Theor Appl Genet (2009) 119:815–825
DOI 10.1007/s00122-009-1091-5

ORIGINAL PAPER

Comparison of the starch synthesis genes between maize and rice: copies, chromosome location and expression divergence

Hong-Bo Yan · Xiao-Xue Pan · Hua-Wu Jiang · Guo-Jiang Wu

Received: 16 December 2008 / Accepted: 8 June 2009 / Published online: 11 July 2009

Abstract Gene duplication and divergence are important evolutionary processes. It has been suggested that a whole genome duplication (WGD) event occurred in the Gramineae, predating its divergence, and a second WGD occurred in maize during its evolution. In this study we compared the fate of the genes involved in the core pathway of starch biosynthesis following the ancient and second WGDs in maize and rice. In total, thirty starch synthesis genes were detected in the maize genome, which covered all the starch synthesis gene families encoded by 27 genes in rice. All of these genes, except *ZmGBSSIIb* and *ZmBEIII*, are anchored within large-scale synteny blocks of rice and maize chromosomes. Previous findings and our results indicate that two of the current copies of many starch synthesis genes (including *AGPL*, *AGPS*, *GBSS*, *SSII*, *SSIII*, and *BEII*) probably arose from the ancient WGD in the Gramineae and are still present in the maize and rice genome. Furthermore, two copies of at least six genes (*AGPS1*, *SSIIb*, *SSIIIb*, *GBSSII*, *BEI*, and *ISA3*) appear to have been retained in the maize genome after its second WGD, although complete coding regions were only detected among the duplicate sets of *AGPS1*, *SSIIb*, and *SSIIIb*. The expression patterns of the remaining duplicate sets of starch synthesis genes (AGPL1/2, AGPS1/2, SSIIa/b, SSIIIa/b, GBSSI/II, and BEIIa/b) differ in their expression and could be classified into two groups in maize. The first group is mainly expressed in the endosperm, whereas the second is expressed in other organs and the early endosperm development. The four duplicate sets of *ZmGBSSII*, *ZmSSIIb*, *ZmSSIIIb* and *AGPS1*, which arose from the second WGD diverged in gene structure and/or expression patterns in maize. These results indicated that some duplicated starch synthesis genes were remained, whereas others diverged in gene structure and/or expression pattern in maize. For most of the duplicated genes, one of the copies has disappeared in the maize genome after the WGD and the subsequent "diploidization".

Communicated by J. Dubcovsky.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1091-5) contains supplementary material, which is available to authorized users.

H.-B. Yan · X.-X. Pan · H.-W. Jiang (✉) · G.-J. Wu (✉)
Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden,
Chinese Academy of Sciences, Xingke Road 723, Tianhe District, 510650 Guangzhou, People's Republic of China
e-mail: hwjiang@scbg.ac.cn

G.-J. Wu
e-mail: wugj@scbg.ac.cn

H.-B. Yan
Bioscience and bioengineering School,
Hebei University of Economics and Business,
050061 Shi Jiazhuang, People's Republic of China

H.-B. Yan · X.-X. Pan
Graduate University of the Chinese Academy of Sciences,
100049 Beijing, People's Republic of China

Introduction

Rice (*Oryza sativa* L.) and maize (*Zea mays* L.) are two of the world's major Graminaceous agricultural crops and form an important part of the diet of humans and livestock. Integration of structural genomic data from a largely assembled genome sequence suggests that a whole genome duplication (WGD) event occurred approximately 70 million years ago (Mya) prior to the divergence of the Gramineae (Paterson et al. 2004). In addition, during their evolution a second WGD occurred in maize approximately 12 Mya (Gaut and

Doebley 1997). Subsequently, maize 'diploidized' by deleting most of the duplicated centromere regions and deleting or tolerating the degeneration of one of most of its paired genes (Song and Messing 2003; Brunner et al. 2005).

Starch, the main carbohydrate storage material in plants, is composed of two distinct types of glucose (Glc) polymers: amylose and amylopectin. The functions of starch depend on the plant tissue in which it is formed. Transient starch is synthesized in leaves or during the early developmental stages of seeds, and is generally rapidly turned over, thereby playing a key role in meeting the energy requirements of plant development. Conversely, storage starch, which accumulates in starchy seeds or tubers, provides a long-term carbon store for the next generation. Starch is also an important contributing factor to the grain yield and quality of maize and rice. In cereals, the core pathway of starch biosynthesis contains ADP-Glc pyrophosphorylase (EC 2.7.7.27, AGPase), starch synthase (EC 2.4.1.21, SS), starch branching enzyme (EC2.4.1.18, BE) and debranching enzyme (EC 2.4.1.41, DBE), ADP-Glc transporter (*BT1*) and starch phosphorylase (PHO, EC 2.4.1.1) (James et al. 2008; Satoh et al. 2008). Plants contain multiple isoforms of these enzymes, most of which are encoded by multi-genes in cereals. Two and four genes encoding AGPS (AGPase small subunit) and AGPL (AGPase large subunit), respectively, ten encoding starch synthase, four encoding BE and four encoding DBE have been detected in the rice genome to date (Dian et al. 2005; Ohdan et al. 2005; Han et al. 2007; Lee et al. 2007; Rösti and Denyer 2007; Wu et al. 2008). The multicopy genes have diverged into two groups in rice: Group I genes are preferentially expressed in the endosperm, while Group II genes are mainly expressed in the vegetative tissues and in the early development of the endosperm (Dian et al. 2005; Ohdan et al. 2005).

The aim of this study was to analyze the fate of the genes involved in the core pathway of starch biosynthesis following the ancient and second WGDs in maize. Nineteen of these genes have been previously characterized in maize (Gao et al. 1997, 1998; Harn et al. 1998; Rahman et al. 1998; Beatty et al. 1999; Knight et al. 1998; Satoh et al. 2003; Rösti and Denyer 2007). In our study, we detected 11 additional genes from maize genomic sequence databases and analyzed their expression patterns. We then compared the chromosome loci, evolution and functions of these genes in rice and maize.

Materials and methods

Plant material

Maize (*Zea mays* L.) seeds of a self-cross line (nanzhi11) were planted at the early season in a field trial in the South China Botanical Garden, Guangzhou, People's Republic of China. Leaves and roots were taken from seedlings at the fourth-leaf stage. Grains (in the middle part of ears) were harvested several times after silking, from 25 June to 20 July, when the ambient temperature was 32–36°C/24–27°C (day/night). The samples were frozen in liquid nitrogen and stored at −80°C until required. All samples were collected between 9:00 am and 10:00 am.

