Bot. Bull. Acad. Sin. (1996) **37:** 41-49

Analysis of 5' region of glutelin genes from wild rice species

Hsin-Kan Wu[1], Tein-chin Chen, and Mei-Chu Chung

Institute of Botany, Academia Sinica, Nankang, Taipei, Taiwan, Republic of China

(Received June 7, 1995; Accepted September 13, 1995)

Abstract. The structure of the 5' flanking region of glutelin genes amplified from the various wild rice species was analyzed by cloning and sequencing. The results showed that beyond the essential boxes (legumin, CAAT, AACA and TATA), the 5' region of rice glutelin genes have numerous putative enhancers (long-direct and short-direct repeats) and putative regulatory segments (RY repeats, -300 bp elements, nuclear protein binding sites) though portions of a few elements have been deleted in some wild species. The possible roles of most of the putative elements in glutelin gene expression remain to be determined. The sequence length and structure of glutelin 5' regions vary among rice species. On the basis of the length, the degree of homolgy, and the corresponding base substitutions and deletions in the 5' regions of glutelin genes, the authors suggest that glutelin genes, in the subfamily *Glua* can be classified into three kinds of members, each with its 5' region of 0.5 kb, 0.9 kb, or 1.2 kb. A member gene in the subfamily may reside at one or more loci in a rice genome. The same member gene that appears among rice species with minor deletion, addition, or substitution may be designated as alleles of that gene.

Keywords: Glutelin gene; 5' Region sturcture; Subfamily *GluA*; Wild rice.

Introduction

A rice glutelin cDNA was first isolated from cultivated rice by Takaiwa et al. (1986). More glutelin cDNAs were isolated and classified into two types, I and II, which can be distinguished respectively by stop coden TAA or TAG and by two and one polyadenylation signals (Takaiwa et al., 1987a). Since these early reports, three genomic clones (*Gt1, Gt2* and *Gt3*) for rice glutelin have been isolated and studied in Okita's laboratory. Comparison of DNA sequences from relevant regions of these clones showed that two of them, *Gt1* and *Gt2*, are closely related. *Gt3* shows little or no homology to *Gt1* and *Gt2*. All three clones had 5' flanking regions of less than 0.9 kb (Okita et al., 1989). Two new glutelin genes, *Glua*-3 and *Glua*-4 were later added to subfamily A (Takaiwa and Oono, 1991) that already contains *Glua*-1 (Type I) and *Glua*-2 (Type II). Takaiwa et al. (1991) proposed a new subfamily of glutelin gene, subfamily B, in which three member genes have been sequenced. Furthermore, direct repeat, enhancer core, and legumin box (Takaiwa et al., 1987b), -300 bp element, RY repeats and inverted repeats (Okita et al., 1989) and nuclear protein binding sites (Kim and Wu, 1990; Takaiwa and Oono, 1990) have been reported to be related to glutelin gene expression in cultivated rice. This paper reports the structure of 5' regions of glutelin genes with special emphasis on that of glutelin subfamily A genes from wild rice species.

Materials and Methods

The species, both cultivated and wild, used in this experiment are shown in Figure 1.

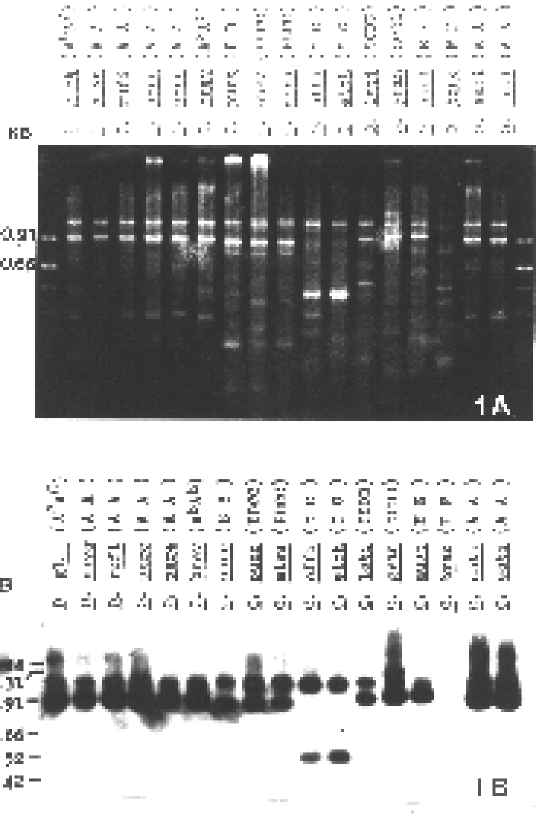

Figure 1. Southern blot analysis of the amplified rice glutelin 5' region. 1A, Gel electrophoresis of the PCR amplified 5' regions of glutelin genes from various rice species each indicated above the line. Each has two to three major bands; 1B, The blotted DNA was probed with the 5' region of a glutelin gene, isolated from rice cultivar Tainung 67 (genome AA). It shows that all wild rice species except *O. brachyantha* (genome FF) have two major bands positively reacted.

[1] Corresponding author.

Polymerase chain reaction (PCR) was used to amplify the 5' region of glutelin genes from various rice species. The two ends of the 5' region of a known glutelin gene (Takaiwa et al., 1986) were used to synthesize two primers, a2–1 (5' CAAGCTTTTGGAAAGGTGCCG3') and a2–2 (5’GCTCTAGAGTTGTTGTAGGACTAATGAA3') each with a *XbalI* or a *Hind*III linker. The amplified 5' regions were cloned into the plasmid M13 *mp*19 to produce recombinant DNA molecules.

Deletion and sizing: Double-stranded recombinant DNA of M13 was extracted and digested with exonuclease III to produce successively shortened insert DNA using the Erase – a Base System (Promega). Transformation of *E. coli* JM101 was carried out with the deleted recombinant DNA using an *E. coli* Pluser Apparatus. Sizing was performed by electrophoresis of the single stranded shortened recombinant DNA from M13.

Sequencing and Analysis: The single-stranded recombinant DNAs were used as template and annealed to the fluorescent primers supplied in the Auto Sequencing Kit (Pharmacia). Sequencing reactions were carried out according to the procedures suggested by the supplier, and the products were loaded in an automated Laser Fluorescent DNA Sequencer (Pharmacia ALF). Data thus obtained were processed by a GCG (Genetics Computer Group) sequence analysis software package.

Results

For PCR amplification of the 5' regions of the rice glutelin genes, two oligomers were synthesized. The amplified DNAs were analyzed by gel electrophoresis. As shown in Figure lA, each lane gave two major bands and several minor ones. Major bands were identified as true glutelin 5' regions of the species by Southern blot analysis using the 5' region from a glutelin gene isolated from Tainung 67 (cultivated rice of genome AA) as probe (Figure lB). There were two major bands, 1.2 kb and 0.9 kb in length, amplified from each rice species except that from *O. eichingeri* and *O. officinalis* in which the 0.9 kb band was substituted by a 0.5 kb band (Figure lB). Only the 0.9 kb band has been shown to appear in glutelin genes of the cultivated rice (Takaiwa et al., 1987b, 1991). The other two bands are being described in thispaper for the first time.

