Kasetsart J. (Nat. Sci.) 43 : 154 - 163 (2009)

Nucleotide Polymorphism of Starch Synthesis Genes in Thai Rice

Siriphorn Jangsutthivorawat[1]* and Hugo Volkaert[2, 3]

ABSTRACT

Starch composition and cooking quality are important rice-grain characteristics of interest to breeders. One hundred and ninety-two accessions of *Oryza*, mostly from Thailand, were assayed using PCR-SSCP and alleles were identified at seven starch-synthesis gene loci (GBSSI, SSSIIIa, SSSIIIb, SSSIVa, SSSIVb, RBE1 and RBE3). The nucleotide sequences were obtained corresponding to each of the SSCP patterns observed and the sequences were submitted to GenBank (accession numbers EF990806 – EF990893). The frequency of nucleotide polymorphisms at these loci was about 1 SNP per 29 bp. A total of 136 nucleotide polymorphisms (SNPs), 16 insertions/deletions (indels) and 4 SSRs were identified. The SSRs were found at the GBSSI (CT, AATT), SSSIVb (GGGTT) and RBE1 (CT) loci. The nucleotide diversity, as measured by π, ranged from 0.00086 (RBE3) to 0.02747 (SSSIVa) and nucleotide polymorphism, as measured by q, ranged from 0.00154 (RBE3) to 0.02337 (SSSIVa). The *Oryza* accessions from Thailand in this study displayed a high nucleotide diversity compared with previous estimates of *Oryza* species based on estimates of multiple loci.

Key words: starch synthesis gene, diversity, polymorphism, SSCP, *Oryza*

INTRODUCTION

Rice (*Oryza sativa* L.) is one of the most important food crops of the world and is the staple food for approximately half of the world's population. In the future, the increase in rice production required to feed the growing human population in the world will rely primarily on genetic improvement of the existing rice cultivars. The quality of the rice grain is a major characteristic that determines price and demand in the market. The eating and cooking qualities of rice are determined by three characteristics specifying the physical and chemical properties of the starch in the endosperm: apparent amylose content, gel consistency and gelatinization temperature.

Starch is composed of two different types of glucan chains: linear amylose and branched amylopectin. The starch synthase isoform responsible for the amylose synthesis is the Granule Bound Starch Synthase (GBSS) (James *et al*., 2003). For amylopectin synthesis, the Soluble Starch Synthases (SSSs) are responsible for the chain extension, while Starch Branching Enzymes (SBEs) and Starch Debranching Enzymes (DBEs) determine the amount of branching of the chains (Nakamura, 2002).

Multi-sequence alignment analysis of the starch synthase proteins from several plant species,

1 Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand.

2 Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen, Nakhon Pathom 73140, Thailand.

3 BIOTEC, Thailand Science Park, Phaholyothin Road, Khlong Luang, Pathum Thani 10120, Thailand.

* Corresponding author, e-mail: s.jang@chiangmai.ac.th

Received date : 03/07/08 Accepted date : 29/08/08

such as rice, maize, barley, wheat, cassava, pea and potato, suggested that they could be grouped into five classes: Soluble-Starch Synthase I (SSSI); SSSII; SSSIII; SSSIV; and Granule-Bound Starch Synthase (GBSS). In rice, ten gene members have been identified in the starch synthase gene family (Hirose and Terao, 2004). The Starch Branching Enzymes (SBEs) form two groups, families A and B, that differ in terms of the lengths of chains transferred *in vitro*, with family A transferring shorter chains than family B (James *et al*., 2003). In rice, family A consists of two isoforms: RBE3 and RBE4. Family B consists of only a single gene in rice (RBE1) (Mizuno *et al*., 2001).

In 1989, Orita *et al*. developed a mobility shift assay to detect DNA polymorphism using single-strand DNA in non-denaturing polyacrylamide gel electrophoresis. The mobility shift was thought to be due to changes in the three dimensional conformation of the single-stranded DNAs which were caused by nucleotide substitutions. SSCPs were found to be allelic variants with true Mendelian inheritance and therefore they should be useful genetic markers. Sequencing is the best method to study the diversity of alleles. Sequencing technology has developed rapidly over the past two decades. The power of the technique has ensured that DNA sequencing has become one of the most-utilized molecular approaches for inferring phylogenetic history (Hillis *et al*., 1996).

