THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Doctor of Philosophy (Agricultural Biotechnology)

DEGREE

Agricultural Biotechnology — **FIELD**

Interdisciplinary Graduate Program — **PROGRAM**

TITLE: Genetic Diversity of Starch Synthesis Genes in Rice

NAME: Miss Siriphorn Jangsutthivorawat

THIS THESIS HAS BEEN ACCEPTED BY

(Mr. Hugo Volkaert, Ph.D.) **THESIS ADVISOR**

(Associate Professor Sontichai Chanprame, Ph.D.) **COMMITTEE MEMBER**

(Mr. Theerayut Toojinda, Ph.D.) **COMMITTEE MEMBER**

(Associate Professor Pongthep Akratanakl, Ph.D.) **GRADUATE COMMITTEE CHAIRMAN**

APPROVED BY THE GRADUATE SCHOOL ON _____

(Associate Professor Gunjana Theeragool, D.Agr.) **DEAN**

THESIS

GENETIC DIVERSITY OF STARCH SYNTHESIS GENES IN RICE

SIRIPHORN JANGSUTTHIVORAWAT

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy (Agricultural Biotechnology)
Graduate School, Kasetsart University
2008

Siriphorn Jangsutthivorawat 2008: Genetic Diversity of Starch Synthesis Genes in Rice. Doctor of Philosophy (Agricultural Biotechnology), Major Field: Agricultural Biotechnology, Interdisciplinary Graduate Program. Thesis Advisor: Mr. Hugo Volkaert, Ph.D. 114 pages.

Starch composition and cooking quality of the rice grain are important characters of interest to breeders. One hundred and ninety-four accessions of *Oryza* were assayed using PCR-SSCP and alleles at seven starch synthesis genes loci (GBSSI, SSSIIIa, SSSIIIb, SSSIVa, SSSIVb, RBE1 and RBE3) were identified. The nucleotide sequences corresponded to each of the observed SSCP patterns. The frequency of nucleotide polymorphism in this study is about one SNP per 29 bp. At the $(CT)_n$ region in the GBSSI locus, four new alleles were observed $(CT)_7$, $(CT)_9$, $(CT)_{10}$ and $(CT)_{12}$. Allele $(CT)_7$ was present only in Thai rice and the ORF-100 deletion types rice. We found two LD block observed in GBSSI and SSIV-1 loci and four LD groups, including polymorphism from different loci, were detected at $r^2 > 0.2$ and $P < 0.05$ indicating relationship between haplotype of seven gene loci. For association analysis, 13, 5, 5 and 1 polymorphisms from GBSSI, SSSIIIa, SSSIVa and SSSIVb, respectively, were significantly associated with AAC, especially, 7 polymorphisms from GBSSI were significant at $P < 0.0001$. Whereas, 10, 9, 7 and 1 polymorphisms from GBSSI, SSSIVa SSSIVb and RBE1, respectively, were significantly associated with GT, especially, one polymorphism from GBSSI were significant at $P < 0.01$. Finally, 11, 3 and 1 polymorphisms from GBSSI, SSSIIIa and SSSIVb, respectively, were significantly associated with GC, especially, 2 polymorphisms from GBSSI were significant at $P < 0.0001$. The observed sequence polymorphisms can also be developed into SNP detection assays for rapid screening in rice breeding programs.

_____ _____ ____/____/____

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENTS

The author would like to thank her advisor, Dr. Hugo Volkaert, who has been an admirable colleague and mentor. His expertise and accurate judgments have been invaluable for her work, their kind support and encouragement has been appreciative. Acknowledgements are extended to her committee, Dr. Theerayut Toojinda and Associate Professor Sontichai Chanprame, for their valuable comments and suggestions, which considerably improved this manuscript.

The author would like to acknowledge these sources of plant, namely International Rice Research Institute (IRRI), The Philippines and Biotechnology Research and Development Office, Ministry of Agriculture and Cooperatives. This study has been supported by Center for Agricultural Biotechnology (CAB), Kasetsart University. The author is grateful for their support. The Department of Biology, Faculty of Science, Chiang Mai University, permitted the author to pursue Ph.D. study with no hesitation, which she is thankful.

The author is deeply in debt to kind supports of members in CAB, and last, but not least, family members of the Jangsutthivorawat owed the author's gratefulness. They all, especially Mr. Viboon and Mrs. Wanatchaphorn Jangsutthivorawat, parents, my sisters and my brothers have been so helpful in completing this degree. Without whose endless patience and encouragement, the author could not have been able to start nor finish this work. The author would like to express her deep gratitude and dedicate this degree to them.

Siriphorn Jangsutthivorawat

October 2008

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF ABBREVIATIONS

π	=	the average number of nucleotide differences per site between two sequences
AAC	=	apparent amylose content
AFLP	=	amplified fragment length polymorphism
ASV	=	alkali spread value
Bp	=	base pairs
°C	=	degree celsius
cDNA	=	complementary DNA
D’	=	standardized disequilibrium coefficient
DBE	=	starch debranching enzyme
DNA	=	deoxyribonucleic acid
dNTP	=	deoxynucleotide triphosphate
EDTA	=	ethylenediamine tetraacetic acid
GBSS	=	granule bound starch synthase
GC	=	gel consistency
GT	=	gelatinization temperature
hr.	=	hour
L	=	liter
LD	=	linkage disequilibrium
M	=	molar
µL	=	microliter
mg	=	milligram
mL	=	milliliter
mm	=	millimeter
min	=	minute
N	=	normality
nm	=	nanometer
O.	=	*Oryza*
PCR	=	polymerase chain reaction

LIST OF ABBREVIATIONS (Continued)

r	=	allele-frequency correlations
RAPD	=	Random amplified polymorphic DNA
RBE	=	rice starch branching enzyme
RFLP	=	restriction fragment length polymorphism
RNase	=	ribonuclease
RT-PCR	=	reverse transcription - polymerase chain reaction
SBE	=	starch branching enzyme
SNP	=	single nucleotide polymorphism
SSCP	=	single strand conformation polymorphism
SSLP	=	simple sequence length polymorphism
SSR	=	simple sequence repeat
SSS	=	soluble starch synthase
TAE	=	tris-acetate-EDTA electrophoresis buffer solution
TBE	=	tris-borate-EDTA electrophoresis buffer solution
θ	=	Watterson's estimate
Wx	=	waxy

GENETIC DIVERSITY OF STARCH SYNTHESIS GENES IN RICE

INTRODUCTION

Rice (*Oryza sativa* L.) is one of the most important food crops of the world and is the staple food of approximately one-half of the world's population. In the future, the increase of rice production required to feed the growing human population in the world will rely primarily on genetic improvement of the existing rice cultivars. Advances in molecular breeding hold tremendous potential for genetic improvement of rice with beneficial genes from wild rice species, which can be introduced in cultivars using marker assisted selection. Marker assisted selection uses markers that are located near the gene of interest. The ability to detect genetic diversity at specific loci of interest to the breeder will be a major component of marker assisted selection. An important character that breeders are interested in is starch composition and cooking quality of the grain. Both genetic and environmental factors are responsible for variation in starch composition and cooking quality of rice.

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, starch branching enzyme genes, and starch debranching enzyme genes. Targeted amplification of DNA fragments has been used to study genetic diversity of specific loci in the rice nuclear genome. PCR amplified fragments 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 in 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.

OBJECTIVES

1. To study the genetic diversity at specific DNA loci known to be involved in starch synthesis in rice. PCR amplification followed by Single-Strand Conformation Polymorphism (PCR-SSCP) has been used to detect different alleles at these loci.

2. To clarify whether SSCP haplotypes of seven loci coding for starch synthesis genes could differentiate the indigenous rice varieties according to starch physico-chemical quality groups.

LITERATURE REVIEW

1. Rice

Rice (*Oryza sativa* L.) is one of the most important food crops of the world and is the staple food of approximately one-half of the world's population. Global rice production is about 590 million metric tons per year. Thailand produces about 25 million tones (unhusked rice). Rice production in Thailand is sufficient for local consumption, and the country exports around 7-8 million tones (milled rice) annually (Office of Agricultural Economics, www.oae.go.th).

The cooking quality of the rice grains is a major character that determines prices and demand in markets. However, different parts of the world have dissimilar tastes. For example, Indians like Basmati, which is dry and not sticky after cooking, and Japanese like their rice sticky but smooth in texture and soft after cooking. The eating and cooking qualities of rice are determined by three characters specifying the physical and chemical properties of the starch in the endosperm i.e. the apparent amylose content (AAC), gel consistency (GC), and gelatinization temperature (GT) (Zhou *et al.*, 2003). Two major types of rice starch are recognized i.e. glutinous rice and non-glutinous rice, which are determined by the structure of the starch molecules. Glutinous rice lacks the straight chain amylose starch component, consisting entirely of amylopectin starch. Non-glutinous rice is more common as the staple food and has an amylose starch content up to 30% of the total starch in the endosperm. The normal non-glutinous rice grain is translucent, while the glutinous rice grain generally has a turbid or opaque appearance.

Rice belongs to the genus *Oryza* L. The genus *Oryza* is small, including only about 23 species, but it is remarkable in the diverse ecological adaptations of its species (Vaughan *et al.*, 2003). Sun *et al.* (2002) using a combination of nuclear, mitochondrial, and chloroplastic DNA markers distinguished seven groups in cultivated rice, while a total of 16 different groups were found in wild rice. They calculated that the number of alleles in cultivated rice is about 60% of that in common

wild rice in terms of isozyme and DNA markers. These results verified that the genetic diversity has decreased during domestication, and some alleles were lost through natural and human selection. Over the period that humans selected cultivated rice from wild rice, they probably also preferred those rice strains that had eating qualities they liked.

2. Starch

Cereal crops accumulate starch in the seed endosperm. This starch serves as an energy reserve for the plant embryo and is the primary carbohydrate component in the diets of humans and livestock. Cereal starch also has numerous important industrial applications. It is widely used in the food and beverage industries as a thickener, as well as for some manufacturing applications in the paper and textile industries.