Sequencing of the seven cloned 5' regions of glutelin genes from the wild species of five genomes revealed their lengths (Table 1) which correspond to their molecular weight estimated from the gel electrophoresis of the PCR products. These sequences can be grouped into three categories according to the length of the PCR amplified DNAs. These are 1.2 kb, 0.9 kb, and 0.5 kb respectively.

All seven 5' region sequences were arranged to give maximum alignment as shown in Figure 2. Each sequence listed in Figure 2 has at its upstream 5' end AGCTT or GCTT that is part of the cutting site of Hind III. This and the recovered primer sequence at both ends of the seven sequences assure that the amplified sequences are the true 5' regions of glutelin genes. The figure also includes the sequences of the *Gt1*, *Gt2,* and *Gt3* reported by Okita et al. (1989). Comparing the regions with each other, it is clear that the 0.9 kb 5' region has a short deletion of 20 bp from -834 bp to -815 bp (from the translation initiation codon ATG) and another long deletion of 204 bp from -470 bp to -267 bp. As for the 0.5 kb 5' region, it has two long deletions; one is 463 bp in length from -1055 bp to -593 bp, and another is 211 bp, from -470 bp to -260 bp, that coincides with the long deletion in the 0.9 kb 5' region. Thus the 0.9 kb 5' region is about 200 bp shorter than the 1.2 kb 5' region, and the 0.5 kb 5' region is about 670 bp shorter than the 1.2 kb 5' region. Such long deletions combined with minor deletions of bases account for the actual length (in bp) of each 5' region sequence listed in Table 1.

There are 209 base-substitutions or -deletions that have occurred to corresponding positions in the sequence of the 1.2 kb and the 0.9 kb 5' regions. The same amount of such corresponding substitutions and deletions can also be

Table 1. 5' region sequence length of glutelin genes from various species of rice.

Species*		Genome	Length (kb) estimated from gel electrophoresis	Actual length (bp)
Oryza perrennis	(W0107)	AA	1.2	1,119
Oryza eichingeri	(W1519)	CC	1.2	1,116
Oryza punctata	(W1564)	BBCC	1.2	1,111
Oryza punctata	(W1564)	BBCC	0.9	911
Oryza grandiglumis	(W1194)	CCDD	0.9	913
Oryza australiensis	(W0008)	EE	0.9	912
Oryza sativa	(*Gt1*)	AA	—	779
Oryza sativa	(*Gt2*)	AA	—	878
Oryza eichingeri	(W1519)	CC	0.5	481
Oryza sativa	(*Gt3*)	AA	—	842