This research aimed to study the genetic diversity at specific DNA loci known to be involved in starch synthesis in rice, such as starch synthase genes and starch branching enzyme genes. Targeted amplification of DNA fragments has been used to study each specific locus in the rice nuclear genome. PCR amplification followed by Single-Strand Conformation Polymorphism (PCR-SSCP) has been used to detect different alleles at these loci. Sequencing of the relevant alleles has been done to interpret the observed genetic diversity from a phylogenetic perspective. Knowledge concerning the genetic diversity of the rice genome and its wild relatives provides basic information for rice breeding. The observed sequence polymorphisms can also be developed into SNP detection assays for rapid screening in rice breeding programs.

MATERIALS AND METHODS

Plant materials

A total of 170 accessions of *O. sativa*, *O. rufipogon*, and *O. nivara* were sampled from the Biotechnology Research and Development Office, Pathum Thani. These accessions were selected from all regions of Thailand, representing different cultivation methods and a wide range in percent amylose. In addition, 3 accessions (KDML105, RD6 and Nipponbare) were obtained from the DNA Technology Laboratory, BIOTEC and 18 accessions of *O. sativa, O. rufipogon*, *O. nivara* and 1 accession of *O. glaberrima* were sampled from the germplasm collection maintained by the International Rice Research Institute (IRRI, Los Baños, Philippines).

DNA extraction

Genomic DNA was extracted from five-day-old seedlings using a MATAB method, which was similar to the CTAB method by Agrawal *et al*. (1992), but MATAB was substituted for CTAB in the extraction buffer. The quality of the extracted DNA was assessed by electrophoresis on 0.8% agarose gel run in 1X TAE buffer at 200 V for 20 min. A set of λ DNA standards of known amount (Fermentas) was used for comparison to estimate the concentration.

Selection of regions to be amplified

Sequences of each of the starch-synthesis genes from *O. sativa* (*japonica* and *indica*) were retrieved from publicly accessible DNA databases (GenBank non-redundant database and Trace Archive database of *O. sativa*). The sequences

were aligned by visual inspection using the GENEDOC program version 2.6.002 (Nicholas and Nicholas, 1997) and the intron-exon boundaries were located. DNA fragments to be amplified were selected in the 5′ upstream regions of the genes where sequence differences between japonica and indica rice could be observed. The 5′ upstream region was preferred because it contained the information for transcription initiation and was a major site through which gene expression was controlled. Furthermore, as a non-protein-coding sequence, it was expected to be more variable than the protein-coding parts of the gene.

PCR-SSCP

DNA fragments were PCR-amplified from genomic DNA using the chosen specific primers. Details of all primers are presented in Table 1. Each PCR amplification reaction (15 μl) contained: 20 ng of template DNA; 10 pmole of each of the primers; 200 μM of each dNTPs (Fermentas); 1X PCR buffer with $(NH_4)_2SO_4$; 2 mM $MgCl_2$; and 0.5 units of *Taq* DNA polymerase (Fermentas). Amplification was performed on a T1 Thermocycler (Biometra™). Cycling started with an initial 94°C for 3 min, followed by 35 cycles of 94°C for 45 sec, 50°C for 45 sec, 72°C for 1.30 min and a final extension at 72°C for 5 min. The quality and size of the amplifications were checked by agarose gel electrophoresis.

The PCR fragments were mixed with two volumes of loading dye (95% (v/v) formamide, 0.025% bromphenol blue, 0.025% xylene cyanol and 10 mM NaOH), then denatured at 95°C for 10 min and immediately placed in ice-cold water to stabilize single strands. The electrophoresis was performed on non-denaturing polyacrylamide gels (Sequagel MD, National Diagnostics, U.S.A.) using the Single-Strand DNA Polymorphism approach (Orita *et al*., 1989). For each aliquot, 3.5 μl was loaded and run in 0.6X TBE buffer at constant power (10 watt) for 14-20 h in a 4°C refrigerator.

After the electrophoresis, silver staining was used to reveal the DNA fragments following the procedure of Bassam and Caetano-Anollés (1993).

DNA sequencing

The banding patterns that appeared in the silver stained SSCP gel were analyzed. Accessions were grouped according to the different patterns and one of more representative individuals of each type were selected for sequencing. PCR-amplified DNA fragments were sent for direct sequencing using specific internal sequencing primers. For amplification for direct sequencing from PCR, the same conditions were used, except that the primer and dNTP concentrations were reduced by 15% and the volume was increased to 50 μl.