Starch is composed of two different types of glucan chains, linear amylose and branched amylopectin. Both types of starch molecules have the same basic structure, but differ in their length and degree of branching, which ultimately affects the physico-chemical properties of these polymers (Sivak and Preiss, 1995). High amylose starches have numerous industrial applications. These starches are used in fried snack products to create crispness and are widely used as thickeners or strong gelling agents in the production of jellies. Moreover, many types of photographic film also have a starch component because of the many features of high amylose starch, including transparency, flexibility, tensile strength and water resistance. But high levels of amylose do present the problem of retrogradation, which occurs when gelatinized starch recrystallizes. One way to overcome this is by the introduction of amylopectin, which will also endow starch with new properties depending on the levels of amylopectin present. Starches with high levels of amylopectin are widely used in the food industry. They improve uniformity, stability, and texture, and impart better freeze–thaw stability in frozen foods. In the paper industry, the binding and bonding properties of amylopectin are exploited to enhance paper strength and printing properties. In addition, they are added into livestock feed industry for easily to compress (Slattery *et al.*, 2000).

Amylose accounts for 0-35% of storage starch in rice endosperm. It consists of predominantly linear chains of α(l-4)-linked glucose residues. Amylose is usually branched at a low leve1 (approximately one branch per 1000 residues) by α(1-6) linkages. Amylose does not adopt a semi-crystalline structure, and is thus not amenable to X-ray diffraction analysis. It is unclear how this polysaccharide interacts with amylopectin in the granule (Ball, 1998). Once extracted from plants and in solution, amylose forms hydrogen bonds between molecules, resulting in rigid gels. However, depending on the concentration, degree of polymerization, and temperature, it may crystallize and shrink (retrogradation) after heating (Shewmaker and Stalker, 1992; Martin and Smith, 1995).

Amylopectin accounts for 65-100% of storage starch in rice endosperm and is the highly branched component of starch. It may contain up to 10^6 glucosyl residue with an average chain length of 20-24 α(1-4)-linked glucosyl units that are linked by α(1-6)-branch linkages. Thus, 4-5% of the linkages in amylopectin molecules represent branch points. Between 1000 and 3000 of these chains are linked together to form the amylopectin molecules. Amylopectin has a defined structure composed of tandem linked clusters. In the cluster structure, the A chains carry no side-chains and are linked to the other chains at their reducing end glucose units. The A chains comprise 40-60% of the total glucosyl in the molecule. The B chains carry one or more branches, mostly of the A chains. The B chains are divided into three groups according to their length. The only chain that contains a reducing terminal in an amylopectin molecule is called the C chain. The multiple cluster structure of amylopectin exhibits two regions that are repeated, the amorphous lamellae and the crystalline lamellae. The regions containing a high density of branch linkages are called amorphous lamellae. The crystalline lamellae, recognizable by their resistance to acid hydrolysis, are thought to comprise tightly associated double helices containing primarily A-chains and unsubstituted spans of B-chains. (Peat, 1952; Imberty, 1991; Dey and Harborne, 1997; Nakamura, 2002). The 9- to 10-nm thickness of the repeating unit of crystalline and amorphous lamellae is generally conserved in the plant kingdom. An average amylopectin molecule is 200 to 400 nm long (20 to 40 clusters) and approximately 15 nm wide (Myers *et al.*, 2000). Some structural

biologists believe that long amylose chains are present in regularly spaced amorphous cavities within the starch granule (Ball *et al.*, 1998).

3. Starch Synthesis

ADP-glucose pyrophosphorylase produces the activated glucosyl donor ADP-glucose (ADPG). The enzyme is now known to be expressed largely extra plastidial (i.e. 85–95% cytosolic) in the cereal endosperm. ADP-glucose synthesized in the cytosol is transported into plastids via ADP-glucose transporters. The synthesis of both amylopectin and amylose depends on the transfer of glucose in an α-1,4 position from ADP-glucose to the non-reducing end of growing chains catalyzed by starch synthases (Myers *et al.*, 2000).

The starch synthase isoform responsible for the amylose synthesis is the granule bound starch synthase (GBSS), which is encoded by the Waxy (Wx) loci in cereals. It utilizes ADP-glucose to elongate linear chains by the formation of α-1,4 linkages, and functions specifically to elongate amylose (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). SSSs utilize ADP-glucose to extend linear chains by the formation of α-1,4 linkages. SSSs are found in different combinations. For example, in vitro measurements suggest that the dominant activity in maize endosperm is SSI, in pea embryos SSSII, and in potato tubers SSSIII, even though each species have genes for all three isoforms. In wheat endosperm, SSSI, SSSII, and SSSIII all are present in the soluble phase, and further complexity is evident because the abundance of SSSII and its partitioning between the granule and soluble phases varies during grain development. Subsequently, SBEs catalyze the cleavage of α-1,4 linkages and produce a new α-1,6 linkage. DBEs play a critical role in removing the few wrong-branched chains that are linked with the C-chain in crystalline lamellae. It is difficult to remove them after the synthesis of the cluster. Finally SSSs play a role in elongating the newly formed short chains. (Li *et al.*, 1999; Myers *et al.*, 2000; Nakamura, 2002).

3.1 Starch synthase families

Starch synthase, described in 1960 by de Fekete *et al.* in 1960 (Ball *et al.*, 1998), catalyzes the transfer of a glucosyl moiety from ADP-glucose to the non-reducing end of elongating glucan chains. In higher plants, there are two types of starch synthase i.e. granule-bound starch synthase (GBSS) and soluble starch synthase (SSS). It has been assumed that GBSS functions specifically to elongate amylose, while SSS contributes to amylopectin synthesis. Hirose and Terao (2004) elucidate the roles of the isogenes encoding starch synthase in rice (*O. sativa* L.). In rice, the starch synthase family has 10 members. Multi-sequence alignment analysis of the starch synthase proteins from rice and other plant species suggested that they were grouped into five classes, soluble starch synthase I (SSSI), SSSII, SSSIII, SSSIV and granule-bound starch synthase (GBSS). The relative activities of these isoforms vary greatly depending on the plant species and tissues (Smith *et al.*, 1997).

3.2 Starch branching enzymes

Starch branching enzymes (SBEs) are the only enzymes that can introduce α-1,6-glucosidic linkages into α-polyglucans in plants (Nakamura, 2002). There are 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 (Guan and Preiss, 1993; James *et al.*, 2003). The family A consists of multiple isoforms, including maize (SBEIIa and SBEIIb), pea (SBEI), and rice (RBE3 and RBE4). In maize, SBEIIb has been detected in the endosperm only, whereas SBEIIa has been found in all organs, such as the developing endosperm, leaf blade, leaf sheath, culms, and root. The family B consists of only a single isoform in maize (SBEI), pea (SBEII), and rice (RBE1).

A phylogenetic analysis of plant SBEs constructed on the alignment of protein sequence confirms the existence of two families. The family A has three groups, the dicot family A (pea SBEI, X80009; potato SBEII, AJ000004; Arabidopsis SBE2.1, U22428; Arabidopsis 2.2, U18817), the monocot seed specific A (rice RBE3, D16210; maize SBEIIb, L08065; barley SBEIIb, AF064561), and the monocot widely

expressed family A (rice RBE4, AB023498; maize SBEIIa, U65948; barley SBEIIa, AF064560; wheat SBEII, Y11282). Also, the family B has two groups: the dicot family B (pea SBEII, X80010; potato SBEI, X69805; cassava SBEI, X77012) and the monocot family B (rice RBE1, D11082; maize SBEI, D11081; wheat SBEI, Y12320) (Mizuno *et al.*, 2001).

Rice RBE1 plays an important, but not exclusive, role in synthesis of B chains. RBE4 accounts for only about 15-20% of the total RBE activity in rice endosperm, but it plays a distinct role in the synthesis of short chains of amylopectin in the assimilatory starch, supporting the function of RBE3 and/or RBE1. RBE3 plays a distinct role in the formation of A-chains, and its role cannot be complemented by RBE4 and/or RBE1 (Nakamura, 2002). The different starch branching enzymes are expressed at different times during the development of the rice grain. RBE4 expression was detected earliest at 3 days after flowering (DAF) and maximally at 5–7 DAF. In contrast, RBE1 and RBE3 transcripts are most abundant later, at 7–10 DAF in rice (Mizuno *et al.*, 2001; Rahman *et al.*, 2001).

3.3 Starch debranching enzymes

In plants, starch debranching enzymes (DBEs) cleave α-1,6 branch linkages in amylopectin. They are divided into two types i.e. isoamylase and pullulanase. They differ in substrate specificity, i.e. isoamylase debranches glycogen, phytoglycogen and amylopectin, but scarcely attacks pullulan, whereas pullulanase can attack pullulan and amylopectin, but not glycogen and phytoglycogen. Both DBEs are involved in the synthesis of amylopectin in rice, selectively removing branches that are inappropriately positioned. Accordingly, DBEs activity would be required for maintenance of the cluster structure of amylopectin, for the dense packing of linear chains, and for growing chains to crystallize onto the granule surface (Ball *et al.*, 1996; Myers *et al.*, 2000; Nakamura, 2002; James *et al.*, 2003)