Sequence retrieval and coding region determination

The genes encoding the proteins involved in the core pathway of starch biosynthesis, e.g. ADP-Glc pyrophosphorylase (AGPase), starch synthase (SS), starch branching enzyme (BE), debranching enzyme (DBE), ADP-Glc transporter (BT1) and starch phosphorylase (PHO), were selected for our study. We retrieved maize genomic sequences of putative starch synthesis genes from the National Center for Biotechnology Information database (NCBI, http://www.ncbi.nlm.nih.gov/) and the Maize Genetics and Genomics Database (MaizeGBD, http://www.maizegdb.org) by BLASTN searches using the cDNA sequences of rice starch synthesis genes. The expressed sequence tags (ESTs) of these genes were retrieved from Compbio (http://compbio.dfci.harvard.edu/), NCBI, and MaizeGBD by BLASTN searches using the obtained genomic sequences. The putative complete coding domain sequences (CDS) of the genes were then determined by alignment of genomic DNA sequences with their corresponding ESTs. The genomic sequences of two putative genes, *SSIIIb-2* and *PHOH*, were located in two separate contigs within their respective bacterial artificial chromosome (BAC) sequences. The two separate contigs encoding *PHOH* can be linked by an EST of DY400651, while the linkage of the two separate contigs encoding *SSIIIb-2* was determined directly by RT-PCR and sequence analysis using the primers: 5′-GAGGAGGAGCGGAAGCG AGG-3′ (forward primer) and 5′-GTCCGGCCACTT CCTTCGAG-3′ (reverse primer). *ZmGBSSIIb* lacks a large fragment in the 5′- region, and the genomic sequences of this region in the BAC was confirmed by genomic PCR and sequence analysis using the primers: 5′-CTGCCATCCC AATCCCTAC-3′ (forward primer) and 5′-TTTAGTCGG CTTACAGTCAAC-3′ (reverse primer).

Phylogenetic analysis

The conserved domains amino acid sequences were aligned by the ClustalW program at the EBI web online server for subsequent phylogenetic analysis (http://www.ebi.ac.uk/cgi-bin/clustalw/) (Thompson et al. 1994). Phylogenetic analyses were conducted with the alignment data (Supplementary Material online) using MEGA version 4 (Tamura

et al. 2007). Bootstrap values were calculated from 1,000 replicate analyses.

Gene loci comparison

The gene loci on rice chromosome pseudomolecules were obtained by BLASTN searches using corresponding cDNA sequences from the NCBI database (http://www.ncbi.nlm.nih.gov/blast). The location of the maize genes on the chromosome pseudomolecules was traced though the anchored corresponding BAC genome sequences or related markers (http://www.maizegdb.org). Finally, the loci of anchored rice and maize starch synthesis genes were compared on the rice/maize synteny physical maps available at Gramene (http://www.gramene.org/).

Semi-quantitative RT-PCR analysis

Total RNA was extracted using Trizol reagent according to the manufacturer's instructions (Invitrogen, http://www.invitrogen.com). First-strand cDNA was synthesized from 3 μg total RNA from each sample using M-MLV reverse transcriptase (Promega, http://www.promega.com). The first-strand cDNA was used as a template for semi-quantitative PCR analysis after normalization using maize *Actin* (accession no. EU970017). Appropriate amounts of template cDNA and numbers of PCR cycles were determined for each gene to ensure that amplification occurred in the linear range and allowed good quantification of the amplified products. The primers and the PCR programs used are listed in Table 1. To ensure the primers were specific for each copy of *GBSSIIb-1/2*, *SSIIb-1/2*, and *SSIIIb-1/2*, all of the reverse primes were designed in the divergent 3′ non-coding region. The amplification products obtained with the different primer pairs were purified and sequenced, respectively, to confirm the specificity of each primer pair. The PCR products were analyzed on 1% agarose gels, stained with ethidium bromide and visualized using the Fluorescence Chemiluminescence and Visible Imaging System.

Results

Starch synthesis genes detected from the maize genome database

After the BLASTN search of the maize genome database, we detected 11 previously uncharacterized putative starch synthesis genes containing open reading frames. Two genes encode the ADP-Glc pyrophosphorylase large subunits of *ZmAGPL3* and *ZmAGPL4*. Six genes encode starch synthase: *ZmSSIIb-2*, *ZmSSIIIb-1*, *ZmSSIIIb-2*, *ZmGBSSIIa*, *ZmGBSSIIb* and *ZmSSIV*. One gene encodes the starch branching enzyme, *ZmBEIII*. Two genes encode starch phosphorylase, *ZmPHOL* and *ZmPHOH*. The information is listed in Tables 2 and 3. The gene encoding *ZmBEIII* is intronless. The detected *ZmGBSSIIb* transcripts encode a shorter protein of 275 amino acids. The other genes have similar genomic structures and polypeptide chain lengths to their orthologs in rice (Table 2). In addition, another *ISA3*-Like (*ISA3-L*) gene was detected in BAC AC194926, and a *BEI*-Like (*BEI-L*) gene in AC194216 (Table 3), but their open reading frames were not detected, and no correct sequences for their putative transcripts were obtained by RT-PCR or sequence analysis (data not shown).

Phylogenetic analysis of starch synthesis genes

We conducted a phylogenetic comparison among *Arabidopsis*, rice and maize for these starch synthesis genes (Fig. 1). Based on the phylogenetic tree of AGPase, AGPase proteins were divided into two clades: AGPL and AGPS. Each of the four AGPL proteins in maize clusters with a corresponding protein in rice. According to the tree (Fig. 1a) and the recently reported phylogenetic tree of plant AGPases (Georgelis et al. 2008), the *AGPL1*/*AGPL2* in maize and rice probably arose from an ancient duplication in the Gramineae. AGPS proteins were divided into two branches (Fig. 1a). No members of the AtAGPS2 branch were detected in either maize or rice. OsAGPS1 and ZmAGPS1a/ZmAGPS1b were grouped into the same subfamily and OsAGPS2/ZmAGPS2 into another subfamily. These findings suggest that *AGPS1*/*AGPS2* in maize and rice originated from the ancient duplication in the Gramineae, while ZmAGPS1a/ZmAGPS1b is likely to have arisen from the second WGD of maize.

