTrpGlyIleProAlaLeuLeuArgArgTyrTyrProSerValThrSerSerGlnGlnGlySerTyrTyrProGlyGlnAl

TTCTCCCCAACAGTCAGGACAAGGACAGCAGCCAGGACAAGAACAGCAACCAGGACAAGGGCAACAAGATCAGCAGCCAGGACAAAGACAACAAGGAT
aSerProGlnGlnSerGlyGlnGlyGlnGlnProGlyGlnGluGlnGlnProGlyGlnGlyGlnGlnAspGlnGlnProGlyGlnArgGlnGlnGlyT
558
ACTACCCAACTTCTCCGCAACAGCCAGGACAAGGGCAACAACTGGGACAAGGGCAACCAGGGTACTACCCAACTTCACAGCAGCCAGGACAAAAGCAG
yrTyrProThrSerProGlnGlnProGlyGlnGlyGlnGlnLeuGlyGlnGlyGlnProGlyTyrTyrProThrSerGlnGlnProGlyGlnLysGln

CAGGCAGGACAAGGGCAACAATCAGGACAAGGACAACAAGGGTACTACCCAACTTCCCCGCAACAGTCAGGACAAGGGCAACAACCGGGACAAGGGCA
GlnAlaGlyGlnGlyGlnGlnSerGlyGlnGlyGlnGlnGlyTyrTyrProThrSerProGlnGlnSerGlyGlnGlyGlnGlnProGlyGlnGlyGl
754
ACCAGGGTACTACCCAACTTCTCCGCAGCAGTCAGGACAATGGCAGCAACCAGGACAAGGGCAACAACCAGGACAAGGGCAGCAATCAGGACAAGGGC
nProGlyTyrTyrProThrSerProGlnGlnSerGlyGlnTrpGlnGlnProGlyGlnGlyGlnGlnProGlyGlnGlyGlnGlnSerGlyGlnGlyG

AACAAGGTCAGCAGCCAGGACAAGGGCAACGACCAGGACAAGGACAACAAGGGTACTACCCAATTTCTCCGCAACAGCCGGGACAAGGGCAACAATCA
lnGlnGlyGlnGlnProGlyGlnGlyGlnArgProGlyGlnGlyGlnGlnGlyTyrTyrProIleSerProGlnGlnProGlyGlnGlyGlnGlnSer
950
GGACAAGGGCAACCAGGTTACTACCCAACTTCTTTGCGGCAGCCAGGACAATGGCAGCAACCAGGACAAGGGCAGCAACCAGGACAAGGGCAACAAGG
GlyGlnGlyGlnProGlyTyrTyrProThrSerLeuArgGlnProGlyGlnTrpGlnGlnProGlyGlnGlyGlnGlnProGlyGlnGlyGlnGlnGl

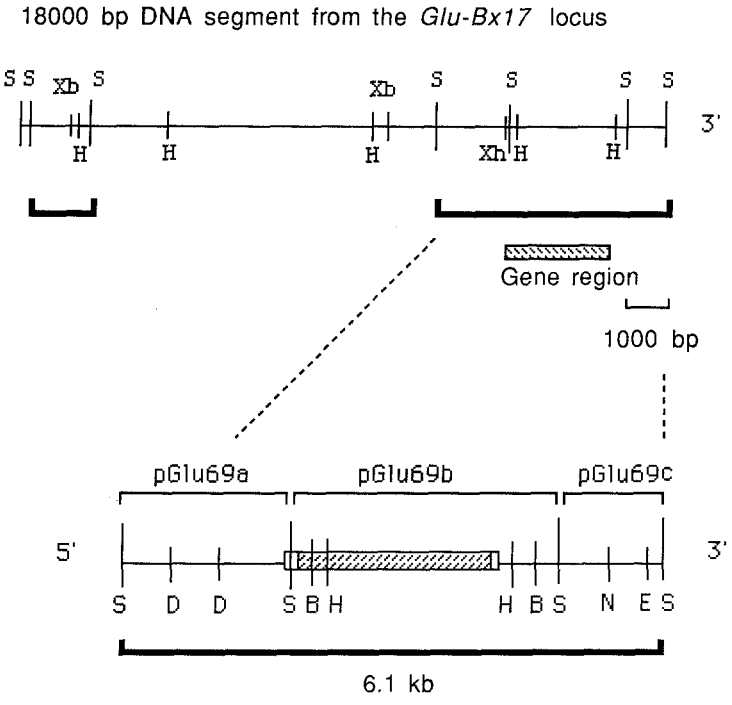
TCAGCAGCCAGGACAAGGACAACAATCAGGACAAGGACAACAAGGATACTACCCAACTTCTCTGCAACAGCCAGGACAAGGGCAACAACTGGGACAAG
yGlnGlnProGlyGlnGlyGlnGlnSerGlyGlnGlyGlnGlnGlyTyrTyrProThrSerLeuGlnGlnProGlyGlnGlyGlnGlnLeuGlyGlnG
1146
GGCAACCAGGGTACTACCCAACTTCGCAGCAGTCGGAACAAGGGCAGCAGCCAGGACAAGGAAAACAACCAGGACAAGGACAACAAGGGTACTACCCA
lyGlnProGlyTyrTyrProThrSerGlnGlnSerGluGlnGlyGlnGlnProGlyGlnGlyLysGlnProGlyGlnGlyGlnGlnGlyTyrTyrPro

