Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 18 (2008) 33–41
Progress in Natural Science

Genetic variations of glycinin subunit genes among cultivated and wild type soybean species

Chunmei Wang, Xiaolei Wu, Faxing Jia, Jinsong Zhang *, Shouyi Chen *

National Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

Received 31 May 2007; received in revised form 30 July 2007; accepted 30 July 2007

Abstract

Glycinin is a predominant storage protein in most soybean accessions. It is a hexamer constituted by five major subunits, which can be classified into two groups. Group I contains G1, G2 and G3, and Group II contains G4 and G5. The genes encoding these subunits have been designated from *Gy1* to *Gy5*, respectively. In the present study, *Gy1* genomic fragments were cloned from wild accessions of subgenera *Glycine glycine*, *Glycine soja* and a cultivar of *Glycine max*. Their sequences and the deduced amino acid sequences were compared. The residues critical for assembling of G1 subunits from the wild perennial accession were conservative. The *Gy4* fragments were cloned from two wild perennial accessions and compared with that from subgenus *Soja*. The intron 3 of *Gy4* had abundant variations between the subgenera *G. soja* and *G. glycine* as well as within the subgenus *G. glycine*. Abundant variations existed in the disordered regions 3 and 4 of G4 subunits from two wild perennial accessions. The genomic organization of glycinin genes was analyzed in 19 accessions from subgenera *Soja* and *Glycine*. The hybridization patterns were identical among the accessions of subgenus *Soja*. On the contrary, abundant polymorphisms existed between the accessions from subgenus *Glycine*. These results indicated that glycinin genes have high degree of conservation within subgenus *Soja* but more variations within subgenus *Glycine*.

Keywords: Genetic variation; Glycinin subunit genes; Soybean

1. Introduction

Glycinin, an 11S globulin, is one of the predominant storage proteins in soybean. In developing seeds, the constituent subunits of 11S globulin are synthesized as a single polypeptide precursor, preproglycinin, the signal sequence of which is removed cotranslationally in the endoplasmic reticulum [1,2]. The resultant proglycinins are assembled into trimers and then transported from the endoplasmic reticulum through the Golgi apparatus to protein storage vacuoles (PSVs), where they are cleaved to form acidic and basic polypeptides that are linked by a disulfide bond [3]. The site of the cleavage is between an asparagine and glycine residue, and has been highly conserved during the evolution of the 11S globulins [4]. It has been demonstrated that the cleavage is a trigger for the formation of the hexamers [4]. The glycinin extracted from cultivated soybean seed has a hexametric structure with a molecular mass of 300–380 kDa. Five major subunits of glycinin have been identified and classified into two groups [5,6]. Group I contains G1 (A1aB1b), G2 (A2B1a) and G3 (A1bB2), and Group II contains G4 (A5A4B3) and G5 (A3B4). Each subunit is composed of an acidic polypeptide with a molecular mass of about 32 kDa and a basic polypeptide with a molecular mass of about 20 kDa. The genes encoding these subunits are named as *Gy1–Gy5*, respectively. The sequence identity is about 80% in each group and about 45% between groups [5,6].

* Corresponding authors. Tel.: +86 10 64886859; fax: +86 10 64873428.
E-mail addresses: jszhang@genetics.ac.cn (J. Zhang), sychen@genetics.ac.cn (S. Chen).

1002-0071/$ - see front matter
doi:10.1016/j.pnsc.2007.07.002

Five variable regions have been proposed based on the sequence alignment among various 11S globulins [7,8]. And five of the six disordered regions correspond roughly to the five variable regions [9]. The crystal structure of a homohexamer of the glycinin G5 subunit has also been determined [10]. The sorting determinants of glycinin, which mediates protein targeting to the protein storage vacuole (PSV), have been investigated in maturing soybean cotyledons by transient expression assays [11]. A C-terminal stretch of 10 amino acids (C-terminal vacuolar sorting determinant, ctVSD), the sequence [ICTMRL] downstream of disordered region 4 and Ile-297 of G1, has been demonstrated to be critical for sorting [11].

Genus *Glycine* is one of the most important genera in legume plants. It consists of subgenera *Soja* and *Glycine*. Subgenus *Soja* includes two species (*Glycine soja* and *Glycine max*). Subgenus *Glycine* includes about 16 wild perennial species. Among them, *Glycine tomentella* is a large group including diploids ($2n = 38$ and 40) and tetraploids ($2n = 78$ and 80). *Glycine tabacina* is also a group of both diploids ($2n = 40$) and tetraploids ($2n = 80$). Previous studies of DNA sequence variations at the histone H3-D and alleles at two isozyme loci [12,13], chloroplast DNA polymorphism [14] and nrDNA internal transcribed spacer ITS loci [15] have provided evidence for multiple origins of the polyploid races.

Weng et al. have cloned *Gy5* cDNA from wild annual soybean *G. soja* and compared with that from *G. max* [16,17]. The comparison revealed 98% similarity at the nucleotide level and 97% identity at the amino acid sequences between them. This indicated that glycinin, as a soybean seed-storage protein, is highly conservative between *G. max* and *G. soja* [17]. Fukuda et al. have found two and five lines containing small A3 and large A4 polypeptides of glycinin, respectively, from 390 lines of wild annual soybeans (*G. soja*) [18]. Mahmoud et al. have investigated the degree of divergence in the DNA sequence of the genes encoding glycinin in the ancestral and modern cultivars released during the past 60 years by Southern blotting and the polymerase chain reaction. The banding patterns were remarkably similar among the ancestral cultivars and those derived from them, suggesting a high degree of conservation of seed-storage protein genes in cultivars [19]. Staswick et al. have conducted a germplasm survey to identify storage protein variants by using SDS–PAGE to compare the electrophoretic mobility of storage proteins among 120 lines of *G. soja* but found no obvious variants. They further analyzed 11S seed-storage proteins from four perennials related to soybean (*Glycine canescens*, *Glycine tomentella*, *Glycine tabacina*, and *Glycine clandestina*). The results indicated that the perennials exhibited variability in their seed proteins at a level not found among the subgenus *G. soja* [20].