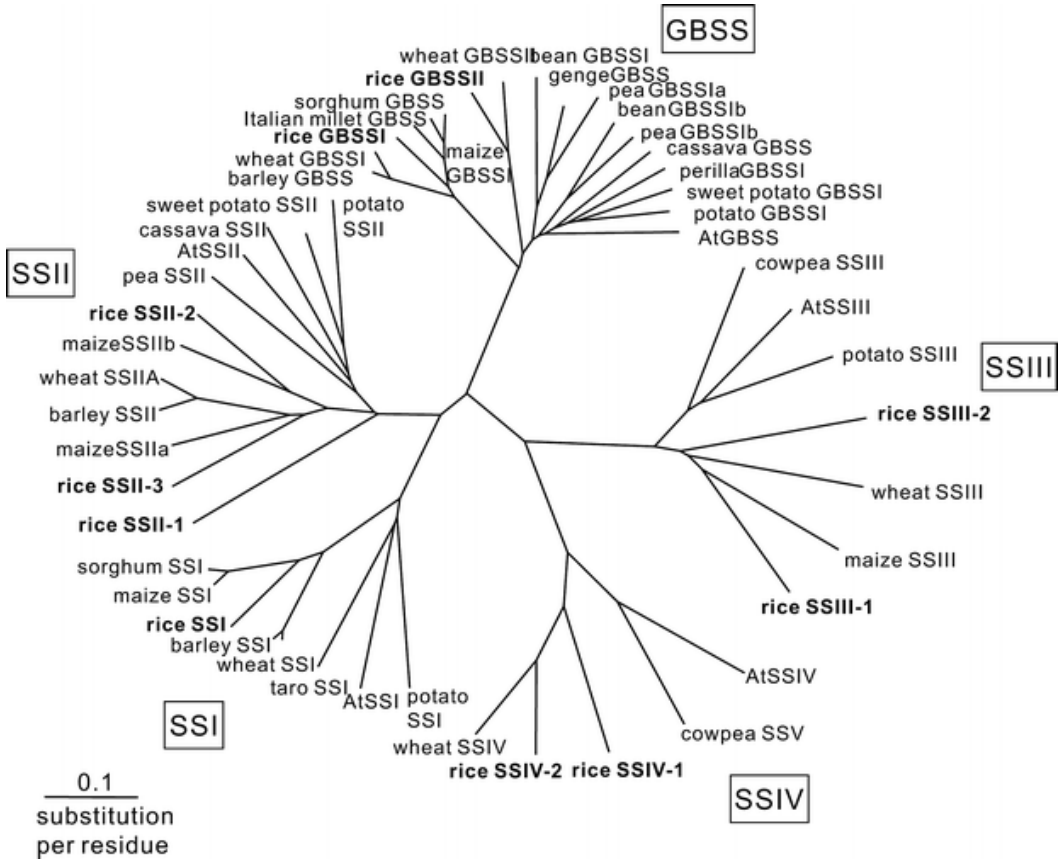


Figure 1 An un-rooted dendrogram of plant starch synthases based on their deduced amino acid sequences (Hirose and Terao, 2004).

4. Techniques to study genetic diversity

Molecular markers provide a powerful tool to study genetic diversity and evolutionary relationships and are used to assist in breeding programs. There are many molecular techniques that can provide markers.

Isozymes are variant forms of an enzyme, with small change(s) in amino acid sequence, resulting in a small change in the enzyme's electric charge. Polymorphisms have been found in isozyme alleles, and many isozyme loci have been identified in various plant species. The analysis of isozyme genes has greatly contributed to advances in different fields of biological research such as gene expression and regulation, evolutionary genetics, population genetics, and developmental genetics. They are very useful as genetic markers (Morishima, 1997), but because of the limited

number of loci and alleles that can be detected they have been largely superseded by DNA based markers. In rice, the intraspecific variation of *O. sativa* has been investigated on a large scale by isozyme analysis (Glaszmann, 1987).

Restriction fragment length polymorphism (RFLP) has been intensively used for the investigation of genetic diversity and relationships of crop species and their wild relatives. As a result, genetic diversity and evolution patterns have been illustrated in many crops such as wheat, barley, and rice (Wang *et al.*, 1992; Zhang *et al.*, 1992; Salvo-Garrido *et al.*, 2001; Sun *et al.*, 2001). In rice, RFLP markers have been used to construct a genomic map (Causse *et al.*, 1994; Mackill, 1999). Nuclear RFLP data were used to study genetic differentiation among landraces of rice in northern Vietnam, which lies in the center of genetic diversity of Asian cultivated rice (Fukuoka *et al.*, 2003).

Markers based on the polymerase chain reaction (PCR) have been developed to overcome some of the disadvantages of RFLP. This opened up the potential to assay markers with DNA extracted from small amounts of tissue. Random amplified polymorphic DNA (RAPD) (Williams *et al.*, 1990) was the first PCR-based marker to be widely applied in germplasm characterization (Mackill, 1999). Martin *et al.* (1997) used RAPD markers to facilitate the identification of *Oryza* species within a germplasm collection. The results indicated that RAPD technology can be used as a fast and accurate method to assist in the validation of the identification of wild *Oryza* species.

More recently, the amplified fragment length polymorphism (AFLP) assay (Vos *et al.*, 1995) has been developed. Similar to RAPDs it randomly amplifies genomic DNA fragments, but more bands are obtained per reaction and it is supposedly more reproducible. AFLP has been used to study genetic diversity in rice such as genetic diversity in cultivars and landraces of *O. sativa* subsp. indica (Prashanth *et al.*, 2002).

Microsatellite markers consist of short (2-6 bp) DNA sequences that are tandemly repeated from 10 to more than 30 times. These markers, sometimes referred to as simple sequence repeat (SSR) or simple sequence length polymorphism (SSLP), can be amplified by PCR with primers flanking the repeats. Polymorphisms are detected as length variations of fragments, usually on polyacrylamide gels. Initial studies in rice indicated that these markers were highly polymorphic. While, RFLPs usually only exhibit 2 or 3 alleles per locus, as many as 25 alleles have been observed at a single microsatellite locus among a diverse set of cultivars (Yang *et al.*, 1994). Microsatellite markers have since become one of the markers of choice for genetic characterization of rice cultivars (Mackill, 1999).

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). Olsen and Purugganan (2002) examined an allele genealogy of the *Waxy* locus to trace the evolutionary and geographical origins of this phenotype and found evidence that the mutation responsible for the waxy phenotype has a single evolutionary origin and that it probably arose in Southeast Asia. In addition, they found that many nonglutinous varieties in Northeast Asia also carry the mutation at the same site, suggesting that partial suppression of this mutation may have played an important role in the development of Northeast Asian nonglutinous rice.

In 1989, Orita *et al.* developed a mobility shift analysis of single-stranded DNAs using non-denaturing polyacrylamide gel electrophoresis to detect DNA polymorphism. In this technique, nucleotide substitutions might lead to a conformational change of single-stranded DNA molecules and result in a change of the electrophoretic mobility. SSCPs were found to be allelic variants of true Mendelian traits, and therefore they should be useful genetic markers.

For rapid and large-scale detection, the discovered allelic diversity can be converted into SNP assays. SNPs are markers that detect alternative nucleotides at a specific position in a DNA sequence. There are usually only two types at a variable

nucleotide position, and thus SNPs are generally bi-allelic. These polymorphisms are very abundant in the genome of plants and animals. SNP in the *waxy* gene have been used to study evolution in rice (Wang *et al.*, 1995; Isshiki *et al.*, 1998; Yamanaka *et al.*, 2004).

The study of crop origins can provide unique insights into the evolution of morphological and developmental adaptations favored by early farming cultures. Phylogeographic analysis of allele genealogies has proved useful for examining crop origins (Olsen and Schaal, 1999; Sanjur *et al*., 2002). When applied to crop systems, this approach typically uses noncoding, selectively neutral genetic variation to draw inferences about the history of population divergence and crop-progenitor relationships. A potential extension of the phylogeographic method is to examine genes directly responsible for the phenotypic variation within a crop species. This candidate-locus approach may provide a means of inferring the origin and dispersal of specific traits that have been favored over the course of domestication (Olsen and Purugganan, 2002).

5. Association analysis

Association mapping, also referred to as linkage disequilibrium (LD) mapping, is a popular method that has been widely used to test the relationship between specific sequence polymorphisms in candidate genes and phenotypic variation (Thornsberry *et al*., 2001; Flint-Garcia *et al*., 2003; Gupta *et al.*, 2005), facilitating identification of actual functional polymorphisms within candidate genes and epistatic interactions (Caicedo *et al.*, 2004; Hagenblad *et al.*, 2004). LD studies conducted in plants have recently been summarized by Gupta *et al.* (2005), and examples using association mapping are still increasing (Bundock and Henry 2004; Simko *et al.*, 2004; Andersen *et al.*, 2005; Szalma *et al.*, 2005).

MATERIALS AND METHODS

1. Plant materials

170 varieties of *O. sativa*, *O. rufipogon*, and *O. nivara* were sampled from the Biotechnology Research and Development Office, Patum Thanee. These accessions represented the germplasm from all parts of Thailand, ecotypes and cooking types. In addition, 3 varieties (KDML105, RD6 and Nipponbare) were obtained from the DNA Technology Laboratory, BIOTEC, and 19 accessions of the genus *Oryza* such as *O. sativa*, *O. nivara*, *O. rufipogon* and *O. glaberrima* were sampled from the germplasm collection maintained by the International Rice Research Institute (IRRI, Los Baños, Philippines).

2. Evaluation of eating and cooking quality

2.1 Determination of amylose content

The apparent amylose content (AAC) in starch granules was determined using a near-infrared reflectance spectroscopy method based on an iodine colorimetric assay (Juliano, 1985). Serial dilutions of purified amylose from potato were used as standards. Twenty milligrams of starch was gelatinized by treatment with 0.5 mL of 95% ethanol and 4.5 mL of 1 M NaOH and stood for 24 hr. at room temperature. After the addition of distilled water up to 50 mL, the solution was homogenized. An aliquot (2.5 mL) of the solution was taken and 35 mL of distilled water, 0.5 mL of 1 M CH_3COOH, and 0.2 mL of 0.2% (w/v) I_2, 2% (w/v) KI solution were added. After adjusting the volume to 50 mL with distilled water, the solution was homogenized and stood for 20 min. at room temperature. The absorbance at 680 nm (the blue value) and at the wavelength of maximal absorbance (λ-max) were measured. The apparent amylose content was calculated using a calibration line obtained from the blue value at 680 nm by serial dilutions of purified amylose from potato in the iodine solution.