ACTTCTCCGCAACAGTCAGGACAAGGGCAACAACTGGGACAAGGGCAACCAGGTTACTACCCAACTTCTCCACAGCAGTCAGGACAAGGACAACAATC
ThrSerProGlnGlnSerGlyGlnGlyGlnGlnLeuGlyGlnGlyGlnProGlyTyrTyrProThrSerProGlnGlnSerGlyGlnGlyGlnGlnSe
1342
AGGACAAGGACAACAAGGGTACTACCCAACTTCTCCGCAACAGTCAGGACAAGGGCAACAACCGGGACAAGGGCAATCGGGGTACTTCCCAACTTCTC
rGlyGlnGlyGlnGlnGlyTyrTyrProThrSerProGlnGlnSerGlyGlnGlyGlnGlnProGlyGlnGlyGlnSerGlyTyrPheProThrSerA

Fig. 1 (for continuation see page 619)

```
GGCAGCAGTCAGGACAAGGGCAGCAGCCAGGACAAGGACAACAGTCGGGACAAGGGCAACAAGGTCAGCAACCAGGACAAGGACAACAAGCGTACTAC
rgGlnGlnSerGlyGlnGlyGlnGlnProGlyGlnGlyGlnGlnSerGlyGlnGlyGlnGlnGlyGlnGlnProGlyGlnGlyGlnGlnAlaTyrTyr
                                                                                                    1538
CCAACTTCTTCGCAACAGTCAAGACAAAGGCAACAGGCAGGACAATGGCAACGACCGGGACAAGGGCAACCAGGGTACTACCCAACCTCTCCACAGCA
ProThrSerSerGlnGlnSerArgGlnArgGlnGlnAlaGlyGlnTrpGlnArgProGlyGlnGlyGlnProGlyTyrTyrProThrSerProGlnGl

GCCAGGACAAGAGCAACAATCAGGACAAGCGCAACAATCAGGACAATGGCAACTAGTGTACTACCCAACTTCTCCGCAACAGCCAGGCCAATTGCAAC
nProGlyGlnGluGlnGlnSerGlyGlnAlaGlnGlnSerGlyGlnTrpGlnLeuValTyrTyrProThrSerProGlnGlnProGlyGlnLeuGlnG
                                                                                                    1734
AACCAGCACAAGGGCAACAACCAGCACAAGGGCAACAATCAGCACAAGAGCAACAGCCAGGACAAGCGCAACAATCAGGACAATGGCAACTAGTGTAC
lnProAlaGlnGlyGlnGlnProAlaGlnGlyGlnGlnSerAlaGlnGluGlnGlnProGlyGlnAlaGlnGlnSerGlyGlnTrpGlnLeuValTyr

TACCCAACTTCTCCGCAACAGTCAGGACAAGGGCAACAAGGGTACTACCCAACTTCTCCGCAACAGTCAGGACAAGGGCAGCAGCCAGGACAAGGACA
TyrProThrSerProGlnGlnSerGlyGlnGlyGlnGlnGlyTyrTyrProThrSerProGlnGlnSerGlyGlnGlyGlnGlnProGlyGlnGlyGl
                                                                                                    1930
ACAGCCAAGACAAGGGCAACAAGGGTACTACCCAATTTCTCCGCAGCAGTCAGGACAAGGGCAACAACCAGGACAAGGGCAACAAGGATACTACCCAA
nGlnProArgGlnGlyGlnGlnGlyTyrTyrProIleSerProGlnGlnSerGlyGlnGlyGlnGlnProGlyGlnGlyGlnGlnGlyTyrTyrProT

CTTCTCCGCAGCAGTCAGGACAAGGGCAACAACCAGGACATGAGCAACAGCCAGGACAATGGCTGCAACCAGGACAAGGGCAACAAGGGTACTATCCA
hrSerProGlnGlnSerGlyGlnGlyGlnGlnProGlyHisGluGlnGlnProGlyGlnTrpLeuGlnProGlyGlnGlyGlnGlnGlyTyrTyrPro
                                                                                                    2126
ACTTCTTCACAGCAGTCAGGACAAGGGCATCAATCAGGACAAGGGCAACAAGGGTACTACCCAACTTCTCTGTGGCAACCAGGACAAGGGCAACAAGG
ThrSerSerGlnGlnSerGlyGlnGlyHisGlnSerGlyGlnGlyGlnGlnGlyTyrTyrProThrSerLeuTrpGlnProGlyGlnGlyGlnGlnGl

CTACGCCAGCCCATACCATGTTAGCGCGGAGTACCAGGCGGCCCGCCTAAAGGTGGCAAAGGCGCAGCAGCTCGCGGCACAGCTGCCGGCAATGTGCC
yTyrAlaSerProTyrHisValSerAlaGluTyrGlnAlaAlaArgLeuLysValAlaLysAlaGlnGlnLeuAlaAlaGlnLeuProAlaMetCysA
                                                                                                    2322
GGCTGGAGGGCAGCGACGCATTGTCGACCAGGCAGTGATAGAACTCTCTGCAGCTTGCATGGTGCTTGGGCATGCATGCACCTTAGCTATACAATAAA
rgLeuGluGlySerAspAlaLeuSerThrArgGln * *

CGTGACGTGTGTTCACAGCTTTTCGTGTAACTAGAGTAAAGCCCAATAATAATGCAAGATGAAAAGCTTCTCCGACTAAAAAAAGAACAAAACTGGTG