Although the glycinin subunit genes in subgenus *Soja* have well been studied, the genetic diversity of these genes in subgenus *Glycine* is less known. In this study, we cloned and compared the partial genomic sequences of the glycinin subunit genes from the cultivated and wild soybean species. The corresponding encoded amino acid sequences were analyzed and the genetic variations of the glycinin genes among different soybean accessions were further investigated.

2. Materials and methods

2.1. Plant materials and DNA extraction

Totally 20 soybean accessions were sampled in this study. The species, collection codes, numbers of chromosomes, karyotypes and origins of the accessions are listed in Table 1. Seeds of the accessions were grown in pots filled with vermiculite. Two-week-old seedlings were harvested and stored at $-70\,^{\circ}C$ for DNA extraction. Genomic DNA extraction was performed as described previously [21]. The genomic DNAs of cultivar Liuyuehuang (*G. max*), wild annual accession Y0074 (*G. soja*), wild perennial accessions PW0031 (*Glycine latifolia*) and PW0052 (*G. tomentella*) were used as templates in the PCR. The genomic DNAs of all the accessions except the cultivar Liuyuehuang (*G. max*) were used in the Southern blotting.

2.2. Cloning and sequencing of the partial genomic sequences of the glycinin subunit genes

The partial genomic fragments of *Gy1* including the entire exon 3, intron 3 and exon 4 were amplified and cloned by PCR. Considering the high similarity of the three

Table 1
Soybean accessions used in this study

Species	Collection code	Number of chromosomes	Karyotype	Origia
Subgenus *Glycine*				
G. latifolia	PW0031	40	B1B1	Australia
G. tabacina	PD0005	40	B2B2	Australia
	PD0021	80	AAB2B2	Australia
	PD0042-3	80	?	China
G. tomentella	PW0049	38	EE	Australia
	PW0052	40	DD	Australia
	PW0054	78	DDEE	Australia
	PW0060	80	AADD	China
	PW0063	80	?	Australia
Subgenus *Soja*				
G. soja	Y0047	40		China
	Y0049	40		China
	Y0069	40		China
	Y0100	40		China
	Y0043	40		China
	Y0073	40		China
	Y0074	40		China
	Y0146	40		China
G. max	Kefeng 1	40		China
	Nannong1138-2	40		China
	Liuyuehung	40		China

subunits in Group I [5,22,23], we designed the forward primer from 48 bp downstream of starting site of the exon 3 according to the *G. max* glycinin gene *Gy1* (X15121) [24]. The specific forward primer was 5′-CAAGAGCAAGG AGGTCATCA-3′ and the reverse primer was 5′-CT AAGCCACAGCTCTCTTCT-3′. The templates were genomic DNAs of wild perennial accession PW0031 (*G. latifolia*) of the subgenus *Glycine*, wild annual accession Y0074 (*G. soja*) and cultivar Liuyuehuang (*G. max*), respectively. PCRs were performed as follows: 35 cycles at 94 °C for 1 min, 56 °C for 1 min, and 72 °C for 1 min.

The partial genomic fragment of *Gy5* including the entire exon 3, intron 3 and exon 4 was amplified and cloned. The primers for amplification of the partial *Gy5* gene were designed according to *G. max* glycinin gene *Gy5* (AB003680) [25]. The forward primer was 5′-GT ATTTTACCTTGCTGGGAA-3′ and reverse primer was 5′-TTATGGGTTGACCACAAGGGC-3′. Because of the high similarity between *Gy4* and *Gy5* [5,25,26], the forward primer was also completely complementary to the first 20 bp of the exon 3 of *Gy4* (X52863) [26]. The reverse primer was partially complementary to the 3′ end of *Gy4*. This pair of primers could amplify specific bands from cultivar Liuyuehuang (*G. max*), the wild perennial accessions PW0031 (*G. latifolia*) and PW0052 (*G. tomentella*), respectively, under different PCR conditions. When the cultivar Liuyuehuang (*G. max*) genomic DNA used as template, PCR was performed as follows: 35 cycles at 94 °C for 1 min, 60 °C for 1 min, and 72 °C for 1 min. When the wild perennial accessions PW0031 (*G. latifolia*) and PW0052 (*G. tomentella*) genomic DNAs were used as templates, PCRs were performed as follows: 35 cycles at 94 °C for 1 min, 54 °C for 1 min, and 72 °C for 1 min 30 s. The amplified fragments were purified and cloned into the pGEM-T easy vector (Promega) and sequenced on the Capillary Automatic DNA Sequencer ABI 3730 (Applied Biosystems).

2.3. Sequence analysis

The sequences were analyzed by the program DNA STAR (Inc. Larsergene). The putative cDNA sequences encoding the subunits were deduced by comparing the cloned genomic sequences with the published cDNA sequences. The accession numbers of the cDNA sequences of *Gy1*, *Gy4* and *Gy5* are E02463, X02626 and AB049440, respectively.

2.4. Southern blotting

Genomic DNA extraction and Southern blotting were performed as described previously [21,27]. Ten micrograms of the genomic DNA was digested completely with TaqI, and separated on a 0.8% agarose gel then transferred onto Hybond N+ nylon membranes. Hybridizations were carried out for 16 h at 65 °C using the ^{32}P-dCTP-labeled fragments of *Gy1* cloned from Liuyuehuang, and the exon 3 of *Gy4* cloned from *G. latifolia* as probes, respectively. The exon 3 of *Gy4* from *G. latifolia* was prepared by PCR from the cloned *Gy4* fragment of*G. latifolia.* The forward primer was the same as that used to amplify the Group II genes. The reverse primer was 5′-CTTGTAGAGGACAACAT ATT-3′. The filters were washed with 2×, 1×, and 0.5× SSC/0.1% SDS for 15 min, 10 min, and 2 min at 65 °C, respectively.