2.2 Determination of gelatinization temperature

The GT was measured on the basis of individual grains and expressed as the alkali spread value (ASV) using the method of Little *et al.* (1958) with minor modifications. Ten intact milled grains from each accession were incubated in 15 mL of 1.7% KOH. The grains were carefully separated from each other using a forceps and incubated at room temperature for 23 hr. to allow partial dissolution of the grains. The spreading value of the grains was measured using the following seven-point semi-quantitative rating scale: 1, grain not affected; 2, grain swollen; 3, grain swollen, collar incomplete and narrow; 4, grain swollen, collar complete and wide; 5, grain split or segmented, collar complete and wide; 6, grain dispersed, merging with collar and 7, grain completely dispersed and intermingled. ASV corresponding to GT are as follows: 1-2, high (GT: 74-80°C); 3-5, intermediate (GT: 70-74°C), and 6-7, low (GT < 70°C).

2.3 Determination of gel consistency

The GC was measured in duplicates according to the method of Cagampang *et al.* (1973). Briefly, 100 mg rice flour was weighed in a 13 mm. x 150 mm. culture tube, to which 0.2 mL of 95% ethanol containing 0.025% Bromthymol blue was added to prevent clumping of the powder during gelatinization. Two milliliter of 0.2 N KOH was added and the solution was vortexed thoroughly. The tubes were covered with parafilm and boiled vigorously in a water bath for 8 min. After standing at room temperature for 5 min, the tubes were put on ice for 20 min, and then laid down horizontally on a table surface. The gel spreading length was measured 1 h later as the distance from the bottom of the tube to the front of the gel migration. The gel length thus obtained provides a measurement of the gel consistency: the longer the spreading distance, the softer the gel.

3. DNA extraction

Genomic DNA was extracted from 5-days-old seedlings using MATAB method, which was similar the CTAB method by Agrawal (1992) but CTAB was

substituted for by MATAB in the extraction buffer. Briefly, the procedure consists of grinding 0.1 g of leaf tissue with a mortar and pestle under liquid nitrogen and transferring the frozen powdered tissue to 0.7 mL isolation buffer (2% MATAB, 1.4 M NaCl, 0.2% 2-mercaptoethanol, 20 mM EDTA, 100 mM Tris-HCl, pH 8.0 with 2 µL of RNaseA added) preheated at 60°C in a water bath. The samples were incubated for 30 minutes at 60°C with occasional gentle swirling. After this they were kept at room temperature for 5 min followed by a single extraction with 0.7 mL of chloroform-isoamyl alcohol (24:1). Cell debris was precipitated by centrifugation at 10,000 *g* for 10 min and the aqueous phase transfered to a new centrifuge tube using a wide-bore pipet tip. Nucleic acids were precipitated by adding 2/3 volumes of cold isopropanol and gently mixing. The DNA was pelleted by centrifugation at 10,000 *g* for 5 min and the pellet was washed with cold 70% ethanol and air-dried. Finally, the pellet was dissolved with 0.1 mL of TE buffer.

4. Selection of PCR primers

BLAST was used to retrieve *O. sativa* (japonica and indica) sequences corresponding to each of the starch synthesis genes 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 is preferred because it contains the information for transcription initiation and is a major site through which gene expression is controlled. In addition, as a non-protein-coding sequence it is expected to be more variable than the protein coding parts of the gene.

5. PCR-SSCP

DNA fragments were PCR-amplified from genomic DNA using specific primers. Details of all primers are presented in Table 2. Each PCR amplification reaction (15 µL) contained 20 ng of template DNA, 10 pmoles 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^{TM}). Cycling started with an initial denaturation at 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 2 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 on 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), 3.5 µL of the aliquots were loaded and run in 0.6X TBE buffer at constant 10 watt for 14-20 hr. in a 4 °C refrigerator.

After the electrophoresis, silver staining was used to reveal the DNA fragments. The gel on glass plates was covered with fixation solution (10% acetic acid) and shaken gently on an orbital shaker for 20 min, then washed 3 times in deionized water, shaken gently 10 min each time. Silver staining was carried out for 30 min on a shaker using 1% silver nitrate (Fisher scientific Ltd., UK) with 1.5 mL/L of 37% formaldehyde. Chilled developing solution, containing 30 g/L sodium bicarbonate, 1.5 mL/L of 37% formaldehyde, and 2 mg/L sodium thiosulphate, was used to develop the bands. The developing process was stopped by 10% acetic acid and the gels were washed thoroughly with deionized water twice for 2 min each time, then left to air dry (Bassam ***et al.***, 1991; Bassam and Caetano-Anollés, 1993).

6. 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 or more representative individuals of each type were selected for sequencing. PCR

amplified DNA fragments were sent for direct sequencing using specific internal sequencing primers. The PCR amplification for direct sequencing followed the same conditions except that primer and dNTP concentrations were reduced 15% and the volume was increased to 50 µL.

7. ORF100 locus

The ORF100 locus was amplified using specific primer sets following the method from Chuayjaeng (2005). The size of the amplified products was checked by electrophoresis on 1% agarose gel with TAE buffer and scored for the differences in size of the amplified fragment.

8. Statistical and Computer Analysis

8.1 Haplotypes and eating-cooking quality

The result was analyzed using the SPSS 15.0 (Computer Pro System Corp, USA). Correlation analysis was conducted to characterize the relationships between AAC, GT and GC and expressed by the Pearson's correlation coefficient (r). The analysis of variance (ANOVA) of the starch physico-chemical properties and haplotypes at different gene loci was performed. Tukey's multiple range tests were conducted for comparison between means of each of the starch physico-chemical properties and each gene locus with $P < 0.05$.

8.2 Evaluation of DNA sequence diversity

Dna-SP Version 4.0 software (Rozas *et al.*, 2003) 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 neutrality of the SNP polymorphism (Tajima, 1989).

8.3 Network reconstruction

Allele networks were reconstructed using the statistical parsimony algorithm (Templeton *et al*., 1992) employed by the program TCS 1.21 (Clement *et al.*, 2000). With this method, unrooted cladograms that have a high probability (>95%) of being true based on a finite-site model of DNA evolution are identified. Each indel, regardless its size, was counted as one mutational event and treated as fifth state. Nucleotide substitutions within indels were treated as additional events.

8.4 Linkage disequilibrium

LD between pairs of sites (both SNPs and indels) was evaluated by using TASSEL Version 2.0.1 (http://www.maizegenetics.net/bioinformatics/tasselindex.htm). LD was estimated by squared allele-frequency correlations (r^2), and standardized disequilibrium coefficients (*D*'). The loci were considered to be in significant LD if $P < 0.001$.

8.5 Haplotype block structure

The plots of LD (r^2) for pairs of loci were drawn from r^2 values calculated by TASSEL. The graph displays the LD between all possible pairs of sites. The default setting was used.

8.6 Population structure

Haplotypes from the seven gene sequences were used for assessment of population structure. The program STRUCTURE version 2.1 (Pritchard *et al.*, 2000) was used to test the hypotheses for one to twelve subpopulations, with admixture model and with correlated allele frequencies (Falush *et al.*, 2003) a burn-in phase of 100,000 iterations, and a sampling of 1,000,000 replicates. At least 5 repetitions were

conducted for each population size. The Q matrix from the optimal result was extracted for later use.

8.7 Association analysis

The general linear model was used in the analysis of associations between nucleotide polymorphisms or marker alleles and different eating-cooking qualities, which was performed with TASSEL Version 2.0.1 software.

9. Gene expression analysis

The level of expression of starch synthesis genes in the endosperm and leaf was checked by RT-PCR. Total RNA was extracted from the endosperm and leaves harvested between 5 to 10 days after flowering using RNAeasy kit (Qiagen). The first-strand cDNA synthesis was done using Omniscript Reverse Transcriptase (Qiagen), and fragments of each gene were PCR amplified with specific primer sets. The fragments were checked by agarose gel electrophoresis, and the expression of each gene in endosperm and leaf was compared.

RESULTS AND DISCUSSION

1. Diversity of starch properties in Thai rice

Across the 160 accessions of *O. sativa* from Thailand, the apparent amylose content (AAC) varied from 1.11% to 30.07% including 59 accessions of waxy rice (36.88% of total) and 101 accessions of non-waxy rice i.e. 31 accessions of intermediate amylose content (19.37% of total) and 70 accessions of high amylose content (43.75% of total) (Figure 2). The gelatinization temperature (GT) alkali score varied from score 2 (GT: 74-80°C) to score 7 (GT < 70°C) including one accession of high GT (74-80°C, score 2), 52 accessions of intermediate GT (70-74°C, score 3-5) and 107 accessions of low GT (< 70°C, score 6-7) (Figure 3). The gel consistency (GC) varied from 1 to 15 cm. including 33 accessions of high GC values (>6 cm.), 34 accessions of medium GC values (4.1 - 6 cm.) and 93 accessions of low GC values (<4 cm.) (Figure 4).

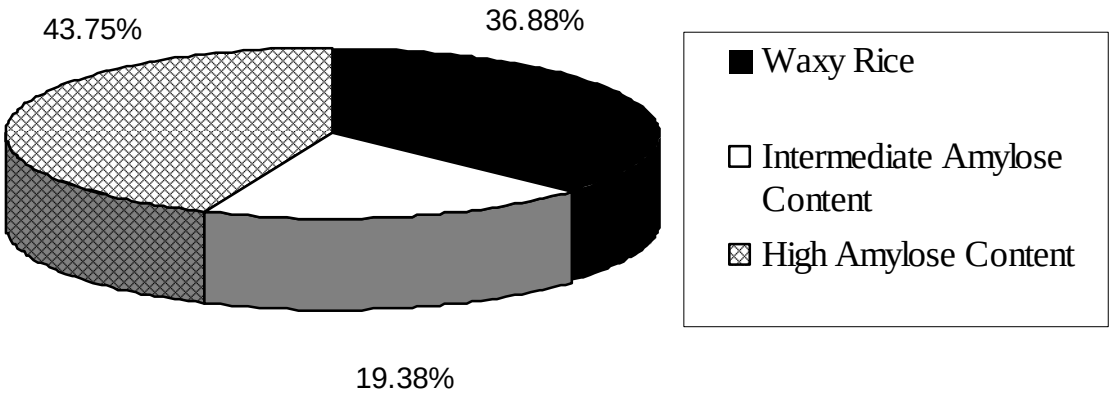


Figure 2 Distribution of the amylose content for the 160 accessions of *O. sativa* from Thailand.