*Only some of the species indicated in Figure 1 was chosen to be cloned and sequenced. The length of the 5' region of *Gt1*, *Gt2,* and *Gt3* was calculated based on the sequences published (Okita, 1989). The sequence of the 5' region of glutelin gene of each wild rice species has been deposited to the DataBank of Japan (DDBJ) Tsukuba, Japan. The given accession number of each sequence is as follows: clone W0107-1.2 : D26363; clone W1564-0.9 : D26364; clone W1564-1.2 : D26365; clone W1519-0.5 : D26366; clone W1519-1.2 : D26367; clone W1194-0.9 : D26368; clone W0008-0.9 : D26369.

```
W0107120  Oryza perennis (AA)                                     AGCT -1151
W1519120  Oryza eichingeri (CC)                                   ..GC
W1564120  Oryza punctata (BBCC)                                   ...T
W1564085  Oryza punctata (BBCC)                                   ...T
W1194095  Orzya grandigiumis (CCDD)                               AGCT
W0008095  Oryza australiensis (EE)                                ..CT
     Gt1  Oryza sativa (AA)                                       ....
     Gt2  Oryza sativa (AA)                                       ....
W1519050  Oryza eichingeri (CC)                                   AGCT
     Gt3  Oryza sativa (AA)                                       ....

W0107120  TTTGGAAAGG TGCCGTGCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG -1101
W1519120  TTTGGAAAGG TGCCGTGCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG
W1564120  TTTGGAAAGG TGCCGTGCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG
W1564085  TTTGGAAAGG TGCCGTGCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG
W1194095  TTTGGAAAGG TGCCGTGCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG
W0008095  TTTGGAAAGG TGCCGTGCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG
     Gt1  .......... .......... .......... .......... ..........
     Gt2  .......... .......... .......... .......... ..........
W1519050  TTTGGAAAGG TGCCGTGCCG TTCAAACAAT TAGTTAGCAG TAGGATGTTG
     Gt3  .......... .......... .......... .......... ..........

W0107120  ATTTTTGCAC ATGGCAATGA GAAGTTAATT ATGGTGTAGG CAACCCAAAT -1051
W1519120  ATTTTTGCAC ATGGCAATGA GAAGTTAATT ATGGTGTAGG CAACCCAAAT
W1564120  A.TTTTGCAC AT.GCAATGA GAAGTTAATT ATGGTGTA.G CAACCCAAAT
W1564085  ATTTTTGCAC ATGGCAATGA GAAGTTAATT ATGGTGTAGG CAACCCAAAT
W1194095  A.TTTTGCAC ATGGCAATGA GAAGTTAATT ATGGTGTAGG CAACCCAAAT
W0008095  ATTTTTGCAC ATGGCAATGA GAAGTTAATT ATGGTGTAGG CAACCCAAAT
     Gt1  .......... .......... .......... .......... ..........
     Gt2  .......... .......... .......... .......... ..........
W1519050  GCTTTTGCTC ACAGCAATGA GAAGTTAATT ATGGTGTACG CGTGA.....
     Gt3  .......... .......... .......... .......... ..........

W0107120  GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA -1001
W1519120  GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA
W1564120  GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA
W1564085  GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA
W1194095  GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA
W0008095  GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA
     Gt1  .......... .......... .......... .....TTCTG TAGTACAGAC
     Gt2  .......... .......... .......... .....TTCTG TAGTACAGAA
W1519050  .......... .......... .......... .......... ..........
     Gt3  .......... .......... .......... .......... ..........

W0107120  TAAACT.AAA GTAATGAAAG AAGA..TGGT GTTAGAAAAT GAAACAATAT -951 Inverted rept
W1519120  TAAACT.AAA GTAATGAAAG AAGA..TGGT GTTAGAAAAT GAAACAATAT      Enhancer core
W1564120  TAAACT.AAA GTAATGAAAG AAGA..TGGT GTTAGAAAAT GAAACAATAT
W1564085  TAAACT.AAA GTAATGAAAG AAGATGTGGT GTTAGAAAAG GAAACAATAT
W1194095  TAAACT.AAA GTAATGAAAG AAGATGTGGT GTTAGAAAAG GAAACAATAT
W0008095  TAAACT.AAA GTAATGAAAG AAGATGTGGT GTTAGAAAAG GAAACAATAT
     Gt1  AAAACTAAAA GTAATGAAAG AAGATGTGGT GTTAGAAAAG GAAACAATAT
     Gt2  TAAACT.AAA GTAATGAAAG AAGA..TGGT GTTAGAAAAT GAAACAATAT
W1519050  .......... .......... .......... .......... ..........
     Gt3  .......... .......... .......... .......... ..........

W0107120  TATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAATAAAA AAAGAACATT -901 RY repeat 7
W1519120  TATGACTAAT GTGTGAGCAT TATGGGACCA CGAAATAAAA AAAGAACATT
W1564120  TATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAATAAAA AAAGAACATT
W1564085  CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAAT  AA AAAGAACATT
W1194095  CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAAT  AA AAAGAACATT
W0008095  CATGAGTAAT GTGTGAGCAT TAT.GGACCA CGAAAT  AA AAAGAACATT
     Gt1  CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAAT  AA AAAGAACATT
     Gt2  TATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAATAAAA AAAGAACATT
W1519050  .......... .......... .......... .......... ..........
     Gt3  .......... .......... .......... .......... ..........

W0107120  TTTATGAGCA GTGTGTTCTC AATGAGCCTT GAATGTTATC TCACCCAGGA -851
W1519120  TTTATGAGCA GTGTGTTCTC AATGAGCCTT GAATGTTATC TCACCCAGGA
W1564120  TTTATGAGCA GTGTGTTCTC AATGAGCCTT GAATGTTATC TCACCCAGGA
W1564085  TTGATGAGTC GTGTATCCTC GATGAGCCTC AAAAGTTCTC TCACCCGGGA
W1194095  TTGATGAGTC GTGTATCCTC GATGAGCCTC AAAAGTTCTC TCACCCGGGA
W0008095  TTGATGAGTC GTGTATCCTC GATGAGCCTC AAAAGTTCTC TCACCCGGGA
     Gt1  TTGATGAGTC GTGTATCCTC GATGAGCCTC AAAAGTTCTC TCACCCGGGA
     Gt2  TTTATGAGCA GTGTGTTCTC AATGAGCCTT GAATCT..TA TCACCCAGGA
W1519050  .......... .......... .......... .......... ..........
     Gt3  ........GT TAGTGTGCAA TGTAACTGTA GCTTCTTATA GCTTAGTGCT

W0107120  TAAGAAACCC TTTAGCAATG AAACATGCAA GCGTTTAATG TGCAAAGTTG -801 RY repeat 6,
W1519120  TAAGAAACCC TTTAGCAATG AAACATGCAA GCGTTTAATG TGCAAAGTTG      -900 bp 6
W1564120  TAAGAAACCC TTTAGCAATG AAACATGCAA GCGTTTAATG TGCAAAGTTG      7 bp rept 6
W1564085  TAAGAAACCC TTAAGC                          AATG TGCAAAGTTT
W1194095  TAAGAAACCC TTAAGC                          AATG TGCAAAGTTT
W0008095  TAAGAAACCC TTAAGC                          AATG TGCAAAGTTT
     Gt1  TAAGAAACCC TTAAGC                          AATG TGCAAAGTTT
     Gt2  TAAGAAACCC TTAAGCAATG AAACATGCAA GCGTTTAATG TGCAAAGTTG