3. Results

3.1. Comparison of the Gy1 partial genomic sequences and their deduced amino acid sequences

The full length of the *Gy1* gene containing four exons and three introns is 2388 bp in length [24]. We amplified the genomic region ranging from 48 bp (nucleotide position 1108) downstream of the starting site of the exon 3 to the stop codon (nucleotide position 2388) of the *Gy1* from cultivated accession Liuyuehuang (*G. max*), wild annual accession Y0074 (*G. soja*), and wild perennial accession PW0031 (*G. latifolia*). Three *Gy1* fragments were 1281 bp, 1281 bp and 1278 bp in length, respectively (Fig. 1). They all included the corresponding exon 3, intron 3 and exon 4 when compared with the published *Gy1* sequence of *G. max*. The similarities of the present three *Gy1* sequences with the published *Gy1* from *G. max* were 100%, 99.3% and 99.8%, respectively. And all of their putative mRNA splicing sites were consistent with the corresponding ones of *G. max*. The corresponding deduced amino acid sequences were 299 Aa, 299 Aa and 298 Aa in length, respectively, ranging from position Gln178 to the last amino acid (Ala476) of the subunit G1 from *G. max* [24]. The similarities of these amino acid sequences with the published one from *G. max* were 100%, 99.7% and 98.3%, respectively (Fig. 2). These results indicated that the *Gy1* genes and the encoded proteins are conserved among the compared accessions.

Gy1 of wild annual accession Y0074 (*G. soja*) had one nucleotide substitution (A to G) in the corresponding exon 3. This event led to the change from Asp266 in *G. max* to Glu266 in *G. soja*, but did not change the acidic characteristic of the residue. This change was located in the hypervariable region IV [6] as well as the disordered region 4 [9] of the subunit (Fig. 2). The partial genomic fragment of *Gy1* cloned from *G. latifolia* showed more variations both in the exons and in the intron, compared with those of the *G. max* and *G. soja*. The exon 3 from *G. latifolia* had a triple nucleotide deletion (GAA) which did not change the ORF of the sequence. This deletion resulted in an acidic residue Glu265 deletion in the hypervariable region IV [6] as well as in the disordered region 4 [9] of the subunit (Fig. 2). The hypervariable region IV is an acidic residue rich region which has eight Glu residues and one Asp residue in *G. max* [6]. The Glu265 deletion in this region may not influence the structure and the function of the sub-

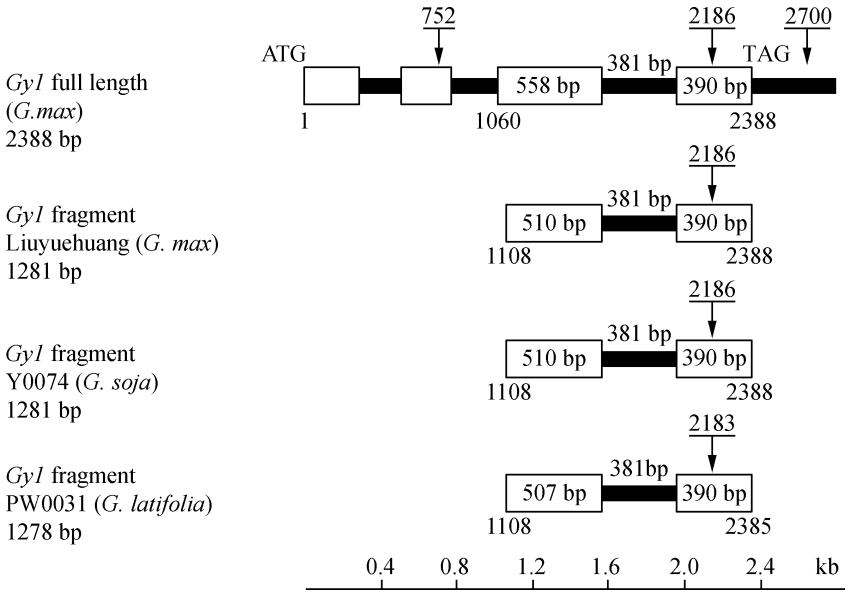

Fig. 1. Comparison of the glycinin *Gy1* genes from various accessions. The full length of the *Gy1* gene is from the other cultivar [24]. *Gy1* fragments were amplified from cultivar Liuyuehuang (*G. max*), wild annual soybean Y0074 (*G. soja*) and wild perennial PW0031 (*G. latifolia*), respectively. The open boxes indicate exons. The filled boxes indicate introns. The arrows mark TaqI cleavage sites and the numbers above the arrows indicate the positions of cleavage. The numbers under each panel indicate positions of the nucleotides.