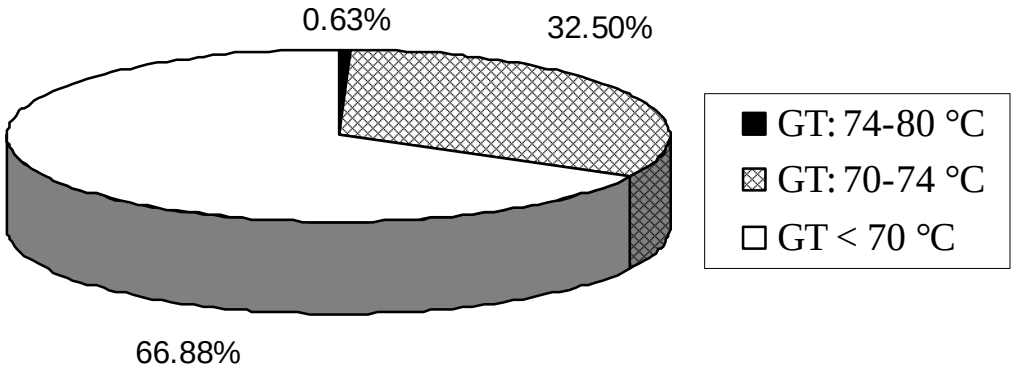


Figure 3 Distribution of the gelatinization temperature for the 160 accessions of *O. sativa* from Thailand.

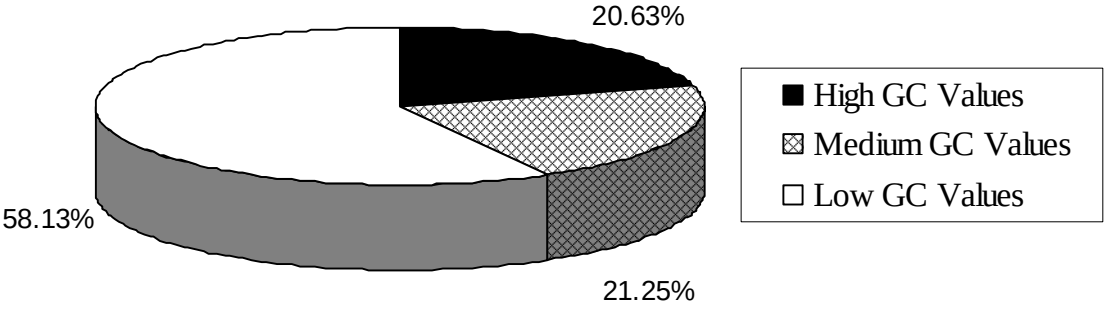


Figure 4 Distribution of the gel consistency for the 160 accessions of *O. sativa* from Thailand.

Statistical analyses were performed to determine whether there were correlations among the starch property measures. The Pearson correlation coefficient (*r*) relating AAC to GC is -0.772 (Figure 5). Likewise an inverse correlation is suggested between AAC and GT (Figure 6) (r = -0.467). The GC measure shows a weak but positive correlation with the GT score (r = 0.325) (Figure 7). The statistical analysis indicates that all correlations are highly significant (Table 1).

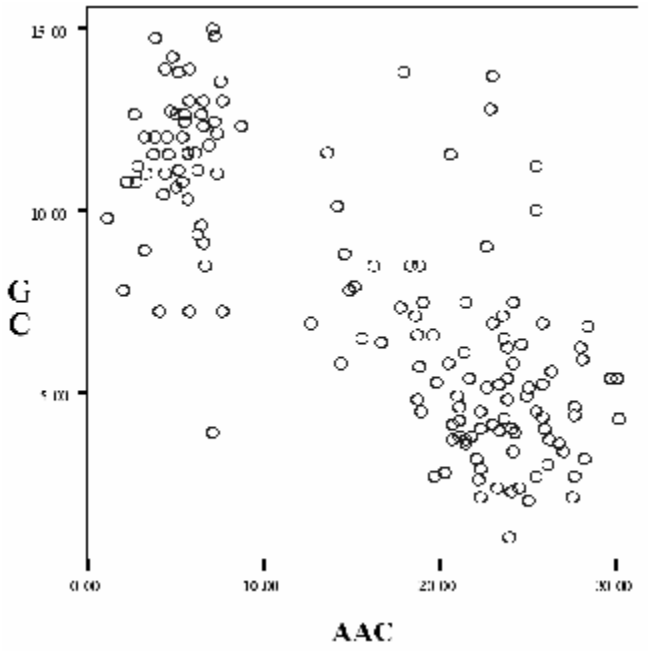


Figure 5 Scattergram showing an inverse correlation between AAC and GC.

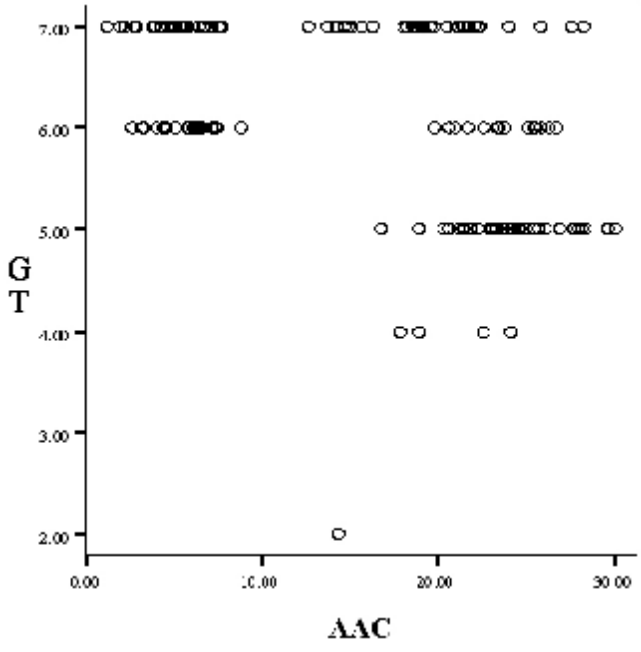


Figure 6 Scattergram showing an inverse correlation between AAC and GT.

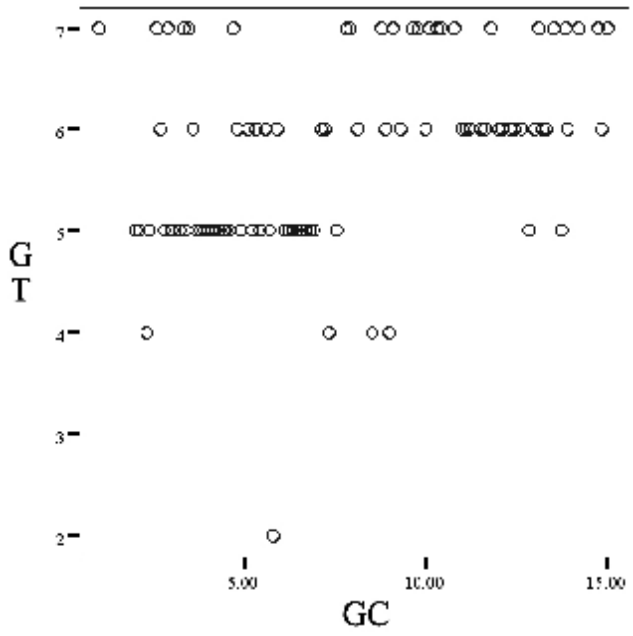


Figure 7 Scattergram showing the correlation between GC and GT.

Table 1 The correlation coefficients among AAC, GC and GT for Thai rice varieties.

		AAC	GC	GT
AAC	Pearson Correlation	1	-0.467**	-0.772**
	Sig. (2-tailed)		0.000	0.000
	N	160	160	160
GC	Pearson Correlation	-0.467**	1	0.325**
	Sig. (2-tailed)	0.000		0.000
	N	160	160	160
GT	Pearson Correlation	-0.772**	0.325**	1
	Sig. (2-tailed)	0.000	0.000	
	N	160	160	160

** Significant at $P < 0.01$

2. The ORF100 locus

The primers for the ORF100 locus (Open reading frame 100) are located in the chloroplast tRNA-Ser (GCU) gene and the psbD gene. The intergenic region between these two genes is amplified and a DNA fragment of approximately 400 bp is obtained. On agarose gel, two different alleles could be distinguished. As previously studied, a deletion of 69 bp in the ORF100 region clearly separated most *indica* type cultivars from most *japonica* cultivars (Kanno *et al.*, 1993; Nakamura *et al.*, 1998; Sun *et al.*, 2002). However, several *indica* cultivars showed the non-deletion haplotype and a few *japonica* cultivars showed the *indica* deletion haplotype. In wild rice the deletion haplotype was found only in some *O. rufipogon* and *O. nivara*, close relatives of Asian cultivated rice. Other accessions of *O. rufipogon* and *O. nivara*, other species with the AA genome and all species with different genomes (BB, CC, EE, GG, BBCC, and CCDD) had the non-deletion haplotype (Chuayjaeng, 2005).

3. Selection of PCR primers

For all genes, *Oryza* DNA sequences were retrieved from the NCBI (http://ncbi.nlm.nih.gov/) database using keyword searches. Nucleotide sequences were first aligned with the GeneDoc program version 2.6.002 (Nicholas and Nicholas 1997) and manually edited. Primers were designed using Fast PCR version 3.2.130 (Kalender, R. Institute of Biotechnology, University of Helsinki, Finland).The primer sequences are listed in Table 2.

Table 2 Gene-specific PCR primers used for SSCP technique.