                                                                                                    2518
CTATATAGTATGTGTTGCATGTCTCAGTTCATTGTCAGCCACTGAATGCAAACATACTCCTATATTTTGCATGCGAAAAAAGAGGGTGTGGCAATATC

CCTAGCTAAAACTAACAACAGTTAGAAGGAAAAAAGAACAGACCAAAAATGGCTTCTTGGTTTGAGACCTACCATCATCCAAGCAAGGGAAAGGGAAA
                                                                                                    2714
ACGCCACATTCTTCGGGTGCAGACTAGGGATACGAGAGAAAAATTGTGCGCTCACCTGTGCCTCCTGCCGCTCCTTCTAACCGCCGGTGAAAATTCTT

ATCCGGTGATCGTAATGATCGGCACGAGAGTTTCGCCCACTCTCGCGTCGATCCAGGCCAATTCGATGGGGGTGCTGCTGCTCGTGGGTTCGTGGTCT
                                                                                                    2910
TCGCACCGTGAAATGGCAGAAGCATACGCCTCCTGAGCTGTGCTTGCTGTTGTTCATGGATCCGTGGCTTGTGCGGCAGGAAGAAGAGGGAGGCTGGG

CAAGAGACATCGACATCAGCGAGAAAGAAAACAAAGAGGAAGTGGAGGCGCAATTTCTGTGGCTGGACGTCACATGAAGCGAAGGAGCCATCAACATC
                                                                                                    3106
AGAGACAAAGAAAACAAAGAGGAAGTGGGGCCGCAATTTCTGTGGCTGGACGTCACATGAAGCGGAGGAGCCATTGACATCAGAGACAAAGAAAACAA

AGAGGAAGTGGGGCCGCAATTTCTGTGGCTGGACGTCACATGAAGCGGAGGAGCCATCGACATCAGAGACAAAGAAGAAGAGGAAGAACGGGGCGTCG
                                                                                                    3302
AGCTCTGAGCCGGACAAGAAAAGGTGGTTCAAGACCAAGATTTGGAGAAAGAAGAAGGCCAAGAACAAACAGTTGGCAACCCTTGTGAAAGAGATTTC

ACTCCCAAGTAATGCCCTCCATCATCTTATTATCCATCTTTGCAGAGAAACTAAATTCAGCAATATCAAATTTATCAATGATTATCAATGAATATTTG
                                                                                                    3498
CCCAACACAGATATGTGACGAGCACTCAGGTACTTACATAGGAAATCTGGTGGTAAGAATGTTTAGTTCAAATACATTGCAACAACAAGTTAGTGATG

TTCAGGGATAGTTGTATGTGCATTGCCCTCTTCTTGGACCTACTTTCACCCTGAAATTTGTAATCCTGGATTAGTAATTCATGTTTGGGCGATTAAGT
                                                                                                    3694
AAGCATTTCTGATTGAGAAGAAATCTTAAAATGCAGATAGTCCCAAGGCCAGGGCTGCGGCAGGGGAAATACTACGGATCGGCAACCACAACATTCCT

AGCAGAGTGTTTACGTACAGTCAACTGTCAGACGCGACCAATTCTTTCAGCCAGGAGAATCTCCTGGGAGAAGGTGGCTTTGGAAGGGTGTACAAAGG
                                                                                                    3890
ATACATTCCAGAGACCATGGAAGTAATAAGCATTCCTTCTCCCAGCTTCCTGTTCTATGTTTCATTAAGTTGGCACCTCAAGATGCTATGCTACATAC

AGCAGCAACACCTTCCTGATGTAAATTATGTGCGTACGTAGGTCATAGCGGTTAAGCAGCTGGACAAGGATGGGTTGCAAGGAAATCGTGAGTTCCTT
                                                                                                    4086
GTCGAGGTGCTGATGCTTAGCCTCCTTCATCACCCAAACCTTGTCACCTTGCTCGGTTACTGCACCGAATGTGACCAAAACATTCTAGTGTATGAGTA

TATGCCACTCGGTTCTTTACAAGATCATCTCCTAGGTAATACAGAACAATAATGTTGTGGTTCTCGAACATATTCCGCAGAACTGAAAAATCTGAAAT
                                                                                                    4147
TATGATAGCACATTTGTATCGATGCTGATTTGGTACTTGAATTCTGCAGATCCTCGAGCTC
```