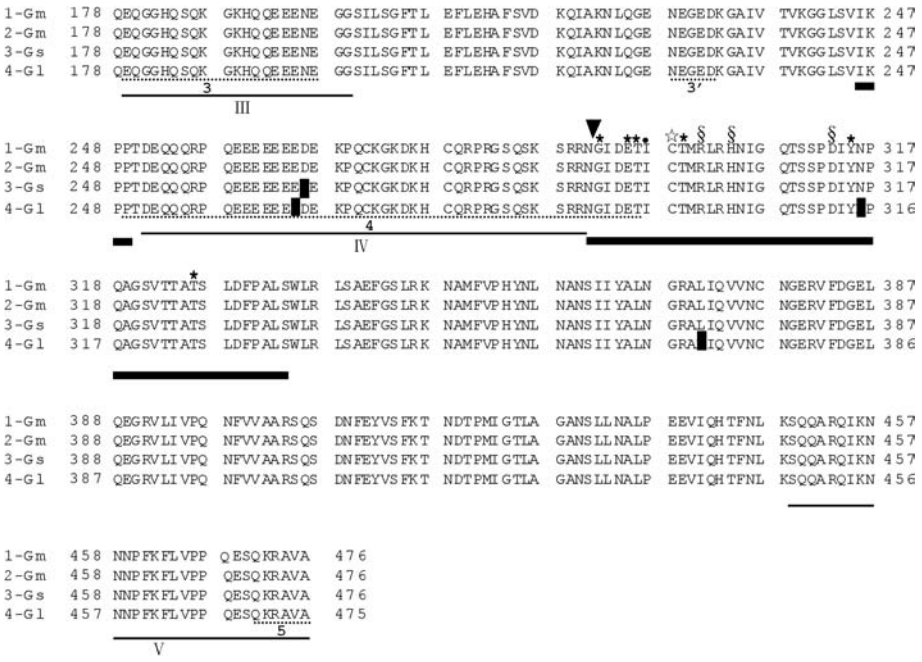

Fig. 2. Alignment of the G1 amino acid sequences from different soybean species. 1-Gm represents the published G1 sequence from *G. max* [24]. 2-Gm represents the G1 sequence from cultivar Liuyuehuang (*G. max*). 3-Gs represents the G1 sequence from the wild annual soybean Y0074 (*G. soja*). 4-Gl represents the sequence from the wild perennial accession PW0031 (*G. latifolia*). The differences are decorated by black shade. Dotted lines indicate the disordered regions 3, 3′, 4 and 5. Thin lines indicate the variable regions III, IV and V. Thick lines indicate the binding regions of the protomers. The arrow indicates the processing site between Asn and Gly. Filled circle indicates the residue Ile297 critical for vacuole sorting. Pentagram indicates Cys residue involved in interchain linkage. * and § indicate residues involved in hydrogen bonds and hydrogen-bonded salt bridges, respectively.

unit significantly. Besides this deletion, the *Gy1* sequence from *G. latifolia* had two, one and one nucleotide changed in the corresponding exon 3 (A to G, C to G), intron 3 and exon 4 (T to C), respectively. Because the first nucleotide change was a synonymous mutation, the corresponding amino acid residue did not change. The second and the fourth nucleotide changes resulted in the changes from Asn316 in *G. max* to Lys315 in *G. latifolia*, and from Leu373 in *G. max* to Ser370 in *G. latifolia*, respectively (Fig. 2).

We further examined the residues critical for the assembling of the G1 subunit from *G. latifolia*. The putative cleavage site between the Asn291 and Gly292 (Fig. 2, arrow) was the same as other glycinin subunits. It has been demonstrated that the cleavage is a trigger for the formation of the hexamers [4]. The Cys298 residue involved in interchain linkage remained conservative. The residues involved in hydrogen bonds and hydrogen-bonded salt bridges were unchanged (Fig. 2). The cloned part of G1 included the third and fourth binding regions of the protomers [10], and they remained conservative too. The known PSV sorting determinants of the G1 subunit from *G. max* [11] remained conserved in the G1 subunit from *G. latifolia*.

3.2. Comparison of *Gy4* and *Gy5* partial genomic sequences and deduced amino acid sequences

Group II subunits of glycinin consist of G4 (A5A4B3) and G5 (A3B4), and the genes encoding these subunits are named as *Gy4* and *Gy5*, respectively [5,6]. The sequence identity between them is about 80% [5]. The full length *Gy4* and *Gy5* from *G. max* are 2600 bp and 2822 bp in length, respectively, and both of them contain four exons and three introns [25,26]. A pair of primers can amplify specific bands from cultivar Liuyuehuang (*G. max*), the wild perennial accessions PW0031 (*G. latifolia*) and PW0052 (*G. tomentella*) under different PCR conditions. Sequence analysis showed that the fragment from Liuyuehuang was 1484 bp in length and shared 100% identity with the *Gy5* of cultivar Williams of *G. max* [25]. The putative splicing sites, the length of the exon 3, intron 3 and exon 4 were exactly the same as those in the *Gy5* of Williams [25], respectively (Fig. 3). However, the fragments cloned from *G. latifolia* and *G. tomentella* were 1517 bp and 1590 bp in length, respectively. Sequence alignment revealed that both of them had higher similarities with *Gy4* (86% and 85%, respectively) than with *Gy5* (72% and 71%, respectively). A sequence of the glycinin gene *Gy4* from wild perennial accession *Glycine microphylla* has been reported ([19], GenBank Accession No.: DQ415544). The *Gy4* fragments from *G. latifolia* and *G. tomentella* had about 94% similarity with the corresponding fragment of the *Gy4* from *G. microphylla*, suggesting that these two sequences are the alleles of *Gy4*, therefore we named them as *Gy4* (Fig. 3).

The similarities at the nucleotide level were compared among the *Gy4* genes from *G. max, G. latifolia*, *G. tomentella*, and *G. microphylla.* For exon 3 of *Gy4*, the similarities between *G. max* and the three wild perennial species were about 86%. The similarities among the wild perennial species were above 95%. For the intron 3 of *Gy4*, the similarities between *G. max* and the three wild perennial species were from 76% to 84%. And the similarities within the three wild species ranged from 75% to 89%. For the exon 4 of *Gy4*, the similarities were above 96% among all the four compared soybean accessions. These results indicated that the intron 3 of the *Gy4* gene had more genetic variations between the subgenera *Soja* (*G. max*) and