Evaluation of DNA sequence diversity

DnaSP Version 4.0 was used to calculate diversity measures for the sequence data. The level of intraspecific genetic variation was calculated based on estimates of the average number of nucleotide differences per site between two sequences (π) (Nei, 1987) and Watterson's estimate (θ), which is similar to π, but focuses on the number of segregation sites (Watterson, 1975). Tajima's *D* test was used to estimate the neutrality of the SNP polymorphism (Tajima, 1989).

RESULTS

Development of nuclear gene primers

For all genes, *Oryza* DNA sequences were retrieved from the NCBI (http://ncbi.nlm.nih.gov/) databases using keyword searches and BLAST. Nucleotide sequences were first aligned with the GENEDOC and then the start codon and introns-exon boundaries were located. DNA regions to be amplified were chosen from regions that exhibited sequence difference between japonica and indica rice in the 5′ upstream region.

This region was preferred, because it contained the information for transcription initiation and was a major site through which gene expression was controlled and as a non-protein-coding sequence it was expected to be more variable than the protein-coding parts of the gene. The primer sequences are listed in Table 1.

PCR-SSCP and evaluation of DNA sequence diversity

A total of 192 accessions of *Oryza* were assayed using PCR-SSCP and alleles at seven starch-synthesis gene loci (GBSSI, SSSIIIa, SSSIIIb, SSSIVa, SSSIVb, RBE1 and RBE3) were identified. Each of the primer sets successfully amplified a single fragment in all samples. On agarose gel, differences among alleles were not visible. All amplified DNA fragments were denatured and electrophoresed on non-denaturing polyacylamide gel (SSCP). Accessions representing each of the SSCP patterns were selected for large-scale PCR amplification and sent for direct sequencing**.** The sequence data for all accessions were obtained by matching the SSCP pattern with sequences of those samples selected for sequencing.

For the GBSSI locus, 16 SSCP patterns were identified and 21 PCR products were sent for direct sequencing. Sequences corresponding to each of the alleles were deposited at GenBank (accession numbers EF990806-EF990821). The length of the sequences ranged between 540 and 563 bp. Nucleotide polymorphisms were found at 23 positions including 19 substitutions, 2 indels and 2 SSRs (Table 2). The first SSR region was 7-19 CT dinucleotide repeats and the second SSR region was 5-6 AATT tetranucleotide repeats. Nucleotide diversity (π) was 0.01713 and Watterson's estimator (θ) was 0.01017. Tajima's *D* test for selection did not show a significant deviation from the neutral expectation (D = 1.97801, $P > 0.05$) (Table 3).

At the $(CT)_n$ SSR region, 10 different $(CT)_n$ repeat alleles were found including $(CT)_7$, $(CT)_8$, $(CT)_9$, $(CT)_{10}$, $(CT)_{11}$, $(CT)_{12}$, $(CT)_{16}$, $(CT)_{17}$, $(CT)_{18}$ and $(CT)_{19}$, of which $(CT)_{17}$ was the most frequent, and $(CT)_7$, $(CT)_9$, $(CT)_{10}$ and $(CT)_{12}$ were identified for the first time. The allele frequencies, haplotypes of the surrounding sequence and the percentage of waxy rice corresponding to each $(CT)_n$ variant are detailed in Table 4.

Table 1 Gene-specific PCR primers used for the SSCP technique.

Locus	Accession (Position)	Primers
GBSSI	AB002542 (2919-3480)	F-ACCATTCCTTCAgTTCTTTgTC
		R-gTTTCTCCAgTggCgAgAg
SSSIIIa	AF432915 (219-828)	F-TCCTAAAAgCTgggCCAATg
		R-CggTggATCggCATCTCTC
SSSIIIb	AP005441 (1900-2516)	F-AAAATAACCTACATATTTCAAAACAgC
		R-TAgCTTAgCTTCATCCgTCgCATC
SSSIVa	AP003292 (812-1414)	F-TTggTTgTgAAACCgTgAAAgC
		R-CggCCCACTCTgACTTTgg
SSSIVb	AC12135 (1371-1865)	F-TCTCAgTAgTCTgCTCCTgC
		R-TCACTggAAACAgATgCTTC
RBE1	D10838 (3004-3586)	F-AgTgTCAgCATAgAAATCTC
		R-gAAACCACgCCCAggCgAAC
RBE3	AP004879 (607-1169)	F-CCCTCCgCTCCTCCTAgCTTC
		R-TCgCCCTCggggATCATCAC

Table 2 Locus and marker information.