Gene	Accession. No.	Chromo-some	Position		Forward primer (5'-3')	Reverse primer (5'-3')
GBSSI	AB002542	6	2919	3480	ACCATTCCTTCAgTTCTTTgTC	gTTTCTCCAgTggCgAgAg
SSSIIIa	AF432915	4	219	828	TCCTAAAAgCTgggCCAATg	CggTggATCggCATCTCTC
SSSIIIb	AP005441	8	1900	2516	AAAATAACCTACATATTTCAAAACAgC	TAgCTTAgCTTCATCCgTCgCATC
SSSIVa	AP003292	1	812	1414	TTggTTgTgAAACCgTgAAAgC	CggCCCACTCTgACTTTgg
SSSIVb	AC121365	5	1371 Intron	1865 Exon2	TCTCAgTAgTCTgCTCCTgC	TCACTggAAACAgATgCTTC
RBE1	D10838	6	3004	3586 Exon2	AgTgTCAgCATAgAAATCTC	gAAACCACgCTCAggCgAAC
RBE3	AP004879	2	607	1169 Exon2	CCCTCCgCTCCTCCTAgCTTC	TCgCCCTCggggATCATCAC

4. PCR-SSCP and evaluation of DNA sequence diversity

One hundred and ninety-two 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). The number of SSCP patterns that could be distinguished is listed in Table 3. 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 that were selected to be sequenced.

The total length of the aligned sequences for the seven genes is 3,975 bp, including 444 bp of coding sequence and 3,531 bp of noncoding sequence. The number of insertion–deletion polymorphisms ranged from 0 to 9 across loci, with a total of 32 indel polymorphisms (Table 3). A detailed examination of these indel polymorphisms showed that all of them occurred in noncoding regions, with 7 (38.86%) being 1-bp indels. The remaining indels included two each of 3 and 7-bp, and each one of 5, 6, 20, 22, 30, 32 and 34-bp. The average estimates of variation calculated with DnaSP Version 4.0 (Rozas *et al.*, 2003) over seven loci were comparable. The π ranged from 0.00086 (RBE3) to 0.02821 (SIV1) and θ ranged from 0.00154 (RBE3) to 0.02385 (SIV1). The levels of polymorphism within all accession and only *O. sativa* from Thailand were not different (Table 4).

Table 3 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 4 Summary of nucleotide polymorphisms and neutrality tests calculated with the DnaSP program*.

Locus	N^a	S^b	π^c	θ^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

Table 4 (Continued)

Locus	N^a	S^b	π^c	θ^d	D^e
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 pairwise nucleotide differences per site (Nei, 1987) calculated on the total number of polymorphic sites

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

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

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

4.1 GBSS locus

The GBSSI gene codes for the granule bound starch synthase (GBSS). The forward and reverse primers for the GBSSI locus amplified a fragment from base 2919 to base 3480 (according to the sequence AB002542), which is part of the non-coding 5' upstream region. The resulting PCR product is approximately 550 bp long. Different alleles could not be identified on agarose gel. The PCR products were separated on non-denaturing polyacylamide gel using the SSCP technique (Orita *et*

al., 1989). Sixteen SSCP patterns were identified (Appendix Figure 1) 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 (Appendix Figure 2). Nucleotide polymorphisms were found at twenty-three positions including nineteen substitutions, two indels and two SSRs (Table 3). 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, Watterson's estimator (θ) was 0.01017, and Tajima's D = 1.97801 ($P > 0.05$), indicating no significant Tajima's D value, where the neutral model was rejected at $P \geq 0.05$ (Table 4).

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 5. 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 the present study, of which $(CT)_{17}$ was the most frequent. Allele $(CT)_{14}$ and $(CT)_{20}$ were not found in the present study. Four new alleles were detected in this study, such as $(CT)_7$, $(CT)_9$, $(CT)_{10}$ and $(CT)_{12}$, that were not detected in previous reports (Ayres *et al.*, 1997; Bao *et al.*, 2002, 2006). However, only a very limited number (three) of waxy rice accessions were included in the work of Ayres *et al.* (1997) and a low number (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.* (2006) concluded that the alleles with fewer repeats ($n \leq 12$) were highly associated with higher apparent amylose content (AAC) and those with more repeats are 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 can be used for marker-assisted selection (Ayres *et al.*, 1997; Bao *et al.*, 2006). However, in the present study, the alleles with fewer repeats ($n \leq 12$) excluding $(CT)_{10}$ were associated with a lower AAC. The result of relationship between allele polymorphism and waxy-rice were thus not conclusive because there was conflict between results from present and previous studies. So, the correlation between (CT) repeat length and amylose content found in previous studies cannot be extended to rice diversity in Thailand.

Table 5 Diversity of the $(CT)_n$ repeat at the GBSSI locus.

$(CT)_n$	Haplotype of flanking sequence	The present 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

4.2 SSSIIIa locus

The SSSIIIa gene codes for the soluble starch synthase III. (SSSIII) preferentially synthesizes long B1 and B2 chains in amylopectin synthesis. In rice endosperms, SSSIIIa was expressed at an early developing stage (Dian *et al.*, 2005). The forward and reverse primers for the SSSIIIa locus span the 5' upstream region. On agarose gel, a single band of approximately 600 bp was visible, but different alleles could not be distinguished. The PCR products were separated on non-denaturing polyacylamide gel using the SSCP technique. The SSCP assay revealed twenty-one patterns (Appendix Figure 3). 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 (Appendix Figure 4). Nucleotide polymorphisms were found at thirty-six positions: eight indels and twenty-eight substitutions (Table 3). 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 4).

4.3 SSSIIIb locus

The SSSIIIb gene codes for soluble starch synthase III. In rice endosperms, SSSIIIb reached maximum expression at the middle stage (Dian *et al.*, 2005). The forward and reverse primers for the SSSIIIb locus span the 5' upstream region and positioned from base 1900 to base 2516 according to the sequence AP005441. A PCR fragment of approximately 600 bp could be seen on agarose gel but different alleles could not be. The PCR products were separated on non-denaturing polyacylamide gel under SSCP conditions. Ten SSCP patterns were observed (Appendix Figure 5) 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 (Appendix Figure 6). Nucleotide polymorphisms were found at ten positions viz. one indel and nine substitutions (Table 3). 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 arises from an excess of low-frequency nucleotide polymorphisms (Table 4).

4.4 SSSIVa locus

The SSSIVa gene encoded functional starch synthase enzyme and were expressed in a variety of tissues, including roots and germinating seeds (Hirose and Terao, 2004). The forward and reverse primers for the SSSIVa locus span the 5' upstream region and positioned from base 812 to base 1414 according to the sequence AP003292. A PCR fragment of approximate size 600 bp could be seen on agarose gel but different alleles could not be distinguished. The fragments were separated on non-denaturing polyacylamide gel using the SSCP technique. Fourteen SSCP patterns could be identified (Appendix Figure 7). 25 PCR products were 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 (Appendix Figure 8). Nucleotide polymorphisms were found at sixty-nine positions, eight indels and sixty-one substitutions (Table 3). 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 4).

4.5 SSSIVb locus

The SSSIVb gene codes for a functional starch synthase enzyme, but it is a different isoform of SSSIVa gene (Hirose and Terao, 2004). The primers for the SSSIVb marker amplified from intron1 to exon2 and positioned from base 1371 (intron1) to base 1865 (exon2) according to the sequence AP003292. On agarose gel, a single band of approximately 400 bp was visible, but different alleles could not be distinguished. The PCR products were separated on non-denaturing polyacylamide gel using the SSCP technique. The SSCP assay revealed nine patterns (Appendix Figure 9). 22 PCR products were 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 (Appendix Figure 10). Nucleotide polymorphisms were found at eight positions including six substitutions, one indel and one SSR (Table 3). 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). Nucleotide diversity indices of the non-coding region (π = 0.00279) was higher than those of the coding region (π = 0.00189), however, the Watterson's estimate of the non-coding region (θ = 0.00323) was lower than those of the coding region (θ = 0.00434). But separate Tajima's D test for each region revealed no significant departure from the neutral expectation (-0.29635 for non-coding region and -0.58161 for coding region, P > 0.05) (Table 4).

4.6 RBE1 locus

The RBE1 gene codes for starch branching enzymes (SBEs) that can introduce α-1,6-glucosidic linkages into α-polyglucans in plants. RBE1 plays an important, but not exclusive, role in synthesis of B chains (Nakamura, 2002). The forward primer for RBE1 locus was located in the 5' upstream region of RBE1 gene and the reverse primer in the second exon and positioned from base 3004 (5' upstream region) to base 3586 (exon2) according to the sequence D10838. The complete exon1 and intron1 were amplified. The resulting PCR product is approximately 550 bp long. Different alleles could not be identified on agarose gel. The PCR products were separated on non-denaturing polyacylamide gel under SSCP conditions. Ten SSCP patterns were observed (Appendix Figure 11) 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 (Appendix Figure 12). Nucleotide polymorphisms were found at thirteen positions viz. four indels, eight substitutions and one SSR (Table 3). 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 (π = 0.00858 and θ = 0.01523) were higher than the 5' upstream region (π = 0.00104 and θ = 0.00344), whereas no polymorphisms were observed in the exon1 and exon2. Separate Tajima's *D* test for intron1 region and 5' upstream region revealed no significant departure from the neutral expectation (-1.00242 for the intron1 region and -1.49545 for the 5' upstream region, $P > 0.05$) (Table 4).