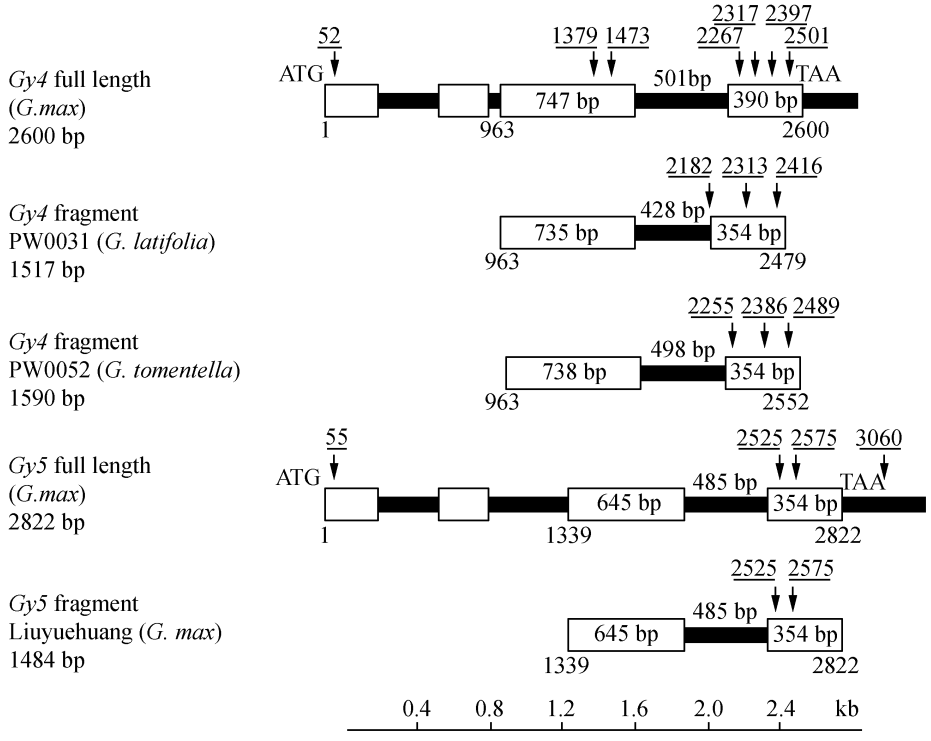

Fig. 3. Comparison of the *Gy4* and *Gy5* genes from various accessions. The full length of the *Gy4* and *Gy5* genes from other cultivars [25,26]. The *Gy5* fragment was amplified from cultivar Liuyuehuang (*G. max*) and *Gy4* fragments were amplified from wild perennial accessions PW0031 (*G. latifolia*) and PW0052 (*G. tomentella*). Other descriptions are the same as those in Fig. 1.

Glycine as well as within the subgenus *Glycine* than the exon 3 and exon 4. The exon 3 had more genetic variations between the subgenera *Soja* and *Glycine* than that within the subgenus *Glycine*. The exon 4 was highly conserved between the subgenera *Soja* and *Glycine* as well as within the subgenus *Glycine*.

The deduced amino acid sequences from the cloned *Gy5* of *G. max* and the *Gy4* of *G. latifolia* and *G. tomentella* were compared with the known sequences. Fig. 4 shows that, in the variable region III (the disordered region 3), the G4 subunits from *G. latifolia*, *G. microphylla* and *G. tomentella* have three or four more Gln residues compared with the G4 and G5 subunits from *G. max* and *G. soja*. More variations were observed in the variable region IV (the disorder region 4). In this region, 31 amino acids (KREQDQDQDE DEDEDEDQPR KSREWRSKKT Q) were present in G4 from accessions of *G. max* and *G. soja* but absent in G5 from these two species [6,25,26] (Fig. 4). In G4 of the wild perennial species *G. latifolia*, *G. tomentella* and *G. microphylla*, the inserted amino acid sequences were seven residues shorter than those from G4 of *G. max* and *G. soja*, and the corresponding amino acid compositions were also different (Fig. 4).

Except for the variable regions as well as the disordered regions of the G4 subunits, other regions were not significantly changed (Fig. 4). The Asn354 and Gly355 residues for the cleavage to separate the acidic and basic polypeptides [4] were conserved. The Cys361 residue involved in