W1519050  .......... .......... .......... .......... ..........
     Gt3  TTACTATCTT CACAAGCACA TGCTATAGTA TTGTTCCAAG ATCAAAGAAT
                            RY rept 6
```

```
                                                                        7 bp rept 5
W0107120  GCATTCTCCA C.GACATAAT GCAAAAGAAG ATATAATCTA TGACATAGCA -751 -900 bp 5,
W1519120  GCATTCTCCA C.GACATTAT GCAAAAGAAG ATATAATCTA TGACATAGCA      Inverted rept,
W1564120  GCATTCTCC. ...ACATAAT GCAAAAGAAG ATATAATCTA TGACATAGCA      Direct rept 4
W1564085  GCATTCTCCA CTGACATAAT GCAAAATAAG ATATCATTGA TGACATAGCA
W1194095  GCATTCTCCA CTGACATAAT GCAAAATAAG ATATCATTGA TGACATAGCA
W0008095  GCATTCTCCA CTGACATAAT GCAAAATAAG ATATCATTGA TGACATAGCA
     Gt1  GCATTCTCCA CTGACATAAT GCAAAATAAG ATATCATCGA TGACATAGCA
     Gt2  GCATTCTCCA C.GACATAAT GCAAAAGAAG ATATAATCTA TGACATAGCA
W1519050  .......... .......... .......... .......... ..........
     Gt3  AATTCATCCT TGCTACCAAC TTGCATGATA TTATATTTGT GAATATCCTA
                                 RY rept 5
          Box VI
W0107120  AGTCATGCAT CATTTCATGC CTCTGTCAAC CTATTCATTT CTAGTCATCT -701 RY repeat 5, 4
W1519120  AGTCATGCAT CATTTCATGC CTCTGTCAAC CTATTCATTT CTAGTCATCT
W1564120  AGTCATGCAT CATTTCATGC CTCTGTCAAC CTATTCATTT CTAGTCATCT
W1564085  ACTCATGCAT CATATCATGC CTCTCTCAAC CTATTCATTC CTACTCATCT
W1194095  ACTCATGCAT CATATCATGC CTCTCTCAAC CTATTCATTC CTACTCATCT
W0008095  ACTCATGCAT CATATCATGC CTCTCTCAAC CTATTCATTC CTACTCATCT
     Gt1  ACTCATGCAT CATATCATGC CTCTCTCAAC CTATTCATTC CTACTCATCT
     Gt2  AGTCATGCAT CATTTCATGC CTCTGTCAAC CTATTCATTT CTAGTCATCT
W1519050  .......... .......... .......... .......... ..........
     Gt3  TCTCTTGGCT TAT...AATG AAATGTGCTG CTGGGTTATT CTGACCATGG
                                                       RY rept 4
                                                        Box V
W0107120  AGGTAAGTAT CTTAAGCTAA AGTGTTAGAA CTTCCCATAC ATAAGTCATA -651 Inverted rept
W1519120  AGGTAAGTAT CTTAAGCTAA AGTGTTAGAA CTTCCCATAC ATAAGTCATA
W1564120  AGGTAAGTAT CTTAAGCTAA AGTGTTAGAA C.TCCCATAC ATAAGTCATA
W1564085  ACATAAGTAT CTTGAGCTAA A.TGTTAGAA CAT.AAACCC ATAAGTCACG
W1194095  ACATAAGTAT CTTGAGCTAA A.TGTTAGAA CAT.AAACCC ATAAGTCACG
W0008095  ACATAAGTAT CTTGAGCTAA A.TGTTAGAA CAT.AAACGC ATAAGTCACG
     Gt1  ACATAAGTAT CTTCAACTAA A.TGTTAGAA CAT.AAAGGC ATAAGTCACG
     Gt2  AGGTAAGTAT CTTAAACTAA AGTGTTAGAA CTTCCCATAC ATAAGTCATA
W1519050  .......... .......... .......... .......... ..........
     Gt3  TATTTGAGAG CCTTTGTATA GCTGAAACCA ACGTATATCG AGCATGGAAC
                                                  RY rept 3

W0107120  ACTGATGACA ATTGGGTGTA ACACATGACA AACCAGAGAG TCAAG..... -601 RY repeat 3
W1519120  ACTGATGACA ATTGGGTGTA ACACATGACA AACCAGAGAG TCAAG.....
W1564120  ACTGATGACA ATTGGGTGTA ACACATGACA AACCAGAGAG TCAAG.....
W1564085  TTTGATGAGT ATTAGGGGTG ACACATGACA AATCAGAGAC TCAAG.....
W1194095  TTTGATGAGT ATTAGGGGTG ACACATGACA AATCACAGAC TCAAG.....
W0008095  TTTGATGAGT ATTAGGGGTG ACACATGACA AATCACAGAC TCAAG.....
     Gt1  TTTGATGAGT ATTAGGGGTG ACACATGACA AATCAGAGAC TCAAG.....
     Gt2  ACTGATGACA ATTGGGTGT. ACACATGACA AACCAGAGAG TCAAG.....
W1519050  .......... .......... .......... .......... ..........
     Gt3  AGAGAAGAAA ATGCAAGGAT TTTTTTATTC TGGTTCATGC CCTGGATGGG
                                         RY rept 2

W0107120  .......CAA GATAAAGCAA AAGGATGTGG TACATAAAAC TACAGAGCTA -551 7 bp rept 4
W1519120  .......CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCTA
W1564120  .......CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCTA
W1564085  .......CAA GATAAAGCAA AATGATGT.G TACATAAAAC TCCAGAGCTA
W1194095  .......CAA GATAAAGCAA AATGATGT.G TACATAAAAC TCCAGAGCTA
W0008095  .......CAA GATAAAGCAA AATGATGT.G TACATAAAAC TCCAGAGCTA
     Gt1  .......CAA GATAAAGCAA AATGATGT.G TACATAAAAC TCCAGAGCTA
     Gt2  .......CAA GATAAAGCAA AATGATGT.G TACATAAAAC TACAGAGCTA
W1519050  .......CAC CATAAAGCAA AAGGATGT.G TACAAAAAAC TCCAGAGCTA
     Gt3  TTAATATCGT GATCATCAAA AAAGATATG. ..CATAAAAT TAAAGTAATA

                                                                         7 bp rept 3
W0107120  TATGTCATGT TGCGAAAAGA GGAGAGCTTA TAAGACAAGC CATGACTCAA -501 -900 bp 4,
W1519120  TATGTCATGT TGCGAAAAGA G..GAGCTTA TAAGAAAAGC CATGACTCAA      Direct rept 3
W1564120  TATGTCATGT TGCGAAAAGA GGAGAGCTTA TAAGACAAGC CATGACTCAA
W1564085  TATGTCATAT TGCAAAAAGA GGAGAGCTTA TAAGACAAGG CATGACTC.A
W1194095  TATGTCATAT TGCAAAAAGA GGAGAGCTTA TAAGACAAGG CATGACTC.A
W0008095  TATGTCATAT TGCAAAAAGA GGAGAGCTTA TAAGACAAGG CATGACTC.A
     Gt1  TATGTCATAT TGCAAAAAGA GGAGAGCTTA TAAGACAAGG CATGACTC.A
     Gt2  TATGTCATGT TGCGA.AAGA GGAGAGCTTA TAAGACAAGG CATGACTC.A
W1519050  TATGTCATAT TGCAAACAGA GGAGAGCCTA TAAGACA.GC CATGACTCAA
     Gt3  AATTTGCTCA TAAGAAACCA AAA......C CAAAAGGACA TATGTCCTAA
                       Direct rept 4           Inverted rept

W0107120  AAAAAATTCA CATGCCTACT GTGGCCCATA TATCATGCAA CAATGCAAAA -451 RY repeat 2,
W1519120  AAAAAATTCA AATGCCTACT GTGGCCCATA TATCATGCAA CAATCCAAAA      Inverted rept
W1564120  AAAAAATTCA CATGCCTACT GTGGCCCATA TATCATGCAA CAATCCAAAA
W1564085  CAAAAATTCA TTTGCCTTTC GTGT.CAAAA
W1194095  CAAAAATTCA TTTGCCTTTC GTGT.CAAAA
W0008095  CAAAAATTCA TTTGCCTTTC GTGT.CAAAA
     Gt1  CAAAAATTCA CTTGCCTTTC GTGT.CAAAA
     Gt2  AAAA...CCA ATCGCCTACT GTGGCCCATA TATCATGCAA TAATCCAAAA
W1519050  GAAAAA.TCA TTTGCCTTTC GTGTACAAAA .......... ..........
     Gt3  ACAAACTGCA TTTTGTTTGT CATGTAGCAA TACAAGAGA. ..........
                            RY rept 1

W0107120  ACTCACAGGT CTCGGTGTTG ATCGTGTCAA CATGTGACCA CCCTAAAAAC -401 Inverted rept
W1519120  ACTCACAGGT CTCGGTGTTG ATCGTGTCAA CATGTGACCA CCCTAAAAAC
W1564120  ACTCACAGGT CTCGGTGTTG ATCGTGTCAA CATGTGACCA CCCTAAAAAC
W1564085                                                   ...
W1194095                                                   ...
W0008095                                                   ...
     Gt1                                                   ...
     Gt2  AGTCACAGGT CTCGGTGTTG ATCGTGTCAA CATGTGACCA CCTAAAA...
W1519050  .......... .......... .......... .......... ..........
     Gt3  .......... .........T AATATATGAC GTGGTTATGA CTTATTCACT
```