4.7 RBE3 locus

The RBE3 gene codes for a starch branching enzyme similar to RBE1, but RBE3 plays an important distinct role in the formation of A-chains, and its role cannot be complemented by RBE4 and/or RBE1 (Nakamura, 2002). The forward primer for RBE3 locus spans the 5' upstream region of RBE3 gene and the reverse primer is located in the second exon and positioned from base 607 (5' upstream region) to base 1169 (exon2) according to the sequence AP004879. The complete exon1 and intron1 were amplified. The resulting PCR product is approximately 550 bp long. Different alleles could not be identified on agarose gel. The PCR products were separated on non-denaturing polyacylamide gel under SSCP conditions. Six SSCP patterns were observed (Appendix Figure 13). Nine PCR products were sent for direct sequencing. The sequences were submitted to GenBank (accession numbers EF990877-EF990882) and the length in all sequences was 557 bp (Appendix Figure 14). Nucleotide polymorphisms were found at five positions, and all of them were substitutions (Table 3). Nucleotide diversity (π) was 0.00086, Watterson's estimator (θ) was 0.00154, and Tajima's D = -0.85901 ($P > 0.05$). The diversity indices of the non-coding regions (π = 0.00113 and θ = 0.00161) were higher than those of the coding region (π = 0.00048 and θ = 0.00146). Within the non-coding regions, the nucleotide diversity indices of the 5' upstream region (π = 0.00205) was higher than those of the intron1 region (π = 0.00014), however, the Watterson's estimate of the 5' upstream region (θ = 0.00103) was lower than those of the intron1 region (θ = 0.00223). Also here, Tajima's *D* test for each region revealed no significant departure from the neutral expectation (-0.47651 for the non-coding 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 4).

A large number of insertion/deletion (InDels) and single nucleotide polymorphisms (SNPs) together with four SSRs were observed in the sequenced regions. The non-coding regions including 5' upstream regions and introns have more polymorphisms than protein coding regions. It was also reported from other species that the noncoding sequences have a much higher divergence than the coding region sequence (Tang *et al.*, 2004). The frequency of nucleotide polymorphism in this study is about one SNP per 29 bp. In the coding regions, the frequency is one SNP per 148 bp, whereas in the noncoding regions the frequency increases to one SNP per 26 bp. This is 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 in different genomic regions and in different samples (Bao *et al.*, 2006).

To test the standard neutral equilibrium model, we performed the Tajima's *D* test (Tajima, 1989) as implemented in DnaSP to address whether the data show evidence that selection has shaped levels of variation for 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 4). 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 low-frequency variant (Zhu *et al.*, 2007).

5. Haplotype networks

Networks were constructed from all obtained sequences by the statistical parsimony algorithm (Templeton *et al.*, 1992) generated by the program TCS Version 1.13 (Clement *et al.*, 2000). With this method, unrooted cladograms that have a high probability (>95%) of being true based on a finite-site model of DNA evolution are identified. Each indel, regardless its size, was counted as one mutational event and

treated as fifth state. Nucleotide substitutions within indels were treated as additional events.

5.1 GBSSI locus

The resulting network of GBSSI is almost completely resolved (Figure 8). About network, lines connecting haplotypes were equivalent to one mutational difference, with empty nodes representing haplotypes were not found in the population. Only one loop indicative of ambiguous connection remains. The biggest outgroup probability is haplotype E including *O. sativa* from Thailand, Myanmar and Indonesia.

The association of haplotypes apparently corresponds more to the planting types (including the floating rice, the irrigation rice, and the upland rice), GT, and GC rather than sub-species or geographical location. For planting types, the upland rice group including haplotype A, C, F, and I separated from other by transversion mutation (T to G) at position-142. Within this group, the haplotype A, C, and F separated from haplotype I by $(CT)_n$ at position-49. However, there is some upland rice in another haplotype. For GC types, two haplotypes showed medium gelatinize temperature are the haplotype L and K. This group separated from other groups by indel 2-bp (TA) at position 228, but there is some medium gelatinize temperature rice in another haplotype. For GC types, there are two groups of the hard and hard-medium gel consistency rice. The first group, including haplotype E and N, separated from other groups by transversion mutation (C to A) at position-37 and transversion mutation (T to A) at position-43. The second group, including haplotype G and J, separated from other groups by transition mutation (T to C) at position-493. However, there is some hard and hard-medium gel consistency rice in the other haplotype.

5.2 SSSIIIa locus

The network resulting of SSSIIIa is shown in Figure 9. The biggest outgroup probability is haplotype A including *O. nivara*, *O. rufipogon* and *O. sativa* from Thailand and Indonesia. The association of haplotypes apparently did not

correspond to the eating and cooking qualities same as species and geographical location

5.3 SSSIIIb locus

The network resulting of SSSIIIb is shown in Figure 10. The biggest outgroup probability is haplotype H including *O. nivara* from Thailand, *O. rufipogon* from Thailand and Malaysia, 2 accessions of *O. sativa* from Lower southern Thailand and *O. sativa* from Sri Lanka, Pakistan, Bangladesh, Philippines and Indonesia. The association of haplotypes apparently did not correspond to the eating and cooking quality, geographical location, and the planting types. However, the wild rice found in the haplotype B, E, H, and I (excluding the haplotype J (*O. glaberrima*). These haplotypes located in the middle of network so the common ancestor of cultivated rice and its wild ancestors could be in here. Maybe the haplotype E is the common ancestor of all *O. sativa* because it links to *O. glaberrima*.

5.4 SSSIVa locus

The network resulting of SSSIVa is shown in Figure 11. The haplotype T (*O. glaberrima*) could not amplified fragment by PCR with SSSIVa primers. The association of haplotypes apparently did not correspond to the eating-cooking quality. The haplotype F could be the common ancestor in this network. The association of haplotype D and L apparently correspond to the planting types, being upland rice. All *O. sativa* from Thailand of the haplotype D were upland rice and four *O. sativa* from other countries did not have data. For the haplotype L, two from fifteen accession of *O. sativa* from Thailand were irrigation rice.

5.5 SSSIVb locus

The network resulting of SSSIVb from this procedure is relatively simple (Figure 12). The biggest outgroup probability is haplotype F including Thai rice from Eastern, RD6 and *O. nivara* from Lao. The first subgroup were the haplotype A, C, F, and H, whereas the second subgroup were the haplotype B, D, E, and I.

The association of the second subgroup apparently corresponds to the planting types, being upland rice. The haplotype D and B were upland rice from the North part of Thailand, but differed in starch type. The haplotype D and B were sticky rice and non-sticky rice, respectively, and separated from haplotype I by transition mutation (T to C) at position-118 and 22-bp indel at position-220. Whereas, the haplotype E including upland rice from the Central and South part of Thailand was separated from haplotype transversion mutation (T to A) at position-281. This subgroup did not have the wild rice and separated from *O. glaberrima* by (GGGTT)n SSR at position-67 and transition mutation (A to G) at position-127. All wild rice accessions were mapped into the second subgroup except haplotype C. The most irrigation and floating rice were located in the haplotype A same as all wild rice from Thailand. This subgroup separated from *O. glaberrima* by transition mutation (G to A) at position-349 and found that *O. rufipogon* from Malaysia separated from the haplotype A by transition mutation (T to C) at position-393. The haplotype A could be the common ancestor of cultivated rice in Thailand whereas the haplotype I could be the common ancestor of upland rice and found Nipponbare in here.

5.6 RBE1 locus

The network resulting of RBE1 from this procedure is not completed resolved (Figure 13). Only one loop indicative of ambiguous connection are present. The biggest outgroup probability is haplotype D. It is sticky rice from Lower southern Thailand. The association of haplotypes apparently did not correspond to the eating and cooking qualities, or to species or geographical location. The haplotype A could be the common ancestor of Thai rice.

5.7 RBE3 locus

The network resulting of RBE3 from this procedure is completely resolved and has only 6 nodes (Figure 14). The haplotype J (*O. glaberrima*) could not be amplified fragment by PCR with RBE3 primers. The biggest outgroup probability is haplotype A. The association of haplotypes apparently did not correspond to the

eating-cooking qualities and geographical location. The haplotype A could be the common ancestor. The wild rice was found in the haplotype A, B and F. The most accessions in the haplotype B and C were upland rice and ORF100 locus presented the non-deletion *japonica* type and Nipponbare was found in the haplotype B. The haplotype B and C separated from the haplotype A by transversion mutation (C to G) at position-135. Whereas the haplotype D and F had only one accession per haplotype and presented the deletion *indica* type. The haplotype D and F separated from the haplotype A by transition mutation (A to G) at position-532 and this mutation was located in Exon2.

5.8 Analysis of the 7 loci combined

The network of the 7 loci includes 111 haplotypes (Figure 15). The haplotype 49 could be the most recent common ancestor. The association of haplotypes did not apparently correspond to the eating and cooking qualities or to species and geographical location