Starch synthase proteins were divided into five branches: GBSS and SSI–SSIV (Fig. 1b). GBSS proteins were classified into two families, GBSSI and GBSSII, which appear to be the ancient duplicate sets in Gramineae. Two copies of GBSSII likely originated during the second WGD, although the GBSSIIb protein lacks the N-terminus in maize (Fig. 1b; Supplementary Fig. 1). SSI is a single copy gene in both rice and maize. SSII proteins were divided into three groups: SSIIa, SSIIb, and SSIIc (Fig. 1b; Yan et al. 2009). In maize and rice, SSIIa/SSIIb may have arisen from the ancient duplicate sets in the Gramineae, while ZmSSIIb-1 and ZmSSIIb-2 probably arose from the second duplication event in maize. SSIII proteins of maize and rice were classified into two groups: SSIIIa and SSIIIb. These groups are likely to have arisen from the ancient duplicate sets in the Gramineae, while ZmSSIIIb-1 and ZmSSIIIb-2 probably arose from the second duplication event in maize (Fig. 1b; Supplementary Fig. 2). SSIV proteins of maize

Table 1 Primers used for RT-PCR

Primer name	Sequence (5′→3′)	Annealing temperature (°C)	Primer position	Product length (bp)
ActF	ACCCAAAGGCTAACCGTGAG	58	439	337
ActR	TAGTCCAGGGCAATGTAGGC		757	
AGPL2F	TGATGATGGGTGCGGATTTG	53	1,489	382
AGPL2R	TCTTGTCCCTCCGACGGCTC		1,851	
AGPL3F	CCAGTCTCTCAATCGCCACA	57	460	857
AGPL3R	GCTGCACTCCCGGTTCTAAC		1,297	
AGPL4F	TGTTCAGAGTCATCGGCAAAG	56	682	635
AGPL4R	CGCCAAGCATTACAGTATCCT		1,296	
GBSSIF	GCCTGTCGCTGGAACGGACT	56	534	223
GBSSI R	CTTTGCGTCCCTGTAGATGC		737	
GBSSIIaF	CTGCAGCCAGCGCGAGGTAC	56	1	1,905
GBSSIIaR	CACCAAGCCCTCACGGAGTG		1,886	
GBSSIIbF	CGGCAGTAGCGTCAACCATC	56	792	524
GBSSIIbR	TGCTTCGGAGTTCGGACCAC		1,140	
SSIIb-2F	CGTGCCACTAATCGGGTTCAT	56	1,624	587
SSIIb-2R	GAAGGGCGTCGGCGTAACTCA		2,190	
SSIIIb-1F	AGTGGTTCAACTCCCTTAGC	56	3,579	308
SSIIIb-1R	CAGTAACGTACAACGGGTGA		3,869	
SSIIIb-2F	CTCGTTTTGCCCAGGTTAGA	56	3,894	221
SSIIIb-2R	CGGGATACGTCGTACTGTTG		4,095	
SSIVF	ACATCCACTTCTGCCCAATCTG	57	796	665
SSIVR	AGACCACCAACCTTTGCGAC		1,441	
BEIIIF	AAGTCTCCCATCAGTCGTATC	53	2,567	274
BEIIIR	TCTACGCAGTTGCATGATTCC		2,840	
ISA2F	AAACGAGGAGTCAGCGAGTG	56	2,336	410
ISA2R	ACATCTCACGGAGTTTAGGT		2,726	
ISA3R	TCAAGTTTCGGCATAACCATC	56	2,135	319
ISA3F	ATGGAAGAGTAAGGAGCAATC		2,433	
PHOLF	AGATGTGGCAGCTACCGTGAA	53	1,959	374
PHOLR	CAAAGACACCACTGCGGACAA		2,312	
PHOHF	CTGGGCTGAGGAAGGATAGA	56	2,189	314
PHOHR	CTTGGCATACTGGGCGATGG		2,483	

and rice were classified into two groups: SSIVa and SSIVb. ZmSSIV fell into the SSIVb group (Fig. 1b; Supplementary Fig. 3). We did not detect SSIVa in maize.

The four BE proteins in maize corresponded to the four proteins in rice. As previously reported by Han et al. (2007), the BE proteins are divided into three clades: BEI, BEII and BEIII (Fig. 1c). BEIIa and BEIIb proteins in maize and rice probably arose from ancient duplication in the Gramineae. The four DBE proteins can be divided into two clades, ISA and PUL, and the ISA clade can be further divided into three branches: ISA1, ISA2 and ISA3 (Fig. 1d). Like rice and *Arabidopsis*, maize also had four distinctive DBE isoforms. The two PHO proteins can be divided into two clades: PHOL and PHOH (Fig. 1e). The two PHO proteins in maize corresponded to the two proteins in rice and *Arabidopsis*.

Comparative chromosome location

Based on the rice/maize synteny (http://www.gramene.org/), the loci of the starch synthesis genes on the physical maps were compared between rice and maize. The 27 starch synthesis genes that we detected in rice were distributed on 10 of its 12 chromosomes, the 30 genes identified in maize were distributed among its 10 chromosomes (Table 3; Fig. 2). All of these genes, except *ZmGBSSIIb* and *ZmBEIII*, are located within or near large-scale synteny blocks of rice and maize chromosomes, respectively (Fig. 2).

Expression analysis

Expression patterns of 16 starch synthesis genes have been previously reported in maize (Fisher et al. 1995;

Table 2 Comparisons of the detected genes in maize with their orthologs in rice

Rice				Maize					
Gene name	Acc no.[a]	Exon number	Amino acids[b]	Gene name	Exon number	Amino acids[b]	Identities[c] (%)	BAC	Position in BAC (bp)
AGPL3	NM_001065811	15	509	*AGPL3*	15	514	88	AC208635	118,858–129,611
AGPL4	NM_001057719	16	511	*AGPL4*	16	505	91	AC211200	97,700–102,080
GBSSII	AY069940	14	608	*GBSSIIa*	14	609	84	AC210377	35,440–41,800
				GBSSIIb	7	275	87[d]	AC177838	148,332–150,431
							95[d, e]		
SSIIb	AF395537	8	694	*SSIIb-2*	8	704	78	AC190571	98,800–103,600
				SSIIb-1[f]	8	698	87[e]	(AF019297)	
SSIIIb	AF432915	16	1,216	*SSIIIb-1*	16	1191	77	AC183945	17,600–26,600
				SSIIIb-2	16	1188	76	AC195583	171,800–175,100, 212,100–215,800
							91[e]		
SSIVb	AY373258	16	915	*SSIV*	16	909	74	AC197339	85,605–94,270
PHOH	NM_001051358	15	841	*PHOH*	15	838	92	AC207564	121,111–126,810, 144,573–178,839
PHOL	AF327055	15	978	*PHOL*	15	984	89	AC191342	1,074,877–116,039
BEIII	NM_001064146	22	903	*BEIII*	1	899	81	AC204884	165,581–162,642

[a] Acc no., GenBank accession number in NCBI

[b] Amino acids, the number of amino acid in the sequence deduced from its open reading frame

[c] Identities of the amino acids between rice and maize proteins

[d] Identities of the amino acids for the 275 amino acids

[e] Identities of the amino acids between the maize gene duplicates

[f] *SSIIb-1* has previously been characterized in maize and AF019297 is its GenBank Acc no. in NCBI (Harn et al. 1998). Its genomic sequences come from two GSS: ZmGSStuc11-12-04.9391.1 and ZmGSStuc11-12-04.9391.2 (MaizeGDB)