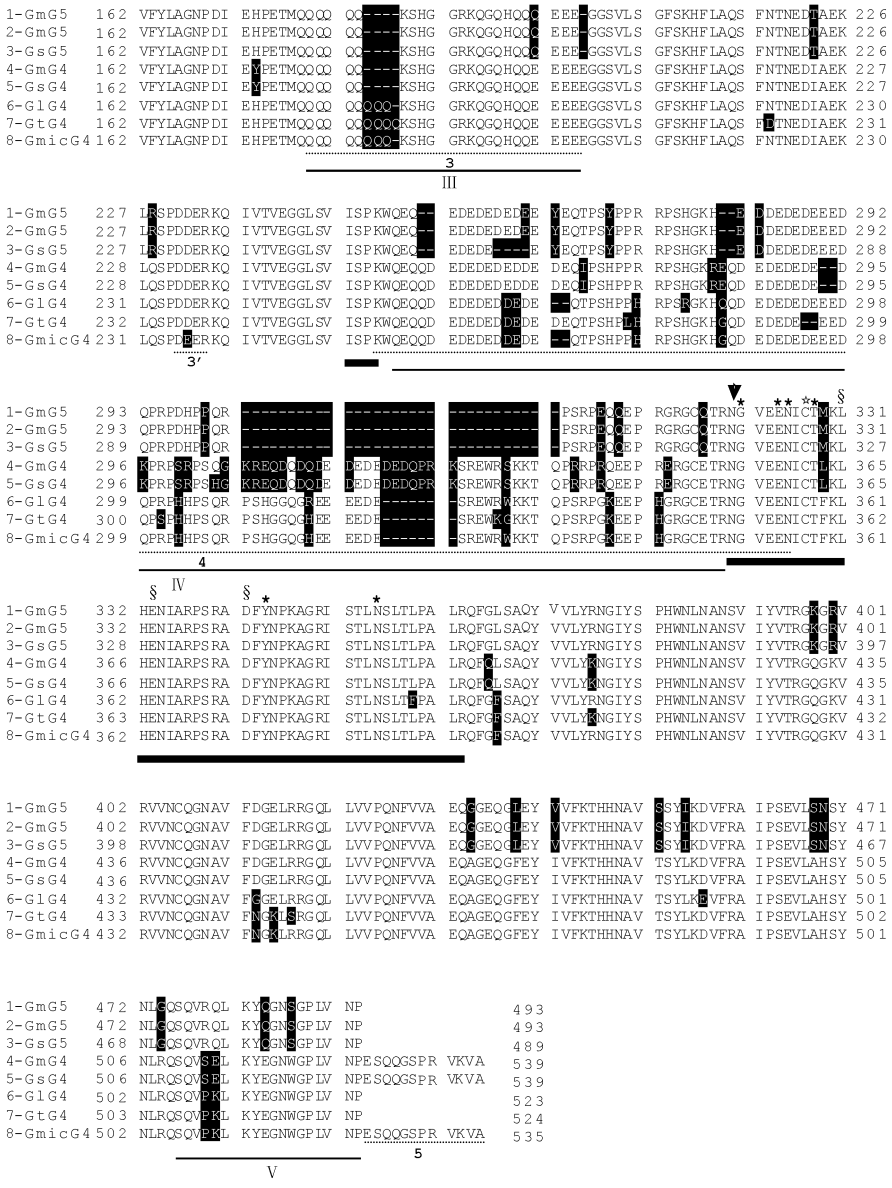

Fig. 4. Alignment of the G4 and G5 amino acid sequences from different soybean species. 1-GmG5 represents a published G5 sequence from *G. max* [25]. 2-GmG5 represents the G5 sequence from cultivar Liuyuehuang (*G. max*); 3-GsG5 represents the G5 from *G. soja* [18]; 4-GmG4 represents a published G4 sequence from *G. max* [26]; 5-GsG4 represents the G4 sequence from *G. soja* [18]; 6-GlG4 represents the G4 sequence from PW0031 of *G. latifolia;* 7-GtG4 represents the G4 sequence from PW0052 (*G. tomentella*); 8-GmicG4 represents the G4 subunit from *G. microphylla* (DQ415544). Other descriptions are the same as those in Fig. 2.

interchain linkage [3] and the residues involved in hydrogen bonds and hydrogen-bonded salt bridges [10] were also unchanged. The third and fourth binding regions of the protomers of G5 subunit from *G. max* [10] remained relatively conservative in the G4 subunit fragments from *G. latifolia*, *G. tomentella* and *G. microphylla* (Fig. 4).

3.3. Genetic diversity of the glycinin subunit genes among accessions from subgenera Soja and Glycine

To understand the genetic variations of the glycinin genes between cultivated and wild soybean species, we analyzed the restriction fragment length polymorphisms of glycinin subunit *Gy1* and *Gy4* genes among 19 accessions of subgenera *Soja* and *Glycine* by Southern blot analysis. Genomic DNA from 19 accessions of 5 species was digested completely with TaqI. The labeled partial genomic sequence of *Gy1* cloned from the cultivar Liuyuehuang (*G. max*) was used as probe. The results in Fig. 5 show that four apparent hybridized bands (1.4 kb, 1.0 kb, 0.7 kb, 0.65 kb) can be observed in all the accessions from *G. max* and *G. soja* of subgenus *Soja*. A smaller and weaker band of 0.5 kb can also be notified. The 1.4 kb and 0.5 kb fragments most likely correspond to the two fragments from the TaqI digestion of the *Gy1* genomic sequence (Fig. 1). Other bands may result from cross-hybridizations of the probe with the *Gy2* and *Gy3* due to the high similarity of these genes [24,25]. The two fragments (1.0 kb and 0.65 kb) probably correlated with the *Gy2* gene and the 0.7 kb fragment likely correlates with the *Gy3* gene. The identical hybridization pattern in all the accessions of the *G. max* and *G. soja* supports that the Group I glycinin subunits have high degree of conservation among most accessions of subgenus *Soja* [19].

The hybridization patterns were different in the subgenus *Glycine* (Fig. 5). One to three apparent hybridization bands were observed in different accessions. This result indicates that although the cloned partial sequence of *Gy1* from PW0031 had high similarity with those of *G. max* and *G. soja*, there were variations that existed in the *Gy1* from PW0031 (*G. latifolia*) or other species in subgenus *Glycine*. Three similar bands were observed in PW0063 and PW0060. Two identical bands were identified in PW0021 and PW0005 of *G. tabacina*, implying a close relationship of the two accessions.

The genetic variations of Group II glycinin subunit genes were also examined among the 19 accessions using the labeled exon 3 fragment of the *Gy4* from PW0031 (*G. latifolia*) as the probe. Fig. 6 shows that three apparent hybridized bands (2.5 kb, 1.2 kb, 0.8 kb) can be identified in all the accessions from *G. max* and *G. soja*. Through the analysis of the probe used for hybridization and the TaqI restriction sites in the *Gy4* and *Gy5* (Fig. 3), we found that the 2.5 kb band corresponded to the fragment from *Gy5*, and the 1.2 kb and 0.8 kb bands corresponded to the fragments from the *Gy4* gene. In the accessions from the subgenus *Glycine*, the hybridization pattern was more varied. A band of ~1.2 kb in PW0031 (*G. latifolia*) and PW0052 (*G. tomentella*) was observed, consistent with the prediction from the TaqI site analysis of the *Gy4* gene. Three accessions PW0052, PW0060 and PW0063 from *G. tomentella* had similar band patterns at ~1.2 kb and ~0.8 kb (Fig. 6). However, their large bands were different, indicating the presence of a polymorphism in this locus. PW0063 had identical band pattern with PW0060 (*G. tomentella*) when *Gy1* fragment was used as the probe (Fig. 5). However, when *Gy4* fragment was used as a probe, PW0063 and PW0060 showed an obvious polymorphism