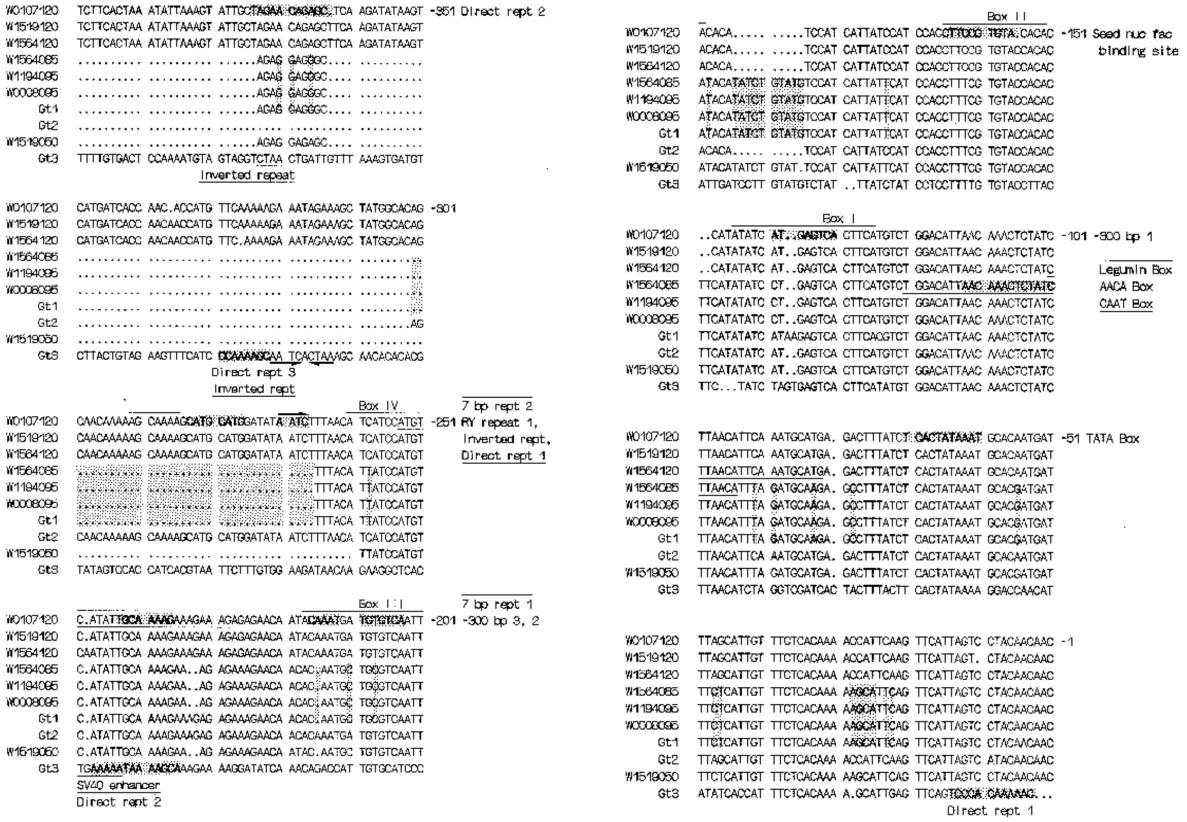

Figure 2. The seven 5' region sequences are maximally aligned with each other (using a GCG sequence analysis software package) including the three 5' regions of Gt genes (Okita et al., 1989) and arranged in descending order starting from the most 5' end. The dot within the sequence denotes base deletion. The shaded, upper, or underlined bases stand for the repeat, element, box, etc., the name of which is listed to the right side. A pair of arrows indicates inverted repeats. The vertical bar represents base substitution or deletion in the 0.9 kb and *Gt1* 5' region with respect to that of the 1.2 kb and *Gt2*. The *Gt3* 5' region has its own repeats, etc. with its name indicated below the sequence.

found between the *Gt1* and *Gt*2 sequences. The substituted or deleted bases in the 5' region sequence of *Gt1* are like those in the sequence of the 0.9 kb 5' region (marked with a vertical bar in Figure 2). The same can be found between the sequences of *Gt2* and the 1.2 kb region. For example, at the -976 bp position the 0.9 kb and *Gt1* sequences have the same T and the 1.2 kb and *Gt2* sequence have a base deleted in common. Furthermore, many boxes, repeats, and elements of the 0.9 kb and *Gt1* sequence and of the 1.2 kb and *Gt2* sequence have similar correspondence (Figure 2, Table 2). As a result, the 5' region of *Gt1* and *Gt2* has a 24.39% and 21.64% higher homology respectively to that of the 0.9 kb and the 1.2 kb than the homology between the 0.9 kb and the 1.2 kb 5' region. Due to the two long deletions in the 0.5 kb 5' region, it has 66 instead of 209 base-substitutions or -deletions. Among them, 36 bases are the same as the correspondent bases in the 0.9 kb 5' region; 9 bases follow that in the 1.2 kb 5' region, and 21 bases follow that in neither of the two 5' regions (Figure 2).

In view of the long direct repeats, four have been identified in *O. sativa* cultivar Mangetsumochi (Takaiwa et al., 1987b) and three in *O. sativa* cultivar M201 (Okita et al., 1989). Figure 2 shows the sequences and positions of the four direct repeats. All four are conserved in the sequences examined except that the 0.5 kb sequence lacks direct repeat 4 and *Gt2* lacks direct repeat 2. *Gt3* sequence has its own four direct repeats, and their positions do not coincide with those identified in the other sequences. Its first direct repeat is located from -15 to -4 bp relative to ATG.

The SV40 enhancer TGAAAAA, identified in the *Gt3* sequence (Okita et al., 1989), is shown to be superimposed with its own direct repeat 2, located from -248 bp to -236 bp, but can not be found in the other 5' region sequences. However, one kind of 7 bp direct repeat (T/AGCA/GAAA/G) with high homology to the SV40 enhancer can be found in the 5' region sequences examined. There are six such short direct repeats in the 1.2 kb sequence (Figure 2), two of them (repeats 2 and 4) are independent. The TGCA motifs of the remaining four are superimposed upon those of their respective -300 bp elements and/or long direct repeats. The 0.5 kb sequence lacks the short direct repeats 6, 5, and 2; the 0.9 kb sequence and *Gt1* lack short direct repeat 2. *Gt3* also has six short direct repeats but with low or very low homology to that of the 7 bp direct repeat located at the corresponding positions in the 5' region of the various wild rice species.

A -300 bp element was reported to be present in the 5' region of barley prolamin gene (Forde et al., 1985). There are six -300 bp elements that can be traced in the 5' re-

Table 2. Base substitution, deletion and homology in the boxes, repeats and elements of glutelin gene flanking 5' region.