Gao et al. 1997, 1998; Harn et al. 1998; Rahman et al. 1998; Beatty et al. 1999; Rösti and Denyer 2007). We determined the expression patterns for an additional 14 starch synthesis genes by semi-quantitative RT-PCR analysis. For comparison, the expression pattern of *ZmGBSSI* was also analyzed (Fig. 3). The transcripts of *ZmAGPL2* were strongly expressed in leaves and early developmental stage endosperm, but expression decreased with endosperm development. *ZmAGPL3* was mainly expressed in the endosperm in the middle and late developing period. *ZmAGPL4* was strongly expressed in leaves, and weakly expressed in the late developing endosperm (Fig. 3). As reported previously by Shure et al. (1983), *ZmGBSSI* (Wx) was mainly expressed in the endosperm in middle and late developmental stages. *ZmGBSSIIa* was strongly expressed in the leaves, but weakly expressed in the endosperm, while *ZmGBSSIIb* was mainly expressed in the developing endosperm. *ZmSSIIb-2* was mainly expressed in leaves and roots, and moderately expressed in the early developing endosperm. *ZmSSIIIb-1* was mainly expressed in leaves, but weakly expressed in the roots and endosperm. *ZmSSIIIb-2* was highly expressed in roots and early developing endosperm, but moderately expressed in leaves. *ZmSSIV* was strongly expressed in leaves, moderately expressed in the late developing endosperm, but weakly expressed in roots (Fig. 3). *ZmBEIII* and *ZmISA2* were constitutively expressed. *ZmISA3* was mainly expressed in leaves and early developing endosperm, but moderately expressed in roots. *ZmPHOL* was strongly expressed in the endosperm, but weakly expressed in leaves and roots, while *ZmPHOH* was constitutively expressed (Fig. 3).

Discussion

Maize and rice share similar starch synthesis gene families

The starch biosynthesis pathway in cereals contains AGPase, ADP-Glc transporter, SS, BE, DBE, and PHO proteins (James et al. 2008; Satoh et al. 2008). There are 30 genes encoding these proteins in maize according to the previous reports and present study. The phylogenetic trees of these genes suggest that maize and rice share the same isoforms of each enzyme and transporter protein for this pathway. There are two classes of AGPL and two classes of AGPS, which occurred in the cytosol (AGPL1 and AGPS1a-1) and the plastid (other proteins) respectively; one ADP-Glc transporter; five isoforms of starch synthase (GBSS, SSI, SSII, SSIII, and SSIV); three isoforms of starch branching enzymes (BEI, BEII, and BEIII); four

Table 3 Chromosome loci of the starch synthesis genes in rice and maize

Rice					Maize					
Gene name	Mutant	Chr	Loc (M)	Acc no.	Gene name	Mutant	Chr	Loc (M)	BAC	Acc no.
AGPL1	*osagpl2*	1	25.3	AY028314	*AGPL1*	*sh2*	3	207.1	AC190966	BT016868
SSIVa		1	30.0	AY373257						
PHOH		1	36.6	NM_001051358	***PHOH***		3	170.0–171.1	AC207564	EU857639
BT1		2	5.7	AK107368	*BT1*	*bt1*	5	117.2	AC206632	BT016796
BEIIb	*ae*	2	19.3	D16201	*BEIIb*	*ae*	5	167.4	AC204651	EF433557
SSIIb		2	31.2	AF395537	*SSIIb-1*		4	~128.2	uaz218a[a]	AF019297
					SSIIb-2		5	206.7	AC190571	EF472249
AGPL4		3	28.9–29.0	NM_001057719	***AGPL4***		1	262.0–262.1	AC211200	EF694839
PHOL	*phol*	3	31.2	AF327055	***PHOL***	*sh4*	1	267.2	AC191342	EU857640
PUL		4	4.3	D50602	*PUL*		2	101.5	AC195146	AF080567
BEIIa		4	20.0–20.2	AB023498	*BEIIa*	*2a*	2	57.0	AC212883	U65948
SSIIIb		4	31.7	AF432915	***SSIIIb-1***		10	126.1	AC183945	EF472250
					SSIIIb-2		2	8.2–8.4	AC195583	EF472251
ISA2		5	19.0	NM_001061991	*ISA2*		6	137.7–137.9	AC204282	AY172633
SSIVb		5	26.2	AY373258	***SSIV***		8	116.8	AC197339	EU599036
AGPL2		5	28.6	D50317	*AGPL2*		6	159.4–159.5	AC190718	Z38111
GBSSI	*wx*	6	1.7	AB425323	*GBSSI*	*wx*	9	28.6–28.7	AC190908	AY109531
SSI		6	3.0	AY299404	*SSI*		9	23.1	AC211152	AF036891
SSIIa	*alk*	6	6.7	AF419099	*SSIIa*	*su2*	6	105.9–106.1	AC194040	AF019296
BEIII		6	15.2	NM_001064146	***BEIII***		8	1340.0	AC204884	EU333945
BEI	*flo2*	6	30.3	EF122471	*BEI*		5	64.4–64.6	AC211441	AY105679
					BEI-Like		6	69.7–69.9	AC194216	
AGPL3		7	8.0	NM_001065811	***AGPL3***		7	21.5–21.6	AC208635	EF694838
GBSSII		7	12.8	AY069940	***GBSSIIa***		7	34.4–34.6	AC210377	EF471312
					GBSSIIb		1	0–0.2	AC177838	EF472248
SSIIIa	*flo5*	8	5.3	AY100469	*SSIIIa*	*du1*	10	43.6	AC212749	AF023159
AGPS1-1	*shr/osagps2*	8	15.6	EF122437	*AGPS1a-1*	*bt2*	4	52.2–52.4	AC193357	AF330035
AGPS1-2		8	15.6	AP004459	*AGPS1a-2*		4	52.2–52.4	AC193357	DQ118038
					AGPS1b		1	209.5–209.7	AC209218	AF334960
ISA1	*su1*	8	25.8	AB015615	*ISA1*	*su1*	4	42.4	AC183521	ZMU18908
AGPS2		9	7.1	AY028315	*AGPS2*		2	165.1	AC177860	AY032604
ISA3		9	17.8	NM_001069968	*ISA3*		7	106.1	AC192599	AY172634
					ISA3-Like		2	178.8	AC194926	
SSIIc		10	15.1	AF383878	*SSIIc*		5	33.6–33.7	AC205670	EU284113

Acc no. GenBank accession number in NCBI, *ae* amylose extender, *bt* Brittle, *Chr* chromosome, *du* dull, *flo* floury, *loc* location, *Sh* shrunken, *su* sugary, *wx* waxy. Genes highlighted in bold are genes identified in this study

[a] The marker uaz218a is encompassed in *SSIIb-1*, and it is close to the marker csu100 on chromosome 4

isoforms of the starch debranching enzyme (ISA1, ISA2, ISA3, and PUL); and two isoforms of starch phosphorylase (PHOL and PHOH), one of which occurred in the cytosol (PHOH) and one in plastids (PHOL) (Fig. 1).