Fig. 1. Nucleotide and deduced amino-acid sequence of the Bx17 glutenin clone (λGlu69). The numbering originates at the first methionine of the open reading frame. The TATA and putative polyadenylation sequences are *boxed* and the stop codons indicated with *asterisks*. The tandem duplication containing the –300 element is *underlined* and the latter enclosed in *brackets*



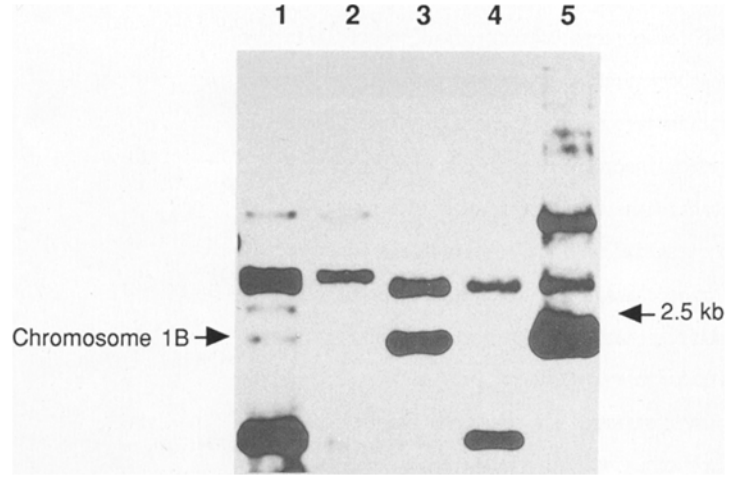


Fig. 2. Top panel: restriction map of the λGlu69 clone and an enlarged section showing the three *Sac*I fragments used for sequencing (labelled pGlu69a–c). The gene region is *hatched* and the restriction sites shown are *B*, *Bam*HI, *D*, *Dra*I; *E*, *Eco*RI; *H*, *Hin*dIII; *N*, *Nco*I; *S*, *Sac*I; *Xb*, *Xba*I; *Xh*, *Xho*I. The sections of λGlu69 were sequenced are indicated (▬). Bottom panel: *Taq*I-digested DNA of the null lines L86-69, -110, -111 and -114 (*lanes 1–4*) and the λGlu69 DNA (*lane 5*) was immobilized on nylon membrane and hybridized with the pGlu69b subclone. The *arrow* on the left indicates the band corresponding to the glutenin gene from the B genome and the number on the right a molecular weight marker. The chromosome-1B band is present in *lane 1* but has come up relatively faintly

Origin of the HMW glutenin clone λGlu69

The chromosomal origin of the clone was determined by comparison of the *Taq*I restriction fragment sizes of DNA from the null subunit lines with that of the λGlu69 clone cut with the same enzyme. Hybridization of DNA from this digest with a *Sac*I fragment of λGlu69 containing most of the gene coding region (pGlu69b) indicated that the clone was derived from the chromosome-1B locus of HMW glutenins (Fig. 2). Since the *Glu-B1* locus in the null lines is derived from the Gabo parent (Lawrence et al. 1988), this clone was expected to correspond to either the Bx17 or the By18 allele. Nucleotide sequence analysis of the λGlu69 clone revealed it to be an x-subunit clone based on its length, N-terminal sequence and repetitive motif patterns (data in next section), thus establishing its correspondence to the Bx17 allele.

Alignment of the Bx17 sequence with genes from chromosomes 1A, 1B and 1D

The amino-acid sequence of the gene for subunit 17 was aligned with that of eight previously published HMW glutenin genes using the multiple sequence alignment algorithm (Smith 1987). This program is equipped to compare sequences of different lengths where it automatically normalizes the alignment scores by dividing with the total alignment length. Alignment scores based on the en-

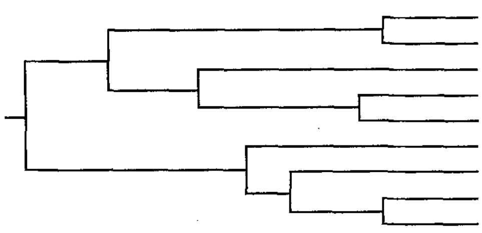


Fig. 3. A phenetic tree derived from the sequence alignment of the nine HMW glutenin genes. The genes are indicated in *parenthesis* and by their allelic nomenclatures