Locus	PCR product size on agarose gel (bp)	Nucleotide sequence (bp)	No. of SSCP haplotypes observed	Nucleotide polymorphism			
				Substitution	Indel	SSR	Total
GBSSI	550	540-563	16	19	2	2	23
SSSIIIa	600	609-616	21	28	8	0	36
SSSIIIb	600	619-620	10	9	1	0	10
SSSIVa	600	577-613	14	61	8	0	69
SSSIVb	400	420-442	11	6	1	1	8
RBE1	550	574-619	10	8	4	1	13
RBE3	550	557	6	5	0	0	5
Total		3905-3976		136	24	4	164

Table 3 Summary of nucleotide polymorphisms and neutrality tests calculated with the DnaSP program*.

Locus	N^a	S^b	p^c	q^d	D^e
GBSSI	174	23	0.01713	0.01017	1.97801
SSSIIIa	178	36	0.00662	0.01027	-1.03672
SSSIIIb	187	10	0.00225	0.00278	-0.45068
SSSIVa	183	69	0.02747	0.02337	0.54053
SSSIVb	179	8	0.00270	0.00334	-0.42674
-Intron1	179	7	0.00279	0.00323	-0.29635
-Exon2	179	1	0.00189	0.00434	-0.58161
RBE1	185	13	0.00217	0.00483	-1.43046
-5′UP	185	6	0.00104	0.00344	-1.49545
-Intron1	185	7	0.00858	0.01523	-1.00242
RBE3	188	5	0.00086	0.00154	-0.85901
-5′UP	188	1	0.00205	0.00103	1.01410
-Exon1	188	1	0.00034	0.00114	-0.71275
-Intron1	188	2	0.00014	0.00223	-1.29335
-Exon2	188	1	0.00073	0.00202	-0.65296
-Coding	188	2	0.00048	0.00146	-0.92124
-Noncoding	188	3	0.00113	0.00161	-0.47651

[a] Total number of sequences

[b] Total number of polymorphic sites

[c] Average number of pair wise nucleotide differences per site (Nei, 1987) calculated on the total number of polymorphic sites

[d] Watterson's estimator of q per base pair (Watterson, 1975) calculated on the total number of polymorphic sites

[e] Tajima's *D* test (Tajima 1989)

* Note that no values are significant at $P < 0.05$

Table 4 Diversity of the $(CT)_n$ repeat at the GBSSI locus.

$(CT)_n$	Haplotype of flanking sequence	This study		Bao *et al.* (2002)		Ayres *et al.* (1997)	
		No. of accessions (%)	No. of waxy rice	No. of accessions (%)	No. of waxy rice	No. of accessions (%)	No. of waxy rice
7	I	21 (12.07)	13	0	0	0	0
8	D	6 (3.45)	4	0	0	1 (1.09)	0
9	F	2 (1.15)	2	0	0	0	0
10	K, L, O, P	11 (6.32)	0	0	0	0	0
11	B	38 (21.84)	16	3 (4.48)	0	6 (6.52)	0
12	H	7 (4.02)	3	0	0	0	0
14		0	0	0	0	8 (8.69)	0
16	N	1 (0.58)	0	4 (5.97)	4	1 (1.09)	1
17	A, G, M	56 (32.18)	16	38 (56.72)	36	5 (5.43)	2
18	C, E	25 (14.37)	8	9 (13.43)	5	27 (29.35)	0
19	J	7 (4.02)	1	12 (17.91)	11	7 (7.61)	0
20		0	0	1 (1.49)	0	37 (40.22)	0
	Total	174 (100)	63	67 (100)	56	92 (100)	3

At the SSSIIIa locus, the SSCP assay revealed 21 haplotypes. A total of 63 PCR products were sent for direct sequencing, resulting in 21 variants corresponding to each of the SSCP patterns. The sequences were submitted to GenBank (accession numbers EF990822-EF990842). The length of the sequences ranged between 609 and 616 bp. Nucleotide polymorphisms were found at 36 positions consisting of 8 indels and 28 substitutions (Table 2). Nucleotide diversity (π) was 0.00662 and Watterson's estimator (θ) was 0.01027. Tajima's *D* test for selection did not show a significant deviation from the neutral expectation (D = -1.03672, $P > 0.05$) (Table 3).