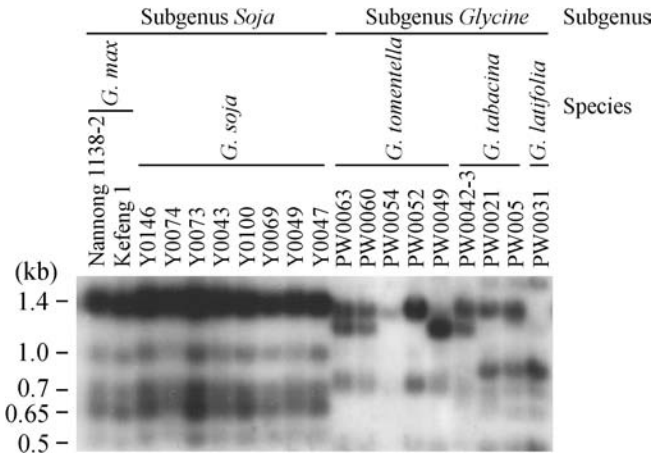

Fig. 5. Genetic variations of Group I glycinin subunit genes among accessions from subgenera *Soja* and *Glycine*. TaqI enzyme was used to digest the genomic DNA. Hybridization was performed using the labeled *Gy1* from Liuyuehuang (*G. max*) as the probe. Due to the high similarity of the Group I glycinin subunit genes, the *Gy2* and *Gy3* gene fragments were also cross-hybridized. The sizes and the corresponding gene fragments are indicated on the left.

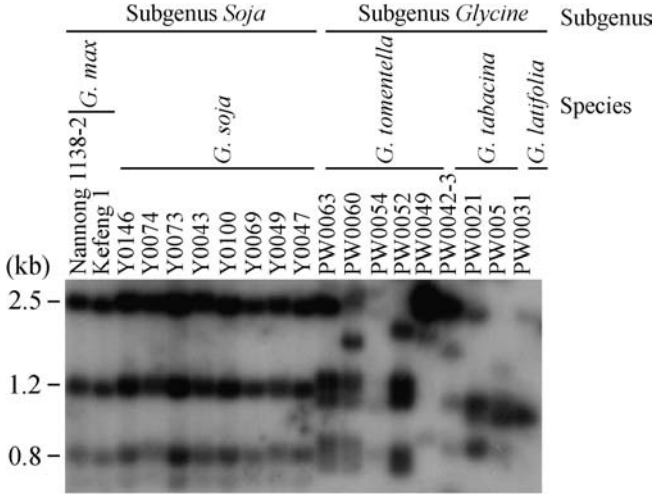

Fig. 6. Genetic variations of Group II glycinin subunit genes among accessions from subgenera *Soja* and *Glycine*. TaqI enzyme was used to digest the genomic DNA. Hybridization was performed using the labeled exon 3 of *Gy4* fragment from PW0031 (*G. latifolia*) as the probe. Due to the high similarity of Group II glycinin subunit genes, the *Gy5* gene fragments were also cross-hybridized. The sizes and the corresponding gene fragments are indicated on the left.

in size of the large bands. These results indicated that the two different groups of glycinin subunit genes may have experienced different evolutionary process in PW0060 and PW0063 of *G. tomentella.*

4. Discussion

We cloned *Gy1* partial genomic sequences from wild perennial accession PW0031 (*G. latifolia*) of subgenus *Glycine*, a wild annual accession Y0074 of *G. soja* and a cultivar of *G. max*, and compared them with the published counterparts from *G. max* [24]. These *Gy1* sequences and the deduced amino acid sequences were conservative, indicating that *Gy1* genes were very conservative during the evolution. The vacuolar sorting determinants [11], the cleavage site between the Asn291 and Gly292 [4], and some residues involved in hydrogen bonds and hydrogen-bonded salt bridges [10] of G1 subunit from *G. max* remained conserved in G1 subunit from *G. latifolia*, indicating that the assembling and sorting mechanisms of G1 subunit were probably conservative during the evolution.

Using a pair of primers designed according to the *Gy5* from *G. max* [25], we cloned *Gy4* fragments from two wild perennial accessions PW0031 (*G. latifolia*) and PW0052 (*G. tomentella*). In a previous study, Staswick et al. have found that the A3 component (acid polypeptide of G5 subunit) of *G. max* was missing in all the studied perennials [20]. In the present study, the *Gy5* gene encoding the G5 subunit (A3B4) appeared to be absent in the PW0031 (*G. latifolia*) and PW0052 (*G. tomentella*) genomes. These studies suggested that the *Gy5* gene may exist in cultivated soybean but not in the wild perennial species. A sequence of the *Gy4* gene from wild perennial accession *G. microphylla* has also been reported ([19], GenBank Accession No.: DQ415544), and this *Gy4* had high similarity with the present *Gy4* genes from *G. latifolia* and *G. tomentella*. The G4 sequences from the three wild perennial accessions of subgenus *Glycine* were seven amino acids shorter than the G4 from *G. max* or *G. soja* in the variable region IV (the disordered region 4). These results may imply that there exists a new allele of the *Gy4* gene, featuring the wild perennial accessions. The intron 3 of the *Gy4* genes had abundant variations between the subgenera *Soja* and *Glycine* as well as within the subgenus *Glycine*.

The genetic variations of the *Glycinin* genes were also compared among accessions from subgenera *Soja* and *Glycine*. We found that the hybridization patterns of the *Glycinin* genes were identical among the accessions from the subgenus *Soja*. However, the hybridization patterns were more diverged among the accessions from the subgenus *Glycine*. These results suggested that the *Glycinin* genes are highly conserved within the subgenus *Soja* but more varified within the subgenus *Glycine*.