		AA CC BBCC	BBCC CCDD EE	*Gt1*	*Gt2*	CC	*Gt3*
		(1.2 kb)	(0.9 kb)			(0.5 kb)	
Direct repeat	4	•GACATAATGCAAAA**G**	T, T	H 0.9 kb[b]	H 1.2 kb	Del	VLH[a]
	3	ATGTCAT**G**TTGCGAAA AGAGGAGAG	A	H 0.9 kb	H 1.2 kb	H 0.9 kb	VLH
	2	TAG**A**ACAG**A**GC	G, G	H 0.9 kb	Del	H 0.9 kb	VLH
	1	ATGTCATATTGCAAAAGAAA**AG**AAAG	••	H 1.2 kb	H 1.2 kb	H 0.9 kb	VLH
Box	VI	**G**TCA	C	H 0.9 kb	H 1.2 kb	Del	VLH
	V	TAAGTCA**TAAC**TGATGA	CGTT	H 0.9 kb	H 1.2 kb	Del	VLH
	IV	AT**C**ATCCATGTCATATTG	T	H 0.9 kb	H 1.2 kb	H 1.2 kb	VLH
	III	AC**A**AATG**A**TG**T**GTCAATTA	•, C, C	H 0.9 kb	H 1.2 kb	H 1.2 kb	VLH
	II	CTTCCGTGTACCACA	Conserved	Conserved	Conserved	Conserved	HH
	I	ATATCAT••GAGTCACTTCA	Conserved	-AA	Conserved	Conserved	HH
-300 bp element	6	TGCAAAGTT	Conserved	Conserved	Conserved	Del	LH
	5	TGCAAAA**G**	T	H 0.9 kb	H 1.2 kb	Del	NH
	4	TGC**G**AAAG	A	H 0.9 kb	H 1.2 kb	H 0.9 kb	LH
	3	TGCAAAAG	Conserved	Conserved	Conserved	conserved	LH
	2	C**A**AA TG**T**GTCA	•, C	H 0.9 kb	H 1.2 kb	H 0.9 kb	LH
	1	**A**T••GAGTCA	C	AA	H 1.2 kb	H 1.2 kb	LH
RY repeat	7	**T**ATG	C	H 0.9 kb	H 1.2 kb	Del	NF
	6	CATGCAAG	Del	Del	H 1.2 kb	Del	H
	5	CATGCATC	Conserved	Conserved	Conserved	Del	H
	4	CATG	Conserved	Conserved	Conserved	Del	H
	3	CATG	Conserved	Conserved	Conserved	Del	H
	2	CATG	Del	Del	H 1.2 kb	Del	H
	1	CATGCATG	Del	Del	H 1.2 kb	Del	H
Enhancer core		•TGGTGTT	G	H 0.9 kb	H 1.2 kb	Del	Del
13 bp AACA box		AACAAACTCTATC	Conserved	Conserved	Conserved	Conserved	
Legumin box		CTTAACATT**C**A**A**ATGA**T**G	T, G, A	H 0.9 kb	H 1.2 kb	H 0.9 kb	H 0.9 kb
Immuture seed nuc fac BS		CTTCCGTGTA	Conserved	Conserved	Conserved	Conserved	HH
CAAT Box		TGGACATTAACAAACTCTATCTTAACA	Conserved	Conserved	Conserved	Conserved	
Inverted repeat		AATC (-271)	Del	Del	H 1.2 kb	Del	AATC (-324)
		AATC (-459)	Del	Del	H 1.2 kb	Del	NF
		AATC (-766)	C	H 0.9 kb	H 1.2 kb	Del	NF
		CTAA(-408)	Del	Del	H 1.2 kb	Del	CTAA (-317)
		CTAA(-684)	Conserved	H 1.2 kb	H 1.2 kb	Del	CTAA (-374)
		CTAA(-996)	Conserved	H 1.2 kb	H 1.2 kb	Del	CTAA (-504)
SV40 enhancer		Not found	Not found	Not found	Not found		TGAAAAA
7 bp direct repeat	6	TGCAAAG	Conserved	Conserved	Conserved	Del	LH
	5	TGCAAAA	Conserved	Conserved	Conserved	Del	VLH
	4	AGCAAAA	Conserved	Conserved	Conserved	Conserved	LH
	3	TGC**G**AAA	A	H 0.9 kb	H 1.2 kb	H 0.9 kb	LH
	2	AGCAAAA	Del	Del	H 1.2 kb	Del	LH
	1	TGCAAAA	Conserved	Conserved	Conserved	Conserved	LH

[a]NH stands for nonhomology; LH, low homology; VLH, very low homology; HH, high homology of *Gt3* 5' region sequence to the corresponding sequence of 1.2 kb 5' region of AA, CC and BBCC genomes. H, homology; NF, not found; Del, deleted. The *Gt3* 5' region has its own direct repeat 4, CCAAAACCAAAAGCA; direct repeat 3, CCAAAAGCA; direct repeat 2, AAAAATAAAAGCA and direct repeat 1, TCCCACAAAAAC and six RY repeats, CATG and four inverted repeats. Positions of these repeats do not coincide with those found in other 5' region sequencces. Boldfaced bases in the lane of 1.2 kb 5' region are shown to have been respectively substituted by the base or deleted shown in the lane of 0.9 kb , for example, **.**-T, **G**-T., **AG**-.. etc.

[b]Sequence homology of *Gt1* 5' region to that of 0.9 kb 5' region.

gion of glutelin genes in the wild rice species. The sixth (located from -810 bp to -802 bp) is independent; five of them, i.e., the elements 5, 4, 3, 2, and 1 are superimposed upon part of the Takaiwa's long direct repeat 4, 3 and upon the sequence of Box IV III and I respectively, which are the bind site of nuclear proteins as reported (Kim and Wu, 1990). In addition to these, superimposition can be found between the -300 bp elements 6, 5, 4, 3, and the four 7 bp repeats. Each of the six -300 bp elements have been conserved with minor base substitutions in the sequences examined but not in the 0.5 kb 5' region, in which it lacks -300 bp elements 6 and 5 . Only low homology remains in most of the elements in *Gt3* (Figure 2, Table 2).

Five protein binding sequences (boxes; Kim and Wu, 1990), have been well conserved because they can be found in all the sequences examined except in the 0.5 kb 5' region that lacks Box V. In addition to the five boxes, we found a new one, the sixth box with its core motif GTCA from -749 bp to -746 bp in the 1.2 kb 5' region, but the first nucleotide G of the motif has mutated to C in the case of 0.9 kb 5' region. In the 0.5 kb 5' region, the fifth and sixth boxes have been deleted. In *Gt*3, only boxes Iand II are conserved (Figure 2 , Table 2).

In the 5' region sequences of 1.2 kb , there are seven RY repeats. The seventh is between -950 bp and -947 bp and has a nucleotide C mutated to T. From -274 bp to -267 bp, the first RY repeat CATGCATG can be traced. Between these two positions, it accommodates the other five RY repeats (repeats 6, 5, 4, 3, and 2). The 0.9 kb 5' region lacks RY repeats 1, 2, and 6. None of the RY repeats can be found in the 5' region sequence of 0.5 kb. Okita et al. (1989) reported that there were respectively one, two and one RY repeat in the 5' region of *Gt1*, *Gt2,* and *Gt3*. In Figure 2, however, it is shown that *Gt1* and *Gt2* have four and seven RY repeats respectively. *Gt3* has six RY (all CATG) repeats dispersed in a segment of 310 bp, from -780 bp to -476 bp. None is located at the same positions shown in the other 5' region sequences (Figure 2 and Table 2). It is interesting to note that the number of RY repeats in the various 5' regions vary substantially among species and that some of the repeats are superimposed with the -300 bp element and the long direct repeat.

Okita et al. (1989) identified two pairs of inverted repeats (AATC and CTAA) in the 5' region of *Gt2* and suggested that the DNA segments between the two components of the two repeat pairs could have been transposed from somewhere else. Figure 2 shows that the position of the three inverted repeat pairs we found in the sequence of 1.2 kb 5' region start at -271 bp, -408 bp; -459 bp, -684 bp; and -766 bp, -996 bp, respectively. It is interesting to note that the sequence between -271 bp and -408 bp including the inverted repeat pair has been deleted in the 0.9 kb and 0.5 kb 5' region (Figure 2, Table 2). In the *Gt3* sequence, we identified one AATC starting at -322 bp but three CTAA at -317 bp, -374 bp, and -504 bp. In the case of 0.5 kb sequences, all the inverted repeats have been deleted.