tire peptide sequence of the nine subunits were then used to produce a phenetic tree (Fig. 3). The dichotomy between the x- and y-type genes is clearly apparent by the tight clustering of the genes from different genomes within a subunit type. Within a cluster, the y subunits (except for the 1Ay silent gene) associate more closely, indicating greater sequence similarity, than to the x genes. Sequence alignment using an alternative computer program developed by SoftGene GmbH (Berlin) produced a similar result where all the y-type sequences aligned closer to

```
                                                                                  80
Bx17      MTKRLVLFAAVVVALVALTAAEGEASGQLQC-----EHELEACQQVVD------------------QQLRDVSPGCRPIT
x-cons    .A.............................ERELQ....K......------------------........E.H.VV
y-cons    .........T..I......T......R..........SS.E..R....QQLAGRLPWSTGLQMRCC.......AK..SVA

                                                                                 160
Bx17      VSPGTRQYEQQPVVASKAGSFYPSETTPSQQLQQMIFWGIPA-LLRRYYPSVTSSQQGSYYPGQASPQQSGQGQQPGQEQ
x-        ...VA......I..PP.G.....G....P.........................P..V..........RP..........
y-        ..QVA....-.T....PP.G.....G..L.....G....TSSQTVQG.......P..............P.......KW-

                                                                                 240
Bx17      QPGQGQQDQQPGQRQQGYYPTSPQQPGQGQQLGQGQPGYYPTS----QQPGQKQQAGQGQQSGQGQQGYYPTSPQ-QSGQ
x-        .............G..............W...E..........PQQP.....L..P.............Q.GQGQP.YYP
y-        --------.EL..G..W.....L........I....Q......LQQP-------------------------------..

                                                                                 320
Bx17      GQQPGQGQPGYYPTSPQQSGQWQQPGQGQQPGQGQQSGQGQQGQQPGQGQRPGQGQQGYYPISPQQPGQGQQSGQGQPGY
x-        TSSQLQPGQLQQ.AQG..GQ.P...Q.......................Q..............................
y-        G..I....Q........HP..R...V....I.....G-------..E...Q...W......T....L.....P..-----

                                                                                 400
Bx17      YPTSLRQPGQWQQPGQ-GQQPGQGQQGQQPGQGQQSG--QGQQGYYPTSLQQPGQGQQLGQGQPGYYPTSQQSEQGQQPG
x-        .....Q.L..G.....G......P..L...A.....SGQ....PGQGQQ.........P...........P.QTSS....
y-        ----WQ.S..G.....GQQGHYLAS.---------------..PGQGQQGHY.ASQQ.P....Q.H..A...QAS-----

                                                                                 480
Bx17      QGKQPGQGQQGYYPTSPQQSGQGQQLGQGQPGYYPTSPQQSGQGQQSGQGQQGYYPTSPQQSGQGQQPGQGQSGYFPTSR
x-        .SQ.........QV.QG...Q....GQ.P.Q...G......T.........P..................L..SAQ....
y-        -QQ........HI.A.Q..P.....GHYP------------ASL..P......H....L..L.....I.-----------

                                                                                 560
Bx17      QQSGQGQQPGQGQQSGQGQQGQQPGQGQQAYYPTSSQQSRQRQQAGQWQRPGQGQPGYYPTSPQQPGQEQQSGQAQQSGQ
x-        ..P...........P.............PG.....P...G....P....Q.............L....G.PGY.PTSPQ.
y-        -.P..K.....TG.GQ.PE.E........G.....--.QG.L..P..G.Q.....Q......L.....G..GHYPASLQ.

                                                                                 640
Bx17      QWQLVYYPTSPQQPGQLQQPAQGQQPAQGQQSAQEQQPGQAQQSGQWQLVYYPTSPQQSGQGQQGYYPTSPQQSGQGQQR
x-        .PGQGQQ.GQL...A.......LA..........V...QQPA..Q...........L.L...........L..P......
y-        .PGQG-Q.GQR.....G.H.E....-----------------------------------...........------...

                                                                                 720
Bx17      PGQGQQPRQGQQGYYPISPQQSGQGQQPGQ-GQQGYYPTSPQQSGQGQQPGHEQQPGQWLQPGQGQQGYYPTSSQQSGQG
x-        ...W..SG.......................WL.P.Q..QSGYYSTSP....QG....................L......
y-        ......LG........T...P........-....HC..........A....QG..I..VQ...............L..P...

Bx17      HQSGQGQ---QGYYPTSLWQPGQGQQGYASPYHVSAEYQAARLKVAKAQQLAAQLPAMCRLEGSDALSTRQ
x-        Q....................................S.....................G....AS.
y-        Q......QSG..HQPGQGQQS...E....D...........SPM......P.T...T...M.......AS.
```

Fig. 4. Alignment of amino-acid sequences of Bx17 and the consensus sequences of the x- and y-type HMW glutenin genes using the multiple sequence alignment algorithm. Residues in the x- and y-consensus that are similar to Bx17 are shown by a *dot* and those that are different by the *relevant letter*. Deletions in all three sequences are indicated by a *dash*

each other than did the x-type genes (B. Wittig, personal communication).

Of particular interest in the phenogram was the close branching of the allelic forms of a subunit type (for example Bx17 and Bx7). The Bx17 and Bx7 genes aligned perfectly with each other except for two amino-acid substitutions and a 36 codon deletion. The 36 amino-acid deletion in Bx17 was located in the repetitive domain and included the loss of a block of three hexapeptide and two nonapeptide repeats. This deletion in Bx17 probably explains the difference in mobilities of the two proteins on an SDS-electrophoresis gel (separation based on protein size) where Bx17 protein runs faster than Bx7 (data not shown).