At the SSSIIIb locus, 10 SSCP patterns were observed and 32 PCR products were sent for direct sequencing. The sequences corresponding to the 10 haplotypes were deposited at GenBank (accession numbers EF990843-EF990852). The length of the sequences ranged between 619 and 620 bp. Nucleotide polymorphisms were found at 10 positions: 1 indel and 9 substitutions (Table 2). Nucleotide diversity (π) was 0.00225, Watterson's estimator (θ) was 0.00278, and Tajima's D = -0.45068 ($P > 0.05$). The negative value of Tajima's *D* arose from an excess of low-frequency nucleotide polymorphisms (Table 3).

At the SSSIVa locus, 14 SSCP patterns could be identified, with 25 PCR products sent for direct sequencing. Sequences corresponding to each of the alleles were deposited at GenBank (accession numbers EF990853-EF990866). The length of the sequences ranged between 577 and 613 bp. Nucleotide polymorphisms were found at 69 positions: 8 indels and 61 substitutions (Table 2). Nucleotide diversity (π) was 0.02747 and Watterson's estimator (θ) was 0.02337. Tajima's *D* test for selection indicated no significant deviation from the neutral expectation (D = 0.54053, $P > 0.05$) (Table 3).

At the SSSIVb locus, the SSCP assay revealed 11patterns, with 22 PCR products sent for direct sequencing. The sequences corresponding to each of the haplotypes were submitted to GenBank (accession numbers EF990883-EF990893). The length of the sequences ranged between 420 and 442 bp.

Nucleotide polymorphisms were found at 8 positions including 6 substitutions, 1 indel and 1 SSR (Table 2). The SSR region was 3-4 GGGTT pentanucleotide repeats. Nucleotide diversity (π) was 0.00270. Watterson's estimator (θ) was 0.00334 and Tajima's D = -0.42674 ($P > 0.05$). While nucleotide diversity indices of the noncoding region ($\pi = 0.00279$) were higher than those of the coding region ($\pi = 0.00189$), the Watterson's estimate of the noncoding region ($\theta = 0.00323$) was lower than for the coding region ($\theta = 0.00434$). But a separate Tajima's D test for each region revealed no significant departure from the neutral expectation (-0.29635 for the noncoding region and -0.58161 for the coding region, $P > 0.05$) (Table 3).

At the RBE1 locus, ten SSCP patterns were identified and 13 PCR products were sent for direct sequencing. Sequences corresponding to each of the alleles were deposited at GenBank (accessions EF990867-EF990876). The length of the sequences ranged between 574 and 619 bp. Nucleotide polymorphisms were found at 13 positions: 4 indels, 8 substitutions and 1 SSR (Table 2). The SSR region was CT dinucleotide repeats. Nucleotide diversity (π) was 0.00217 and Watterson's estimator (θ) was 0.00483. Tajima's D test for selection did not show a significant deviation from the neutral model (D = -1.43046, $P > 0.05$). The diversity indices of the intron1 region ($\pi = 0.00858$ and $\theta = 0.01523$) were higher than for the 5′ upstream region ($\pi = 0.00104$ and $\theta = 0.00344$), whereas no polymorphisms were observed in the exon1 and exon2. Separate Tajima's D tests for the intron1 region and the 5′ upstream region revealed no significant departures from the neutral expectation , with values of -1.00242 for the intron1 region and -1.49545 for the 5′ upstream region, $P > 0.05$ (Table 3).

At the RBE3 locus, only 6 SSCP patterns were observed, with 9 PCR products sent for direct sequencing. The sequences were submitted to GenBank (accession numbers EF990877-EF990882) and the length of all sequences was 557 bp. Nucleotide polymorphisms were found at five positions and all of them were substitutions (Table 2). Nucleotide diversity (π) was 0.00086, Watterson's estimator (θ) was 0.00154, with Tajima's D = -0.85901 ($P > 0.05$). The diversity indices of the noncoding regions ($\pi = 0.00113$ and $\theta = 0.00161$) were higher than those of the coding region ($\pi = 0.00048$ and $\theta = 0.00146$). Within the noncoding regions, the nucleotide diversity indices of the 5′ upstream region ($\pi = 0.00205$) were higher than those of the intron1 region ($\pi = 0.00014$), however, the Watterson's estimate of the 5′ upstream region ($\theta = 0.00103$) was lower than for the intron1 region ($\theta = 0.00223$). Also here, Tajima's D test for each region revealed no significant departure from the neutral expectation (-0.47651 for the noncoding regions, -0.92124 for the coding regions, 1.01410 for the 5′ upstream region and -1.29335 for the intron1 region, $P > 0.05$) (Table 3).