Proximal to the translation initiation codon, ATG, within a range of -1 to -261 bp, the 13 bp AACA Box (AACAAACTCTATC, Takaiwa and Oono, 1991), the legumim Box (CTTAACATTTAGATGCAAG, Takaiwa et al., 1987b), the CAAT Box (TGGACATTAACAA ACTCTATCTTAACA, Okita et al., 1989) the immature seed nuclear factor binding site (CTTTCGTGTA, Takaiwa and Oono, 1990), and Boxes I, II, III and IV of Kim and Wu (1990) all reside in and are well conserved in each of the 5' region sequences examined, in addition to the TATA Box (TCACTATAAAT). The first three of the above mentioned boxes are superimposed.

Discussion

Rice glutelin genes have numerous putative enhancers dispersed though the 5' flanking region within a span of about 950 bp. Beyond the four long direct repeats, the 7 bp short direct repeat may have a putative function as enhancer because its sequences are similar to that of the SV40 enhancer that was superimposed upon the direct repeat 2 in the 5' region of *Gt3* (Okita et al., 1989). In this study, the 7 bp direct repeat 3 and the long direct repeat 3 are superimposed upon each other in all the sequences examined. SV40 enhancer core homology was observed in gliadin (wheat storage protein) genes (Reeves and Okita, 1987). Chen et al. (1986) reported that 4 short direct repeats of AA/GGCCA in the 5' region of β-conglyccinin α subunit could increase expression 20 fold. It would be worthwhile to see how each of the many direct repeats enhances the expression of rice glutelin genes.

Two -300 bp elements were found in the 5' region of prolamin (B1 hordein) genes of barley. The elements are composed of the conserved core motif GTCATG and were proposed to be endosperm specific (Forde et al., 1985). In the 5' region of rice glutelin genes, two (elements 1 and 2, Table 2) of the six -300 bp elements have GTCA motif; the other four (elements 3 to 6, Table 2) are similar to the two -300 bp elements (TGTAAAGT and TGTAAAAG) that are endosperm specific in wheat prolamin (LMW glutenin) genes (Colot et al., 1987). The differential functions of the two groups of the sequences that are similar to that of barley and wheat respectively in glutelin gene expression are not known. Superimposition of the four -300 bp elements (elements 3, 4, 5, and 6) respectively upon the four 7 bp short direct repeats (repeats 1, 3, 5, and 6) and the two long direct repeats combined with the fact that the elements 4, 5, and 6 are dispersed at relatively far 5' upstream localities (extending from -530 bp to -810 bp) may suggest that the four -300 bp elements play a role as enhancer. A sequence similar to that of the SV 40 enhancer however, did not significantly increase the level of expression of soybean storage protein genes (Chen et al., 1986). The superimposition of -300 bp element 1 and 2 upon that of nuclear protein binding boxes I and IV (Kim and Wu, 1990) may imply that some nuclear protein molecules are needed to bind GTCA core motif and enable the two -300 bp elements to carry out endosperm specific expression.

In the 5' region of 1.2 kb, we identified one more pair of inverted repeats that was not found in *Gt2* by Okita et al. (1989). The pair, located at -408 bp and -271 bp and found in the 1.2 kb 5' region (Figure 2), might be responsible for a ~200-base deletion occuring in the 0.9 kb and 0.5 kb 5' regions. The fact that the 0.5 kb and 0.9 kb 5' regions lost three and one and a half pairs of inverted repeats respectively (Table 2) and that the large blocks of deletions border on the inverted repeat may reflect the involvement of the lost inverted repeat pairs in deletion. In CC genomic species (*officinalis* and *eichingeri*) the 0.9 kb 5' region was replaced by that of the 0.5 kb (Figure 1B). It might be inferred that the 0.5 kb 5' region was derived from that of the 0.9 kb due to the occurrence of large blocks of deletion.

The RY repeat (CATGCATG) plays a role in the regulation of legume seed protein gene expression (Dickinson et al., 1988). The seven RY repeats (four of them have the CATG motif only) in the 5' region of glutelin genes are not the only case of multiple copies of the RY repeat present in the 5' region of several legume seed protein genes (Dickinson et al., 1988). Whether the 0.5 kb 5' region of glutelin genes that totally lacks all seven RY repeats— beyond one long direct repeat, one nuclear protein binding site, one enhancer core, two -300 bp elements, and three 7 bp direct repeats (Table 2)—could still perform its normal regulation of glutelin gene transcription is not known.

Legumin, CAAT, AACA, and TATA boxes in all the ten 5' region sequences of glutelin genes observed (including *Gt1*, *Gt2,* and *Gt3*) are located between nucleotide -1 and -150. They are well conserved and aligned. Furthermore, superimposition of the first three boxes upon each other is shown in Figure 2. Legumin box is an element that may have a function in the regulation of legumin gene expression in pea (Baunnlein et al., 1986). The function of this box in rice glutelin genes has not been determined. The CAAT box (TGTTGACAATTT) was designated as the site where the interaction between specific factor and RNA polymerase occurs (Benoist et al., 1980). In rice glutelin genes, the CCAAT sequence was shown to have no significant homology to the eukaryotic model sequence TGTTGACAATTT (Okita et al., 1989). However, the CCAAT-like sequence is important for maximal gene expression in *Kalanchoe* and tobacco plant (Shaw et al., 1984; Odell et al., 1985). CACA box consists mostly of C and A (GTGCCACCAAACACAACATACCAAAA) and was observed in the 5' region of wheat gliadin genes (Reeves and Okita, 1987) though its function was not known. The 13 bp AACA box (Takaiwa and Oono, 1991) is also CA rich and well conserved. Superimposition of the AACA box upon the legumin box and the CAAT box, located proximal to the TATA box in the 5' region of glutelin genes, may mean that the AACA box has an essential function in the expression of glutelin genes.

Study on expression of glutelin genes has been done by comparing the amount of mRNA at developmental stages of rice endosperm (Okita et al., 1989), and by detecting CAT enzyme (Leisy et al., 1989) or GUS activities (Zhao et al., 1994) regulated by a glutelin gene 5' region in transgenic tobacco. It has also been carried out by transient expression assay using immature rice seed protein (Kim and Wu, 1989). Our sequence analysis showed that the 5' region glutelin genes are different in nature from each other. It would be worthwhile first to compare the expression capacity of the three naturally existing 5' regions. Then an artificial 5' region containing well-designed sets of regulating motifs in the glutelin gene 5' region should be constructed and tested to define the expression capacity of the motifs. A thorough understanding of this may be essential to the genetic engineering of rice glutelin genes sooner or later.