According to the analyses of the genomic sequences databases, two copies of some of the starch synthesis genes (*AGPL1*/*AGPL2*, *AGPS1*/*AGPS2*, *GBSSI*/*GBSSII*, *SSIIa*/*SSIIb*, *SSIIIa*/*SSIIIb* and *BEIIa*/*BEIIb*) have been retained in both the rice and maize genome, which probably remain from the ancient whole genome duplication (WGD) in the Gramineae (Figs. 1, 2; Table 3). Maize *BEIII* gene is intronless suggesting that it likely originated from a retro-transcription event and subsequently formed a new gene (Table 2), a hypothesis also supported by its non-colinear location. In addition, two copies of at least six genes (*ZmAGPS1*, *ZmSSIIb*, *ZmSSIII*, *ZmGBSSII*, *ZmBEI* and *ZmISA3*) have also been detected in the maize genome, although their transcripts and open reading frame of *BEI-L*

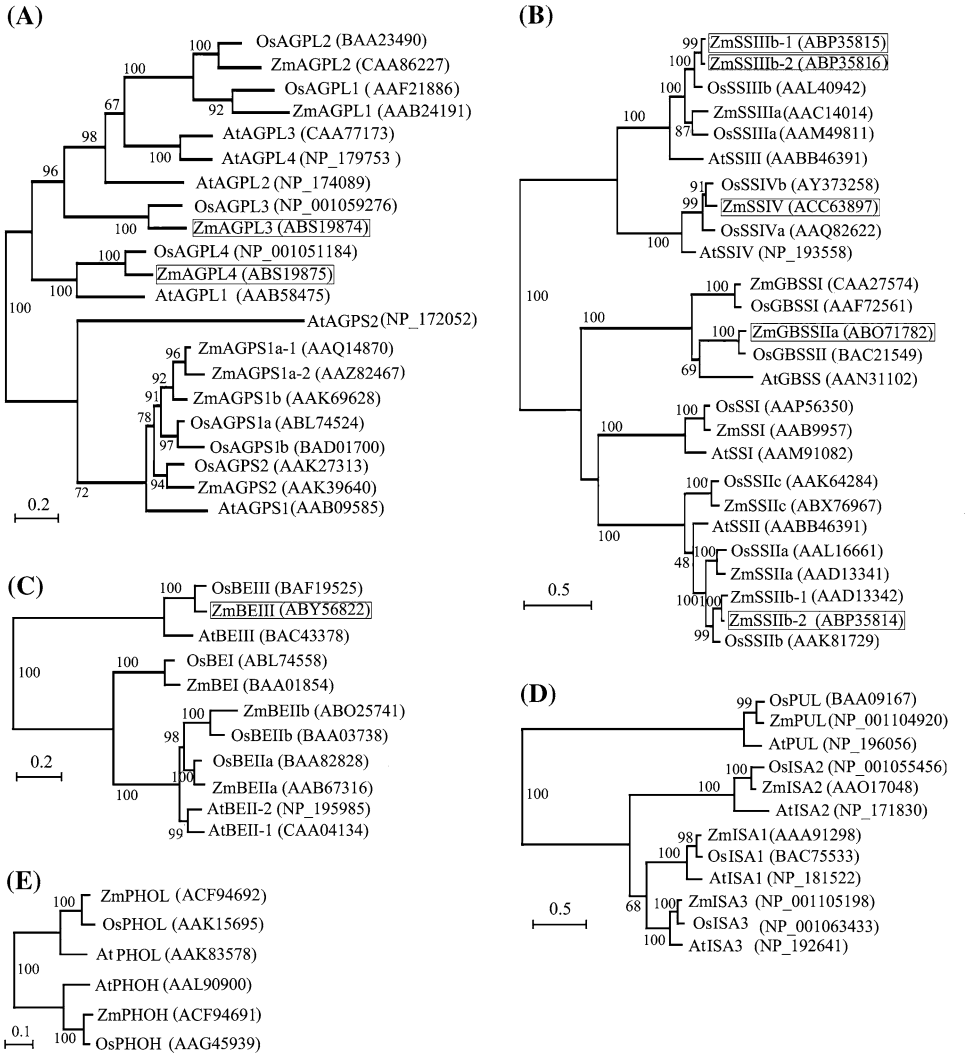

Fig. 1 Phylogenetic tree derived from the full amino acid sequences of starch synthesis proteins. The tree was drawn according to the results generated by PhyML Online analysis (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=phyml) with the JTT model. The *rectangles* around genes indicate genes identified in this study. Bootstrap values calculated for 1,000 replicates are indicated at corresponding nodes. The *scale bar* represents the branch length corresponding to the indicated substitutions per site. The GenBank accession numbers are shown in *brackets*

and *ISA3-L* were not detected in the present study. The duplicates, with the exception of *ZmGBSSIIb*, are located within the large-scale rice/maize synteny blocks (Fig. 2), suggesting that they originate from its second WGD. The non-colinear location of *ZmGBSSIIb* may be the result of chromosome rearrangements or other events, and it has diverged from *ZmGBSSIIa* in both gene length (Table 2) and expression pattern (Fig. 3).

The undetected duplicate copies of some starch synthesis genes in the maize genome may be due to the following reasons. First, the sequence of the full genome is not complete yet and some additional genes might be found later. Second, many duplicated gene copies have been lost or become non-functional during the "diploidization" of the maize genome. On the other hand, several genes (including *SSI*, *ISA1*, *ISA2*, *PUL*, *PHOL*, *PHOH*, and *BT1*) seem to exist in single copies in both the maize and rice genomes (Table 3). The divergence in tissue or time of expression (sub-functionalization) may provide one possible explanation why some of the duplicated genes were retained and others were lost. For example, *ISA1*, *PHOL* and *BT1* are mainly expressed in the endosperms and have a role infunction on the storage starch synthesis (James et al. 1995; Sullivan and Kaneko 1995; Satoh et al. 2008; Fig. 3), whereas *ISA3* and *PHOH* are mainly expressed in other tissues and the early developing endosperm (Ohdan et al. 2005; Fig. 3).