An alignment of the consensus sequence of the x- and y-type genes is shown in Fig. 4. A marked difference exists between the two types in the central domain, with several deletions, duplications and substitutions separating their amino-acid sequence. In contrast, the N- and C-termini are very similar between the two types with only an 18-residue-long deletion distinguishing the N-terminus of the x subunits relative to the y. Figure 4 also shows the clear alignment of λGlu69 to the x-type genes.

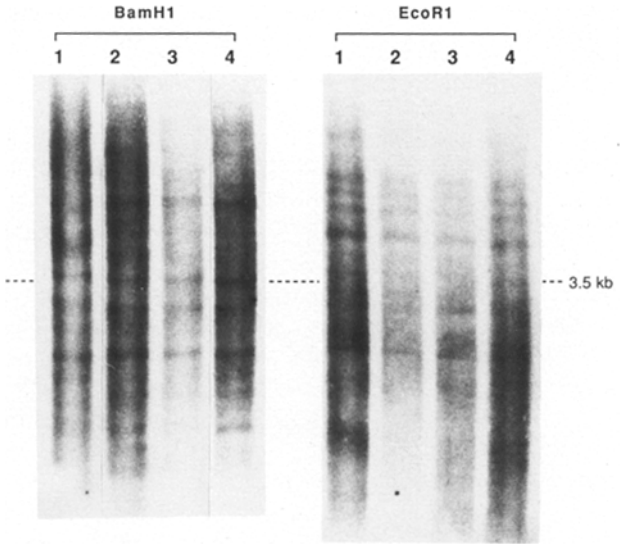


Fig. 5. DNA from varieties Skua (*lane 1*), Canna (*lane 2*), Suneca (*lane 3*) and Warigal (*lane 4*) was digested with *Bam*HI and *Eco*RI individually, immobilized on nylon membrane and probed with the upstream sequence subclone pGlu69a

Variability in wheat varieties assayed with λGlu69

The three subclones, pGlu69a–c (see Fig. 2), spanning the upstream, coding and downstream flanking regions of the Bx17 clone were used as probes in the restriction analysis of a set of 59 Australian and 15 International wheat varieties. Five different restriction endonuclease enzymes, viz. *Bam*HI, *Hin*dIII, *Eco*RI, *Eco*RV and *Taq*I, were used to assay variation at the glutenin loci in these wheats. The following profiles were observed:

Using the 5′ flanking regions as probe: When subclone pGlu69a (from the 5′ flank) was used as a probe on the wheats digested with the above five enzymes, no variation in banding patterns was evident among the cultivars. Although each enzyme produced a distinct banding pattern the bands were uniform in all varieties tested (Fig. 5).

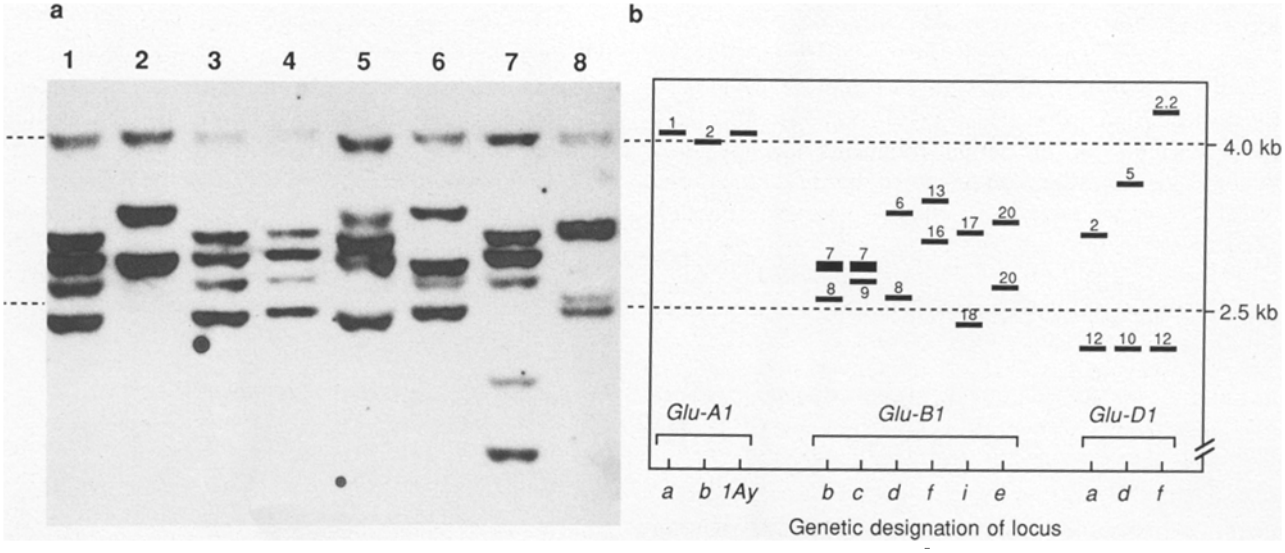


Fig. 6. a *Taq*I-digested DNA of eight wheat cultivars Osprey, Olympic, Oxley, Banks, Songlen, Dagger and Quarrion in *lanes 1–8* respectively was blotted onto nylon membrane and hybridized with the coding region subclone pGlu69b. **b** Diagrammatic representation of the relative positions of the *Taq*I bands produced by using pGlu69b as an RFLP marker on 74 wheat cultivars. The bands are shown for the three glutenin loci and the allelic designation is written at the bottom of each lane with the numerical denotation at each band