DISCUSSION

A large number of insertion/deletion (InDels) and single-nucleotide polymorphisms (SNPs) together with four SSRs were observed in the sequenced regions. The noncoding regions, including 5′ upstream regions and introns, had more polymorphisms than the protein-coding regions. It has also been reported from other species that the noncoding sequences had a much higher divergence than the coding-region sequences (Tang *et al.*, 2004). The frequency of nucleotide polymorphism in this study was about one SNP per 29 bp. In the coding regions, the frequency was one SNP per 148 bp, whereas in the noncoding regions the frequency increased to one SNP per 26 bp. This was consistent with the prediction of purifying selection against deleterious or slightly-deleterious nucleotide substitutions in the coding region. It has been reported that the frequency of SNP polymorphism

changes with genomic region and with different samples (Bao *et al*., 2006a).

At the SSR located close to the 5′-leader intron splice site of the *wx* gene (GBSSI), previous studies have reported eight $(CT)_n$ microsatellite length polymorphisms (Ayres *et al*., 1997; Bao *et al*., 2002). Ten $(CT)_n$ repeat variants were observed in this study, of which $(CT)_{17}$ was the most frequent. Alleles $(CT)_{14}$ and $(CT)_{20}$ were not found in this study. Four new alleles were detected in this study: $(CT)_7$; $(CT)_9$; $(CT)_{10}$; and $(CT)_{12}$; that were not detected in previous reports (Ayres *et al*., 1997; Bao *et al*., 2002, 2006b). However, only a very limited number (three) of waxy rice accessions were included in the work of Ayres *et al*. (1997) and a low portion (16.42%) of non-waxy rice accessions were included in the work of Bao *et al*. (2002). The present study included 63 accessions of waxy rice (36.21% of total) and 111 accessions of non-waxy rice including 34 accessions of intermediate amylose content and 77 accessions of high amylose content.

Based on their accessions, Bao *et al*. (2006b) concluded that the alleles with fewer repeats ($n \leq 12$) were highly associated with higher apparent amylose content (AAC) and those with more repeats were highly associated with a lower AAC in US rice and Chinese rice germplasm. The *wx* microsatellite was deemed polymorphic enough to distinguish most rice cultivars in different amylose classes and stable enough to be easily traced through multiple generations of the US rice pedigree, so it could be used for marker-assisted selection (Ayres *et al*., 1997; Bao *et al*., 2006b). However, in this study, the alleles with fewer repeats ($n \leq 12$) excluding $(CT)_{10}$ were associated with a lower AAC. The results of the study on the relationship between allele polymorphism and waxy-rice were thus not conclusive, because there was conflict between the results from the present and previous studies. So the correlation between the (CT) repeat length and the amylose content found in previous studies cannot be extended to rice diversity in Thailand.

To test the standard neutral equilibrium model, Tajima's *D* test (Tajima, 1989) was used as implemented in DnaSP to address whether the data showed any indication that selection had shaped the levels of variation at any individual locus. The values of Tajima's *D* varied widely across the seven loci, but no significant value was observed at any locus. Thus the neutral model could not be rejected at $P < 0.05$ (Table 3). The GBSSI locus had the highest Tajima's *D* value and Tajima's *D* should be higher in populations that have experienced a recent bottleneck because of the preferential loss of the low-frequency variant (Zhu *et al*., 2007).

CONCLUSION

A total of 192 accessions of cultivated and wild *Oryza* were assayed using PCR-SSCP and alleles at seven starch-synthesis gene loci (GBSSI, SSSIIIa, SSSIIIb, SSSIVa, SSSIVb, RBE1 and RBE3) were identified. The nucleotide sequences corresponding to each of the observed SSCP patterns were obtained. Both indel-included SSR and SNP were present in the sequenced regions. A total of 24 indel polymorphisms and 136 SNPs were found. An average of 1 SNP per 29 bp was found, or 1SNP per 26 bp in the noncoding and 1SNP per 148 bp in the coding regions. This large number of SNPs opens the prospect of using whole-genome association studies in diverse populations to discover markers closely linked to genes contributing to quantitative traits such as starch properties. No significant Tajima's *D* value was observed at any locus.