Expression divergence of duplicated starch synthesis genes in maize

Force et al. (1999) proposed that expression divergence is the first step towards the functional divergence of duplicate genes, which increases the chance of duplicate gene retention in a genome. The starch synthesis gene duplicates have diverged into groups with distinct expression characteristics in rice (Dian et al. 2005; Ohdan et al. 2005), and the expression patterns we detected indicate that the duplicates in maize have also diverged into two similar groups. Group I—including *ZmAGPL1* (Giroux and Hannah 1994), *ZmAGPL3* (Fig. 3), *ZmAGPS1a-1* (Rösti and Denyer

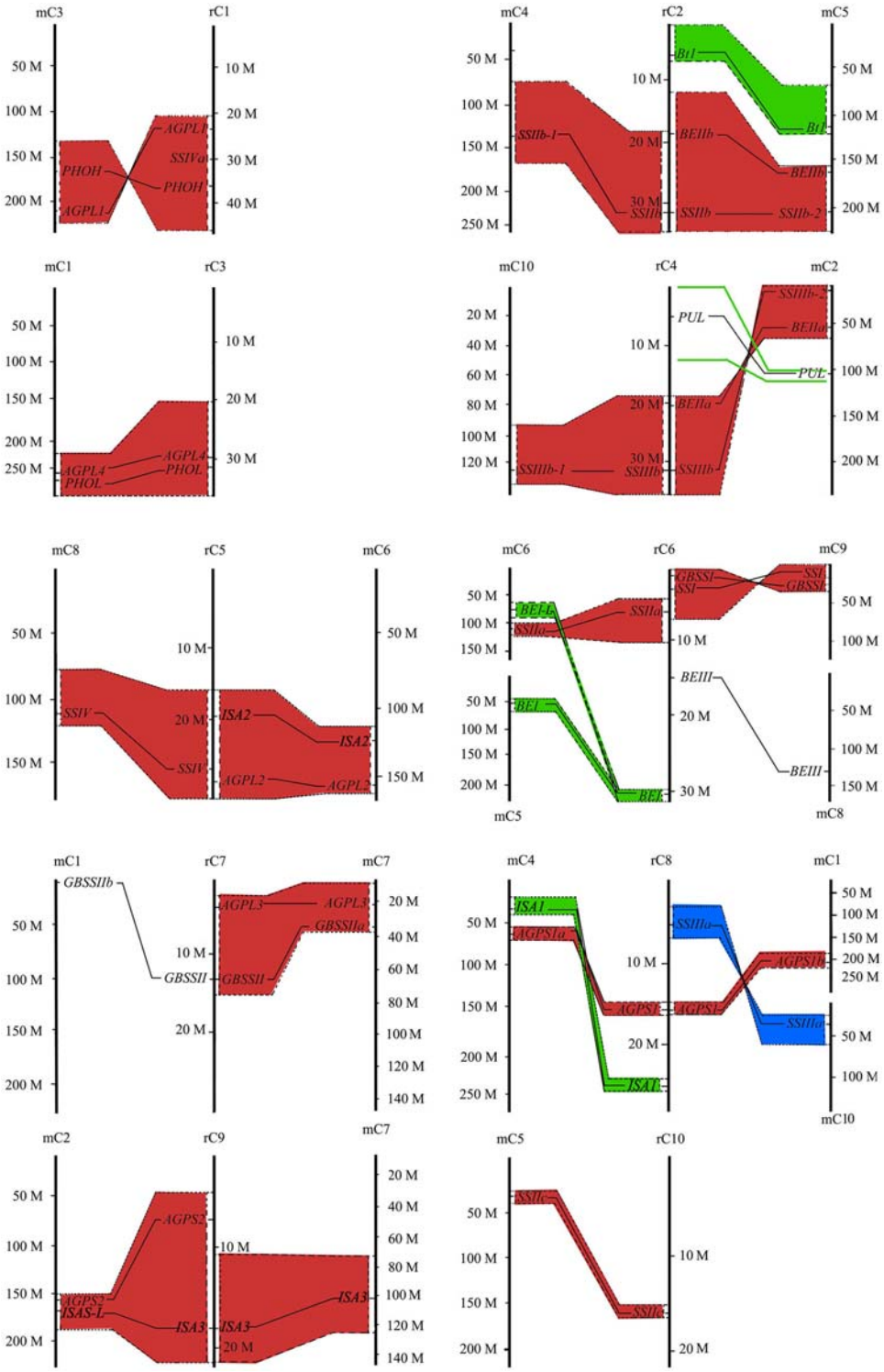

Fig. 2 Comparison of the starch synthesis genes on the rice/maize synteny physical map. *rC* Rice chromosome, *mC* maize chromosome. The *color blocks* refer to the reported rice/maize synteny (http://www.gramene.org/) containing the anchored starch synthesis genes

2007), *ZmAGPS2* (Rösti and Denyer 2007), *ZmGBSSI/Wx* (Shure et al. 1983), *ZmSSIIa* (Harn et al. 1998), *ZmSSIIIa* (Gao et al. 1998), *ZmBEIIb* (Gao et al. 1997), *ISA1* (James et al. 1995) and *PHOL*—appear to be preferentially expressed in the endosperm. Group II genes—including *ZmAGPL2* (Fig. 3), *ZmAGPL4* (Fig. 3), *ZmAGPS1b* and *ZmAGPS1a-2* (Rösti and Denyer 2007), *ZmGBSSIIa* (Fig. 3), *ZmSSIIb* (Fig. 3 and Harn et al. 1998), *ZmSSIIIb* (Fig. 3), *ZmBEIIa* (Gao et al. 1997) and *ISA3*—seem to be mainly expressed in other tissues, with some expression in the endosperm during early developmental phases in maize. The duplicate sets of *ZmAGPSI*, *ZmGBSSII*, *ZmSSIIb* and *ZmSSIIIb*, which may have arisen from the second WGD, also showed expression divergence in maize. *ZmAGPSIa-1*