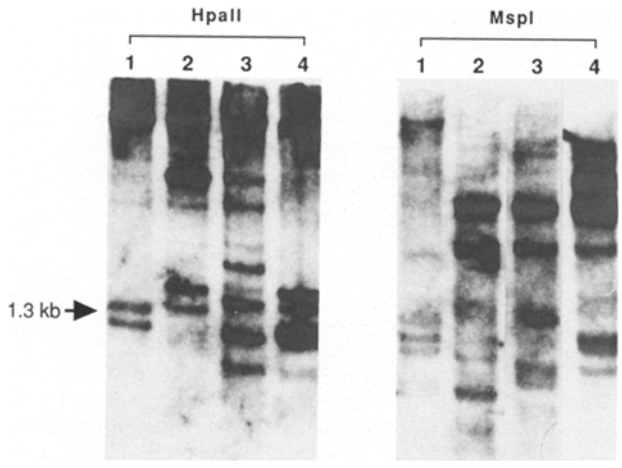


Fig. 7. DNA of cultivars Cranbrook, Gutha, Oxley and Tincurrin restricted with *Hpa*II (*lanes 1–4*) and its isoschizomer *Msp*I (*lanes 5–8*) was blotted and probed with pGlu69b. These two pairs of cultivars share identical glutenin composition but can be differentiated by RFLPs produced by the *Hpa*II enzyme

Multiple bands were arranged in irregular ladders from *Bam*HI, *Eco*RI and *Hin*dIII digests in addition to a smear of bands in the background.

The coding region probe: Restriction analysis of the various wheats with the *Sac*I fragment of λGlu69 that includes the coding region produced a totally different banding pattern compared to the 5′ flanking probe. Discrete bands of mobilities ranging from 1.8 to 9.3 kb in length were observed with the different enzymes (Fig. 6). Although some RFLPs were evident among the varieties with all the enzymes used, maximum variation was detected with *Taq*I.

On comparing the RFLPs of *Taq*I with the HMW glutenin composition based on SDS-PAGE separation of the proteins, a direct correlation was established between the occurrence of restriction fragments and protein bands. Thus *Taq*I digestion produced discriminate bands for HMW glutenin alleles in all the 73 varieties tested. Figure 6 shows the mobilities of the *Taq*I fragments that could be allocated to the different alleles tested.

It was also of interest to differentiate between varieties that where indistinguishable from SDS-PAGE profiles. The enzyme *Hpa*II consistently demonstrated polymorphisms between cultivars that shared identical glutenin alleles (Fig. 7). Differences could also be shown for these wheats when digested with *Msp*I (an isoschizomer of *Hpa*II that is methylation-insensitive) indicating restriction site polymorphism at the 'CCGG' sequence in the cultivars.

Using 3′ flanking region as probe: Patterns produced by the pGlu69c fragment as a hybridization probe on wheat DNA digested with a particular restriction endonuclease, revealed a uniform pattern of bands in all varieties tested (data not shown).

Discussion

Evolution of the glutenin genes

A comparison of the nucleotide sequence of the Bx17 HMW glutenin gene cloned in the present study with that of eight previously published glutenin genes has shown that several structural changes, such as point mutations, deletions and duplications in the coding and flanking regions, separate these genes. The changes are particularly apparent when the x- and y-type subunits are compared. These two lineages of glutenin genes have diverged from an ancestral sequence prior to the separation of the wheat genomes (Shewry et al. 1989). From the present study, it is also obvious that the two lineages are evolving at differential rates. An alignment of the sequences showed that the y-type genes (except for 1Ay) clustered together more closely than the x genes. The silent gene 1Ay (Thompson et al. 1985) exhibited maximum change in its amino-acid composition relative to the other y genes, and the mutations were localized mainly in the central repetitive region. Apart from the silent 1Ay gene, the y-subunit genes from the B and D genomes are more similar to each other in sequence composition than the corresponding x-subunit genes from these genomes. These results suggest that except for the event/s that would have rendered the 1Ay gene silent in bread wheats, the y-subunit genes have accumulated fewer changes and have evolved at a slower rate than the x-type genes during evolution of the wheat family.

An alignment of the consensus sequence of the two subunit types (Fig. 4) also suggests that the deletions in the y-subunits, relative to the x, are comprised of complete blocks of repeat units that include the hexapeptide (PGQGQQ), nonapeptide (GYYPTSP/LQQ) and tripeptide (GQQ) repeat units. It can be concluded that although the x- and y-type genes evolved from the same basic sequence, duplications of this sequence have occurred independently in the two lineages, with the x-type genes having incorporated a tripeptide sequence within its basic sequence.