ACKNOWLEDGEMENTS

This research was supported by the Center for Agricultural Biotechnology under the Higher Education Development Project, Commission on Higher Education, the Ministry

of Education.

Thanks are also due to the Biotechnology Research and Development Office, Pathum Thani, the DNA Technology Laboratory, BIOTEC and the International Rice Research Institute (IRRI, Los Banos, Philippines) for supplying seeds.

LITERATURE CITED

Agrawal, G.K., R.N. Pandey and V.P. Agrawal. 1992. Isolation of DNA from *Choerospondias asillaris* leaves. **Biotech. Biodiv. Lett.** 2: 19-24.

Ayres, N.M., A.M. McClung, P.D. Larkin, H.F.J. Bligh, C.A. Jones and W.D. Park. 1997. Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. **Theor. Appl. Genet.** 94: 773-781.

Bao, J.S., Y.R. Wu, B. Hu, P. Wu, H.R. Cui and Q.Y. Shu. 2002. QTL for rice grain quality based on a DH population derived from parents with similar apparent amylose content. **Euphytica** 128: 317-324.

Bao, J.S., H. Corke and M. Sun. 2006a. Nucleotide diversity in starch synthase IIa and validation of single nucleotide polymorphisms in relation to starch gelatinization temperature and other physico-chemical properties in rice (*Oryza sativa* L.). **Theor. Appl. Genet**. 113: 1171-1183.

Bao, J.S., H. Corke and M. Sun. 2006b. Microsatellites, single nucleotide polymorphism and a sequence tagged site in starch-synthesizing genes in relation to starch physicochemical properties in nonwaxy rice (*Oryza sativa* L.). **Theor. Appl. Genet**. 113: 1185-1196.

Bassam, B.J. and G. Caetano-Anollés. 1993. Silver staining of DNA in polyacrylamide gels. **App. Biochem. Biotechn.** 42: 181-188.

Hillis, D.M., B.K. Mable, A. Larson, S.K. Davis and E.A. Zimmer. 1996. Nucleic acids IV: sequencing and cloning, pp. 321-381. *In* D.M. Hillis, C. Moritz and B.K. Mable (eds.). **Molecular systematic** 2nd ed. Sinauer Associates, Inc., Massachusetts.

Hirose, T. and T. Terao. 2004. A comprehensive expression analysis of the starch synthase gene family in rice (*Oryza sativa* L.). **Planta** 220: 9-16.

James, M.G., K. Denyer and A.M. Myers. 2003. Starch synthesis in the cereal endosperm. **Curr. Opin. Plant Biol.** 6: 1-8.

Mizuno, K., E. Kobayashi, M. Tachibana, T. Kawasaki, T. Fujimura, K. Funane, M. Kobayashi and T. Baba. 2001. Characterization of an isoform of rice starch branching enzyme, RBE4, in developing seeds. **Plant Cell Physiol.** 42: 349-357.

Nakamura, Y. 2002. Toward a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue. **Plant Cell Physiol.** 43: 718-725.

Nei, M. 1987. **Molecular Evolutionary Genetics.** Columbis Univesity Press, New York

Nicholas, K.B. and H.B. Nicholas. 1997. GeneDoc: a tool for edition and annotating multiple sequence alignments. Available Source: http//www.psc.edu/biomed/genedoc/, January 20, 2005.

Orita, M., H. Iwahana, H. Kanazawa, K. Hayashi and T. Sekiya. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. **Proc. Natl. Acad. Sci. U.S.A.** 86: 2766-2770.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics** 123: 585-595.

Tang, J., H. Xia, M. Cao, X. Zhang, W. Zeng, S. Hu, W. Tong, J. Wang, J. Wang, J. Yu, H. Yang and L. Zhu. 2004. A Comparison of Rice

Chloroplast Genomes. **Plant Phisiol**. 135: 412-420.

Watteson, G.A. 1975. On the number of segregating sites in genetical models without recombination. **Theor. Popul. Biol.** 7: 256-276. *Cited in* K.M. Olsen and M.D. Purugganan. 2002. Molecular evidence on the origin and evolution of glutinous rice. **Genetics** 162: 941-950.

Zhu, Q., X. Zheng, J. Luo, B.S. Gaut and S. Ge. 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. **Mol. Biol. Evol.** 24: 875-888.