Overall, we cloned the partial genomic sequences of the *Gy1* and *Gy4* genes from the cultivated and wild soybean species, and found that *Gy1* gene was highly conserved whereas the *Gy4* gene was more diverged among the compared species at both the nucleotide level and the amino acid level. In addition, Group I and Group II genes of the glycinin subunits were highly conserved among the cultivated accessions within the subgenus *Soja*. On the contrary, the genetic variations of the glycinin subunit genes were drastic among the wild perennial soybean accessions. This study may have significance in elucidating the evolution of the soybean species.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant No. 30490254), 863 Project 2006AA10Z1F4 and National Program on Key Basic Research Project (2002CB111303). The authors thank Prof. Gai Junyi for providing the seeds of Kefeng and Nannong1138-2 (*G. max*).

References

[1] Tumer NE, Thanh VH, Nielsen NC. Purification and characterization of mRNA from soybean seeds. J Biol Chem 1981;256(16):8756–60.
[2] Tumer NE, Richiter JD, Nielsen NC. Structural characterization of the glycinin precursors. J Biol Chem 1982;257(8):4016–8.
[3] Staswick PE, Hermodson MA, Nielsen NC. Identification of the cystines which link the acidic and basic components of the glycinin subunits. J Biol Chem 1984;259(21):13424–30.
[4] Dickinson CD, Hussein EHA, Nielsen NC. Role of posttranslational cleavage in Glycinin assembly. Plant Cell 1989;1(4):459–69.
[5] Scallon BJ, Thanh VH, Floener LA, et al. Identification and characterization of DNA clones encoding group-II glycinin subunits. Theor Appl Genet 1985;70(5):510–9.
[6] Nielsen NC, Dickinson CD, Cho TJ, et al. Characterization of the glycinin gene family in soybean. Plant Cell 1989;1(3):313–28.
[7] Wright DJ. The seed glubulins. In: Developments in food proteins. London: Elsevier; 1987. p. 81–157.
[8] Wright DJ. The seed glubulins. In: Developments in food proteins. London: Elsevier; 1988. p. 119–78.
[9] Adachi M, Takenaka Y, Gidamis AB, et al. Crystal structure of soybean proglycinin A1aB1b homotrimer. J Mol Biol 2001;305(2):291–305.
[10] Adachi M, Kanamori J, Masuda T, et al. Crystal structure of soybean 11S globulin: glycinin A3B4 homohexamer. Proc Natl Acad Sci USA 2003;100(12):7395–400.
[11] Maruyama N, Mun LC, Tatsuhara M, et al. Multiple vacuolar sorting determinants exist in soybean 11S globulin. Plant Cell 2006;18(5):1253–73.
[12] Doyle JJ, Doyle JL, Brown AHD. Origins, colonization, and lineage recombination in a widespread perennial soybean polyploid complex. Proc Natl Acad Sci USA 1999;96(19):10741–5.
[13] Doyle JJ, Doyle JL, Brown AHD, et al. Genomes, multiple origins, and lineage recombination in the *Glycine tomentella* (Leguminosae) polyploid complex: histone H3-D gene sequences. Evolution Int J Org Evolution 2002;56(7):1388–402.
[14] Doyle JJ, Doyle JL, Brown AHD, et al. Multiple origins of polyploids in the *Glycine tabacina* complex inferred from chloroplast DNA polymorphism. Proc Natl Acad Sci USA 1990;87(2):714–7.
[15] Rauscher JT, Doyle JJ, Brown AHD. Multiple origins and nrDNA internal transcribed spacer homologue evolution in the *Glycine tomentella* (Leguminosae) allopolyploid complex. Genetics 2004;166(2):987–98.
[16] Fukazawa C, Momma T, Hirano H, et al. Glycinin A3B4 mRNA. Cloning and sequencing of double-stranded cDNA complementary to a soybean storage protein. J Biol Chem 1985;260(10):6234–9.

[17] Weng WM, Gao XS, Zhuang NL, et al. The glycinin A3B4 mRNA from wild soybean *Glycine soja* Sieb. et ZUCC. Plant Physiol 1995;107(2):665–6.
[18] Fukuda T, Maruyama N, Kanazawa A, et al. Molecular analysis and physicochemical properties of electrophoretic variants of wild soybean *Glycine soja* storage proteins. J Agric Food Chem 2005;3(9):3658–65.
[19] Mahmoud AA, Natarajan SS, Bennett JO, et al. Effect of six decades of selective breeding on soybean protein composition and quality: a biochemical and molecular analysis. J Agric Food Chem 2006;54(11):3916–22.
[20] Staswick PE, Broue P, Nielsen NC. Glycinin composition of several perennial species related to soybean. Plant Physiol 1983;72(4):1114–8.
[21] Chen SY, Zhu LH, Hong J, et al. Molecular biological identification of a rice salt tolerant line. Acta Bot Sin 1991;33:569–73.
[22] Thanh VH, Tumer NE, Nielsen NC. The glycinin *Gy2* gene from soybean. Nucleic Acids Res 1989;17(11):4387.
[23] Cho TJ, Nielsen NC. The glycinin Gy3 gene from soybean. Nucleic Acids Res 1989;17(11):4388.
[24] Sims TL, Goldberg RB. The glycinin Gy1 gene from soybean. Nuleic Acid Res 1989;17(11):4386.
[25] Chen SF, Fukazawa C. Molecular cloning and nucleotide sequence of soybean glycinin gene Gy5 (A3B4). Chin J Biotechnol 2000;16(2):215–7, [in Chinese].
[26] Xue ZT, Xu ML, Shen W, et al. Characterization of a Gy4 glycinin gene from soybean *Glycine max* cv. forrest. Plant Mol Biol 1992;18(5):897–908.
[27] Wu XL, He CY, Chen SY, et al. Phylogenetic analysis of the interspecies in the genus Glycine through SSR markers. Acta Genet Sin 2001;28(4):359–66.