Potential utilization of glutenin RFLPs

When DNA fragments from the coding and flanking regions of the Bx17 clone were used as probes to detect the extent of restriction fragment length variability at the glutenin loci in a set of 59 Australian wheat cultivars, maximum variation was defined by the coding region probe pGlu69b (see Fig. 6). With this probe, diagnostic *Taq*I bands could be assigned to the various allelic combinations in the wheats. This correlation between

glutenin RFLPs and protein composition also held true when a more diverse set of international wheat cultivars was analysed. The potential usefulness of the RFLP procedure, in providing an increased level of discrimination among glutenin compositions in breeding programs to select for dough properties, was demonstrated by using the enzyme *Msp*I.

Previous studies using cDNA probes for the hordeins in barley (Bunce et al. 1986) and the HMW glutenins in wheat (Harberd et al. 1986) have shown that restriction fragment length variation exists at these loci. In addition, Bunce et al. (1986) detected restriction polymorphism between cultivars with the same hordein alleles. In the present study, the restriction enzyme *Hpa*II proved useful in discriminating between cultivars with identical glutenin alleles. Thus a combination of *Taq*I and *Hpa*II digestion would not only reveal the extent of variation among cultivars but also aid in varietal differentiation.

Acknowledgements. The authors would like to thank Helen Carpenter and Rosalind Batten for technical support and Dr. G. Donovan for generating the lambda libraries. We are grateful to Dr. G. Lawrence for seeds of null subunit lines and to Dr. C. E. May for supplying wheat DNA for the RFLP analysis. Also thanked is Dr. P. Shewry for providing the p1By9 probe. P. R. was funded by a grant from the Australian Grain Research and Development Corporation.

References

Anderson OD, Greene FC (1989) The characterization 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

Bunce NAC, Forde BG, Kreis M, Shewry PR (1986) DNA restriction fragment length polymorphism at hordein loci: application to identifying and fingerprinting barley cultivars. Seed Sci Technol 14:419–429

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

Galili G, Feldman M (1983) Genetic control of endosperm proteins in wheat. 2. Variation in high molecular weight glutenin and gliadin subunits of *Triticum aestivum*. Theor Appl Genet 66:77–86

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

Harberd NP, Bartels D, Thompson RD (1986) DNA restriction fragment variation in the gene family encoding high molecular weight (HMW) glutenin subunits of wheat. Biochem Genet 24:579–596

Kreis M, Shewry PR, Forde BG, Forde J, Miflin BJ (1985) Structure and evolution of seed storage proteins and their genes, with particular reference to tose of wheat, barley and rye. Oxford Surveys Plant Mol Cell Biol 2:253–317

Lawrence GJ, Shepherd KW (1980) Variation in glutenin protein subunits of wheat. Aust J Biol 33:221–233

Lawrence GJ, MacRitchie F, Wrigley CW (1988) Dough and baking quality of wheat lines deficient in glutenin subunits controlled by the *Glu-A1, Glu-B1* and *Glu-D1* loci. J Cereal Sci 7:109–112

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

Masumara T, Hibino T, Kidzu K, Mitsukawa N, Tanaka K, Fujii S (1990) Cloning and characterisation of a cDNA encoding a rice 13 KDa prolamin. Mol Gen Genet 221:1–7

Miles MJ, Carr HJ, McMaster TC, l'Anson KJ, Belton PS, Morris WJ, Field JM, Shewry PR, Tatham AS (1991) Scanning tunneling microscopy of a wheat seed storage protein reveals details of an unusual supersecondary structure. Proc Natl Acad Sci USA 88:68–71

Moonen JHE, Scheepstra A, Graveland A (1983) The positive effects of the high molecular weight subunits 3+10 and 2* of glutenin on the breadmaking quality of wheat cultivars. Euphytica 32:735–742

Okita TW, Cheeseborough V, Reeves CD (1985) Evolution and heterogeneity of the a/b and g-type gliadin DNA sequences. J Biol Chem 260:8203–8213

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

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

Reddy P, Appels R (1989) A second locus for the 5S multigene family in *Secale* L.: sequence divergence in two lineages of the family. Genome 32:456–467

Shewry PR, Halford NG, Tatham AS (1989) The high-molecular-weight subunits of wheat, barley and rye: genetics, molecular biology, chemistry and role in wheat gluten structure and functionality. Oxford Surveys Plant Mol Cell Biol 6:163–219

Smith DK (1987) Concept and use of software for pairwise and multiple sequence alignment. Techn Rep No. 11. Canberra, CAE, Australian Capital Territory

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

Tatham AS, Miflin BJ, Shewry PR (1985) The beta turn conformation in wheat gluten proteins: relationship to gluten elasticity. Cereal Chem 62:405–411

Thompson RD, Bartels D, Harberd NP (1985) Nucleotide sequence of a gene from chromosome 1D of wheat encoding a HMW-glutenin subunit. Nucleic Acids Res 19:6833–6846

Woodcock DM, Crowther PJ, Diver WP, Graham M, Bateman C, Baker DJ, Smith SS (1988) Rg1B-facilitated cloning of highly methylated eukaryotic DNA: the human L1 transposon, plant DNA and DNA methylated in vitro with human DNA methyltransferase. Nucleic Acids Res 16:4465–4482