The results of our analysis also help to clarify classification of rice glutelin genes. Okita et al. (1989) postulated three gene subfamilies: *Gt1*, *Gt2,* and *Gt3* for glutelin. Takaiwa et al. (1991) classified the glutelin genes so far that they were isolated and sequenced into two subfamilies, A and B. From our PCR experiments (Figure 1A, B), it can be inferred that there are at least three distinct sequence lengths for the 5' regions of glutelin genes. Southern blot analysis, using specific segments as probes, showed that these 5' region sequences can each be accommodated at specific locations in a genome (data not shown) and may represent three glutelin gene loci. Beyond the length differences, corresponding substitution and deletion of the 209 bases in the 0.9 kb and *Gt1* with respect to the 1.2 kb enabled us to find a closer relationship between the 1.2 kb and *Gt2* and between the 0.9 kb and *Gt1* sequences. Total homology at all these 209 bases between the 0.9 kb and *Gt1*, and between the 1.2 kb and *Gt2* support our suggestion that the 5' region of the 1.2 kb and *Gt2* can be assigned as one locus and that of the 0.9 kb and *Gt1* as another locus in various rice genomes. The same is true of the 5' region of the 0.5 kb in wild rice species. When the coding region sequences of glutelin genes with the known 0.5 kb, 0.9 kb, or 1.2 kb 5' region were compared , it revealed that those coding sequences were highly homologous (90–95%) to each other. This suggests that all these glutelin genes can be grouped into one subfamily, i.e., the *Glua* subfamily.

The length of the three 5' regions of 1.2 kb examined in our analysis varies from 1,111 to 1,119 bp among three species, i.e., *O. perrennis, O. eichingeri,* and *O. punctata* (Table 1). This is due to minor deletion or addition of bases varying in number and position occuring in a unique 5' region. For example, in the 1.2 kb 5' region one-base deletion can be examined at position -364 bp in *O. perrennis*, two-base deletion at -529 bp in *O. eichingeri* but four-base deletion at position -791 in *O. punctata*, etc. (Figure 2). In the case of the 0.9 kb 5' region, a similar situation can be observed. We suggest designating these sequences distributed in different species with minor base deletion, addition, or substitution as alleles of a glutelin locus.

Literature Cited

Baumlein, H., U. Wobus, J. Pustell, and F.C. Kafatos. 1986. The legumin gene family: structure of a B type gene of *Vicia* *faba* and a possible legumin gene specific regulatory element. Nucl. Acid Res. **14:** 2707–2719.

Benoist, C., K. O'Hare, R. Breatnach, and P. Chambon. 1980. The ovalbumin gene - sequence of putative control regions. Nucl. Acids Res. **8:** 127–142.

Chen, Z.L., M.A. Schuler, and R.N. Beachy. 1986. Functional analysis of regulatory elements in a plant embryo-specific gene. Proc. Natl. Acad. Sci. USA **83:** 8560–8564.

Colot, V., L.S. Robert, T.A. Kavanagu, M.W. Bevan, and R.D. Thompson. 1987. Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO **6:** 3559–3564.

Dickinson, C.D., R.P. Evens, and N.C. Nielsen. 1988. RY repeats are conserved in the 5'-flanking regions of legume seed-protein genes. Nuc. Acid. Res. **16:** 371.

Forde, B.G., A. Heyworth, J. Pywell, and M. Kreis. 1985. Nucleotide sequence of a Bi Hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. Nucl. Acids Res. **13:** 7327–7337.

Kim, S.Y. and R. Wu. 1990. Multiple protein factors bind to a rice glutelin promoter region. Nucl. Acids Res. **18:** 6845–6852.

Leisy, D.J., J. Hnilo, Y. Zhao, and T.W. Okita. 1989. Expression of a rice glutelin promoter in transgeneic tobacco. Plant Mol. Biol. **14:** 41–50.

Odell, J.T., F. Nagy, and N.H. Chua. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature **313:** 810–812.

Okita, T.W., Y.S. Hwang, J. Hnilo, W.T. Kim, A.P. Aryan, R. Larson, and H.B. Krishnan. 1989. Structure and expression of the rice glutelin multigene family. J. Biol. Chem. **264:** 12573–12581.

Reeves, C.D. and T.W. Okita. 1987. Analyses of α/β type gliadin genes from diploid and hexaploid wheats. Gene **52:** 257–266.

Shaw, C.H., C.H. Cater, M.D. Watson, and C.H. Shaw. 1984. A functional map of the nopaline synthase promoter. Nucl. Acid. Res. **12:** 7831–7846.

Takaiwa, F., S. Kikuchi, and K. Oono. 1986. The structure of rice storage protein glutelin precursor deduced from cDNA. FEBS Lett. **206:** 33–35.

Takaiwa, F., S. Kikuchi, and K. Oono. 1987a. A rice glutelin gene family- a major type of glutelin mRNAs can be divided into two classes. Mol. Gen. Genet. **208:** 15–22.

Takaiwa, F., H. Ebinuma, S. Kikuchi, and K. Oono. 1987b. Nucleotide sequence of a rice glutelin gene. FEBS Lett. **221:** 43–47.

Takaiwa, F. and K. Oono. 1990. Interaction of an immuture seeds specific trans-acting factor with the 5' upstream region of a rice glutelin gene. Mol. Gen. Genet. **224:** 289–293.

Takaiwa, F. and K. Oono. 1991. Genomic DNA sequence of two new genes for new storage protein glutelin in rice. Jap. J. Genet. **66:** 161–171.

Takaiwa, F., K. Oono, D. Wing, and A. Kato. 1991. Sequence of three members and expression of a new major subfamily of glutelin genes from rice. Plant Mol. Biol. **17:** 875–885.

Zhao, Y., D.J. Leisy, and T.W. Okita. 1994. Tissue-specific expression and temporal regulation of the rice glutelin *Gt3* gene are conferred by at least two spatially separated *cis*-regulatory elements. Plant Mol. Biol. **25:** 429–436.

野生稻穀蛋白基因 5' 區的分析

吳信淦 陳添進 鍾美珠

中央研究院植物研究所

野生稻穀蛋白基因 5' 區構造曾用選值及定序的方法予以定出。該基因的 5' 區除了與基因表現有關的主要序列片段（如豆素、 CAAT 、 AACA 、 TATA 等）之外，尚有很多推論為可增進表現的片段（如長及短的單向重複片段）及調控片段（如 RY 重複片段、 300 bp 片段及能與核蛋白結合的片段等)。各野生稻物種中該基因 5' 區的構造不盡相同，主要是若干大片段的缺失以及序列片段中 209 個對應氮基的取代。從穀蛋白基因 5' 區序列的長度、同源程度以及各序列對應氮基的取代及缺失數據，作者等建議水稻穀蛋白基因族 *Glua* 可區別為三種成員基因，其 5' 區長度各為 0.5 kb 、 0.9 kb 及 1.2 kb 。每一成員基因在各染色體組各占一或多個基因座。野生稻物種中同一成員基因則有因其 5' 區序列中少數氮基發生缺失、外加或取代而形成等位基因 (allele)。

關鍵詞：穀蛋白基因； 5' 區構造；野生稻； *GluA* 次族。