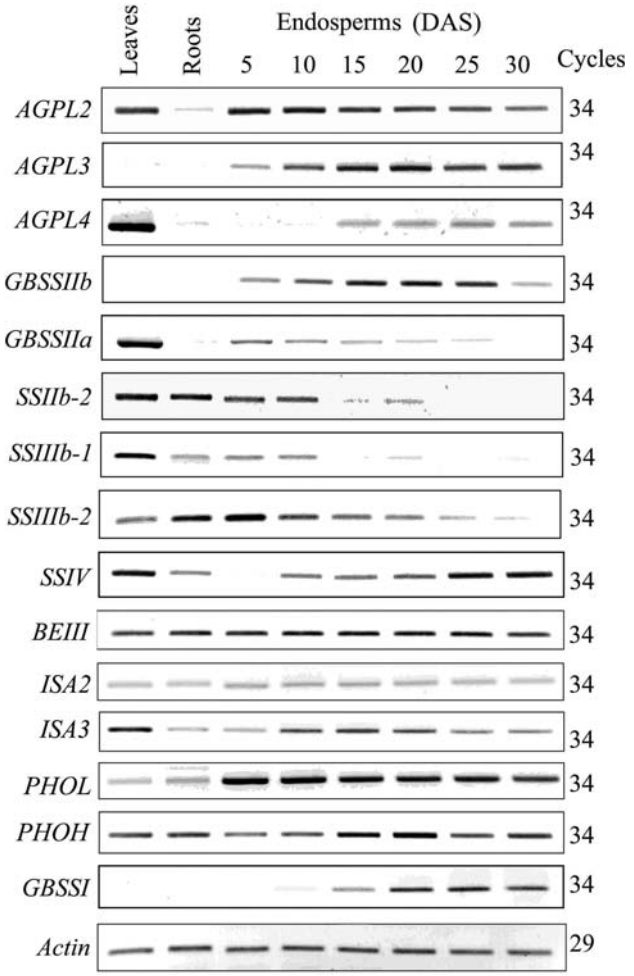

Fig. 3 Results of semi-quantitative RT-PCR analysis of partial genes in maize Total RNA samples obtained from the leaves and developing endosperm for RT-PCR were used to measure RNA expression levels of the partial starch synthesis genes in maize. Thermocycling times and temperatures were as follows: 94°C for 4 min, followed by cycles of 94°C for 30 s, respective annealing temperature (Table 1) for 45 s, 72°C for 40 s or 2 min (*GBSSIIa*), and a final extension step of 72°C for 4 min. The *bottom panel* (*Actin*) shows the loading control of *ZmActin* transcripts. *DAS* Days after silking

and *ZmAGPSIb* were previously reported to be mainly expressed in the endosperm and leaves, respectively (Rösti and Denyer 2007). *ZmSSIIb-1* transcripts are mainly expressed in leaves and have not been detected in roots or endosperm (Harn et al. 1998), while the transcripts of *ZmSSIIb-2* are mainly expressed in roots and leaves (Fig. 3). *ZmSSIIIa* is mainly expressed in the endosperm (Gao et al. 1998). *ZmSSIIIb-1* transcripts were detected mainly in leaves, and at low abundance in roots and endosperm, while *ZmSSIIIb-2* was expressed mainly in roots and early developing endosperm, but at lower abundance in leaves (Fig. 3). These results support the hypotheses that expression divergence is the first step in the functional divergence of duplicate genes and that this process increases the chance of retention of duplicate genes in a genome (Force et al. 1999).

Functional similarities of genes affecting starch yield and structure in the endosperm of maize and rice

Classical genetics has contributed to the study of starch metabolism by identifying specific molecules involved in metabolic processes and alterations in starch properties that affect the nutritional quality of starch-containing tissues and their potential industrial applications. Examples of these genes, found in both maize and rice, include: *shrunken-1* (*sh1*) and *brittle-2* (*bt2*), which code for AGPL1 and AGPS1, respectively (Giroux and Hannah 1994; Kawagoe et al. 2005; Lee et al. 2007); *waxy* (*wx*), which codes for GBSSI (Shure et al. 1983; Sano 1984); *su2/alk*, which codes for SSIIa (Takeda and Preiss 1993; Umemoto et al. 2002; Zhang et al. 2004); *dull1/flo5*, which codes for SSIIIa (Gao et al.1998; Ryoo et al. 2007; Fujita et al. 2007); *amylose extender* (*ae*), which codes for BEIIb (Yano et al. 1985; Stinard et al. 1993; Nishi et al. 2001); and *sugary-1* (*su1*), which codes for ISA1 (James et al. 1995; Nakamura et al. 1996). *Sh4/Pho1* encodes plastidial starch phosphorylase, and loss of its function results in shriveled, opaque endosperm in maize (Yu et al. 2001), and smaller rice seeds when plants are grown at low temperatures (Satoh et al. 2008). However, there is an incomplete overlap in the roles of some of these genes. *Brittle-1*(*bt1*) affects the content of endosperm starch, codes for the ADP-Glc transporter and is located in the amyloplast envelope (Sullivan and Kaneko 1995). The maize pullulanase mutant (*zpu1*-204) accumulates branched maltooligosaccharides in the endosperm (Dinges et al. 2003). Amylopectin in the endosperm of *SSI*-deficient rice mutants has reduced DP 8–12 chain content, increased content of DP 6–7 and 16–19 chains, and the endosperm starch of such rice has increased gelatinization temperatures (Fujita et al. 2006). Thus, since the same starch biosynthesis pathways operate in maize and rice, these results suggest that *Bt1* and *PUL* may affect endosperm starch content and starch structure in rice, and that *SSI* may play similar roles in maize.

Conclusion

These and previously studies indicated that the same gene subfamilies encoding the different enzymes involved in the starch biosynthesis pathway were present in both the maize and rice genomes. The rice starch synthesis gene duplicated during the WGD, except *SSIV*, were remained in maize and diverged in their transcription profiles. At least three starch synthesis gene duplications were retained in the maize genome after its second WGD.

Acknowledgments This research was supported by the Key Innovation Program of the Chinese Academy of Sciences (KSCX2-YW-G-027-3), the CAS '100 Talents' program and the CAS/SAFEA International Partnership Program for Creative Research Teams.

References

Beatty MK, Rahman A, Cao H, Woodman W, Lee M, Myers AM, James MG (1999) Purification and molecular genetic characterization of ZPU1, a pullulanase-type starch-debranching enzyme from maize. Plant Physiol 119:255–266

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17:343–360

Dian WM, Jiang HW, Wu P (2005) Evolution and expression analysis of starch synthase III and IV in rice. J Exp Bot 56:623–632

Dinges JR, Colleoni C, James MG, Myers AM (2003) Mutational analysis of the pullulanase-type debranching enzyme of maize indicates multiple functions in starch metabolism. Plant Cell 15:666–680

Fisher DK, Kim KN, Gao M, Boyer CD, Guiltinan MJ (1995) A cDNA encoding starch branching enzyme I from maize endosperm. Plant Physiol 108:1313–1314

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Fujita N, Yoshida M, Asakura N, Ohdan T, Miyao A, Hirochika H, Nakamura Y (2006) Function and characterization of starch synthase I using mutants in rice. Plant Physiol 140:1070–1084

Fujita N, Yoshida M, Kondo T, Saito K, Utsumi Y, Tokunaga T, Nishi A, Satoh H, Park JH, Jane JL, Miyao A, Hirochika H, Nakamura Y (2007) Characterization of SSIIIa-deficient mutants of rice: the function of SSIIIa and pleiotropic effects by SSIIIa deficiency in the rice endosperm. Plant Physiol 144:2009–2023

Gao M, Fisher DK, Kim KN, Shannon JC, Guiltinan MJ (1997) Independent genetic control of maize starch-branching enzymes IIa and IIb. Plant Physiol 114:69–78

Gao M, Wanat J, Stinard PS, James MG, Myers AM (1998) Characterization of dull1, a maize gene coding for a novel starch synthase. Plant Cell 10:399–412

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci 94:6809–6814

Georgelis N, Braun EL, Hannah LC (2008) Duplications and functional divergence of ADP-glucose pyrophosphorylase genes in plants. BMC Evol Biol 8:232

Giroux MJ, Hannah LC (1994) ADP-glucose pyrophosphorylase in shrunken-2 and brittle-2 mutants of maize. Mol Gen Genet 243:400–408

Han Y, Sun FJ, Rosales-Mendoza S, Korban SS (2007) Three orthologs in rice, Arabidopsis, and Populus encoding starch branching enzymes (SBEs) are different from other SBE gene families in plants. Gene 15:123–130

Harn C, Knight M, Ramakrishnan A, Guan H, Keeling PL, Wasserman BP (1998) Isolation and characterization of the zSSIIa and zSSIIb starch synthase cDNA clones from maize endosperm. Plant Mol Biol 37:639–649

James MG, Robertson DS, Myers AM (1995) Characterization of the maize gene sugary1, a determinant of starch composition in kernels. Plant Cell 7:417–429

James GM, Denyer K, Myers AM (2008) Starch synthesis in the cereal endosperm. Curr Opin Plant Biol 6:215–222

Kawagoe Y, Kubo A, Satoh H, Takaiwa F, Nakamura Y (2005) Roles of isoamylase and ADP-glucose pyrophosphorylase in starch granule synthesis in rice endosperm. Plant J 42:164–174

Knight ME, Harn C, Lilley CE, Guan H, Singletary GW, MuForster C, Wasserman BP, Keeling PL (1998) Molecular cloning of starch synthase I from maize (W64) endosperm and expression in *Escherichia coli*. Plant J 14:613–622

Lee SK, Yu SK, Wang H, Han M, Eom JS, Kang HG, Han Y, Choi SB, Cho MH, Seong HB, Gynheung A, Hahn TR, Okita TW, Jeon JS (2007) Identification of the ADP-glucose pyrophosphorylase isoforms essential for starch synthesis in the leaf and seed endosperm of rice (*Oryza sativa* L.). Plant Mol Biol 65:531–546

Nakamura Y, Umemoto T, Takahata Y, Komae K, Amano E, Satoh H (1996) Changes in structure of starch and enzyme activities affected by sugary mutations in developing rice endosperm. Possible role of starch debranching enzyme (R-enzyme) in amylopectin biosynthesis. Physiol Plant 97:491–498

Nishi A, Nakamura Y, Tanaka N, Satoh H (2001) Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. Plant Physiol 127:459–472

Ohdan T, Francisco PBJR, Sawada T, Hirose T, Terao T, Satoh H, Nakamura Y (2005) Expression profiling of genes involved in starch synthesis in sink and source organs of rice. J Exp Bot 56:3229–3244

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci 101:9903–9908

Rahman A, Wong K, Jane J, Myers AM, James MG (1998) Characterization of SU1 isoamylase, a determinant of storage starch structure in maize. Plant Physiol 117:425–435

Rösti S, Denyer K (2007) Two paralogous genes encoding small subunits of ADP-glucose pyrophosphorylase in maize, Bt2 and L2, replace the single alternatively spliced gene found in other cereal species. J Mol Evol 65:316–327

Ryoo N, Yu C, Park CS, Baik MY, Park IM, Cho MH, Bhoo SH, An G, Hahn TR, Jeon JS (2007) Knockout of a starch synthase gene OsSSIIIa/Flo5 causes white-core floury endosperm in rice (*Oryza sativa* L.). Plant Cell Rep 26:1083–1095

Sano Y (1984) Differential regulation of waxy gene expression in rice endosperm. Theor Appl Genet 68:467–473

Satoh H, Nishi A, Yamashita K, Takemoto Y, Tanaka Y, Hosaka Y, Sakurai A, Fujita N, Nakamura Y (2003) Starch-branching enzyme I-deficient mutation specifically affects the structure and properties of starch in rice endosperm. Plant Physiol 133:1111–1121

Satoh H, Shibahara K, Tokunaga T, Nishi A, Tasaki M, Hwang SK, Okita TW, Kaneko N, Fujita N, Yoshida M, Hosaka Y, Sato A, Utsumi Y, Ohdan T, Nakamura Y (2008) Mutation of the plastidial alpha-glucan phosphorylase gene in rice affects the synthesis and structure of starch in the endosperm. Plant Cell 20:1833–1849

Shure M, Wessler S, Fedoroff N (1983) Molecular identification and isolation of the waxy locus in maize. Cell 35:225–233

Song R, Messing J (2003) Gene expression of a gene family in maize based on noncollinear haplotypes. Proc Natl Acad Sci 100:9055–9060

Stinard PS, Robertson DS, Schnable PS (1993) Genetic isolation, cloning, and analysis of a mutator-induced, dominant antimorph of the maize amylose extender1 locus. Plant Cell 5:1555–1566

Sullivan TD, Kaneko Y (1995) The maize brittle 1 gene encodes amyloplast membrane polypeptides. Planta 196:477–484

Takeda Y, Preiss J (1993) Structures of B90 (sugary) and W64A (normal) maize starches. Carbohydr Res 240:265–275

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4 Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Umemoto T, Yano M, Satoh H, Shomura A, Nakamura Y (2002) Mapping of a gene responsible for the difference in amylopectin structure between japonica-type and indica-type rice varieties. Theor Appl Genet 104:1–8

Wu Y, Zhu Z, Ma L, Chen M (2008) The preferential retention of starch synthesis genes reveals the impact of whole-genome duplication on grass evolution. Mol Biol Evol 25:1003–1006

Yan H, Jiang H, Pan X, Li M, Chen Y, Wu G (2009) The gene encoding starch synthase IIc exists in maize and wheat. Plant Sci 176:51–57

Yano M, Okuno K, Kawakami J, Satoh H, Omura T (1985) High amylose mutants of rice, *Oryza sativa* L. Theor Appl Genet 69:253–257

Yu Y, Mu HH, Wasserman BP, Carman GM (2001) Identification of the maize amyloplast stromal 112-kD protein as a plastidic starch phosphorylase. Plant Physiol 125:351–359

Zhang X, Colleoni C, Ratushna V, Sirghie-Colleoni M, James MG, Myers AM (2004) Molecular characterization demonstrates that the *Zea mays* gene *sugary2* codes for the starch synthase isoform SSIIa. Plant Mol Biol 54:865–879