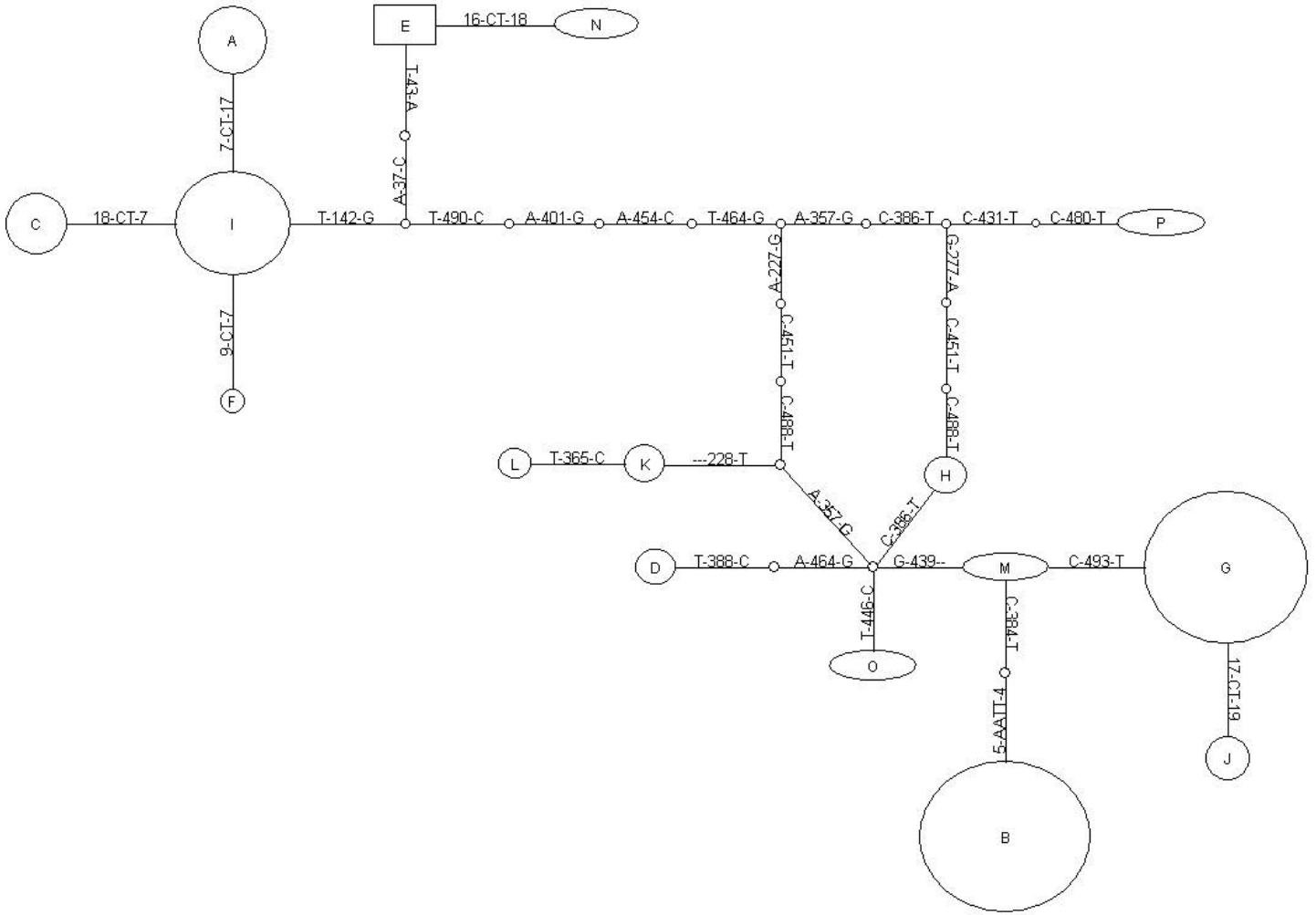


Figure 8 Phylogenetic parsimonious network of *O. sativa* and its wild relatives based on GBSSI locus.

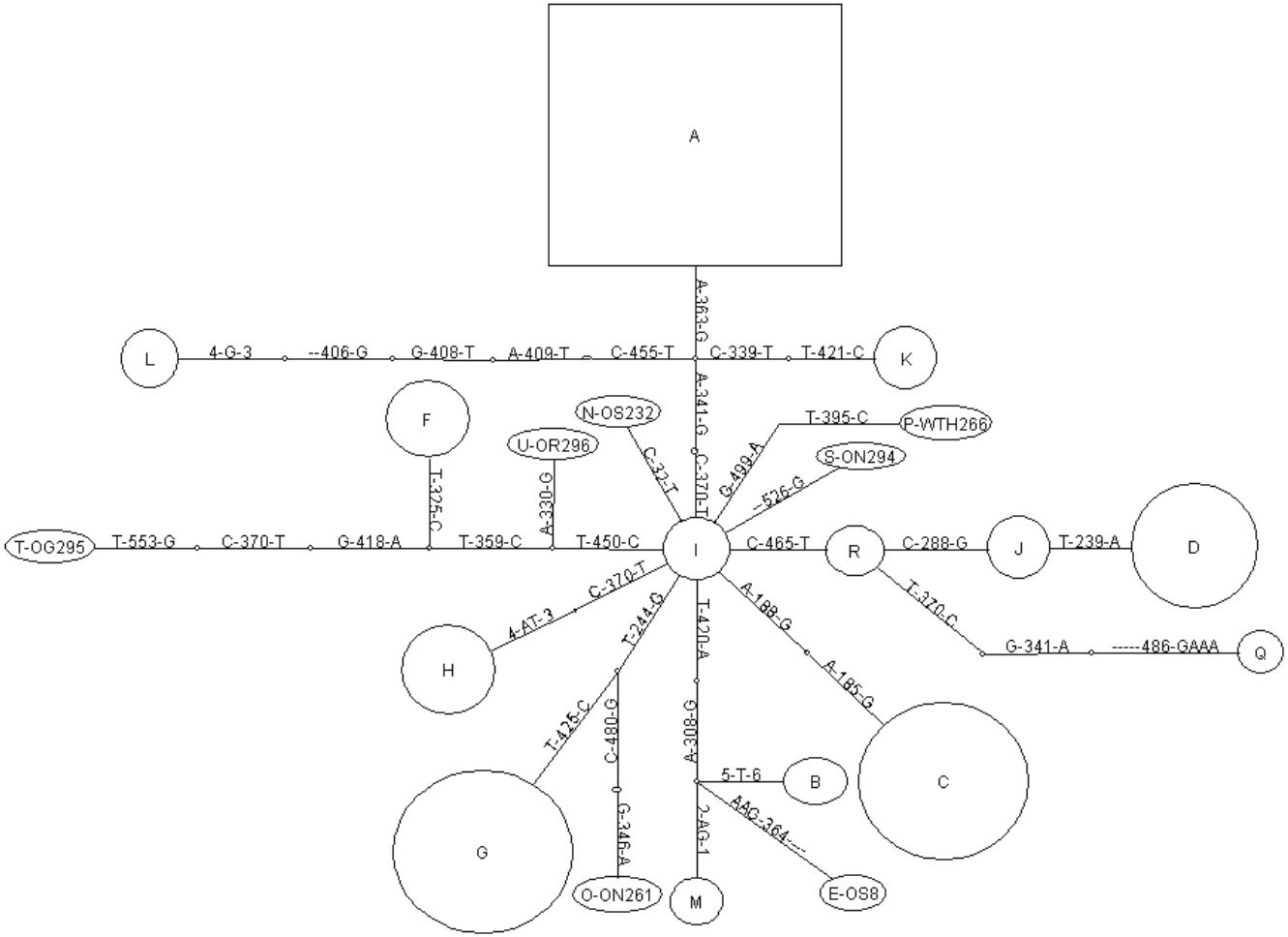


Figure 9 Phylogenetic parsimonious network of *O. sativa* and its wild relatives based on SSSIIIa locus.

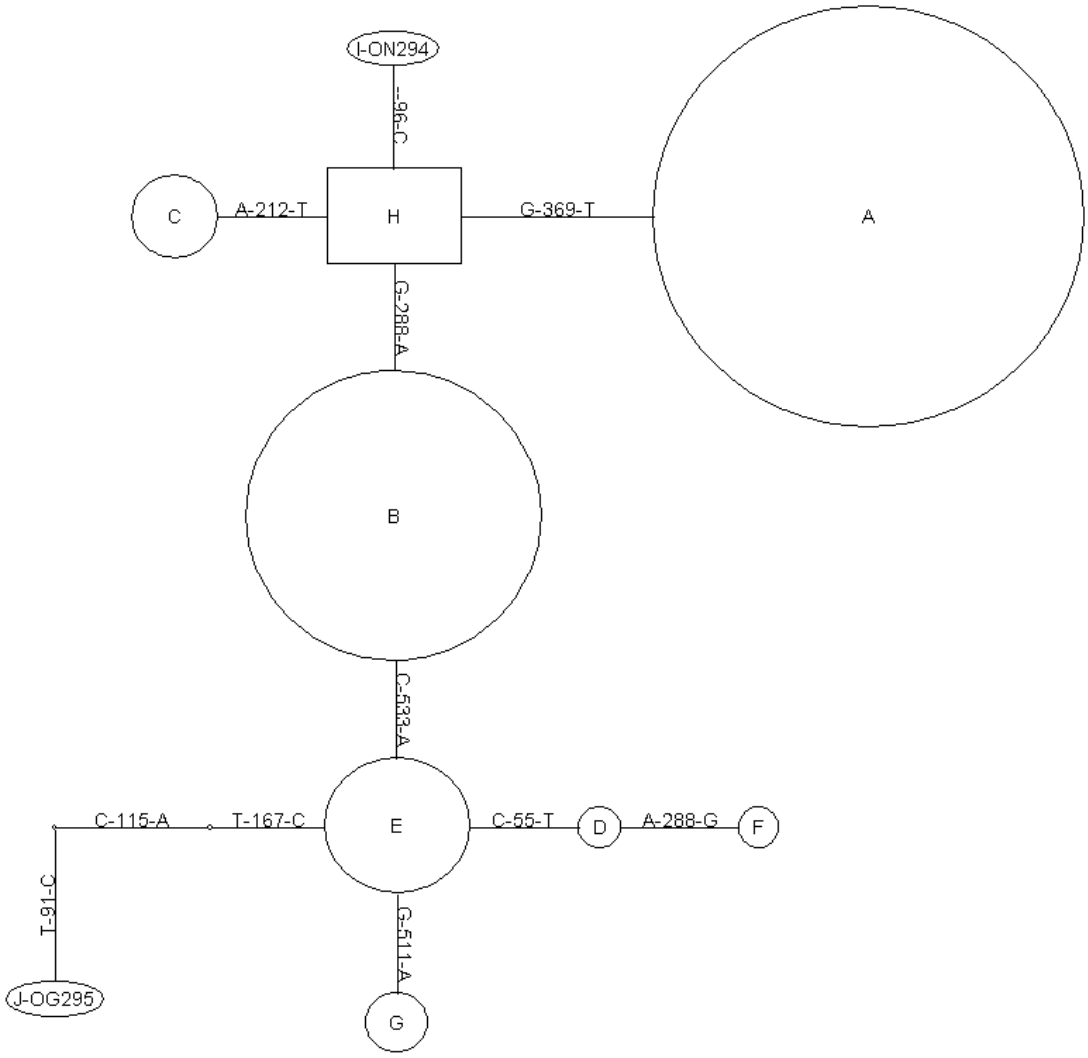


Figure 10 Phylogenetic parsimonious network of *O. sativa* and its wild relatives based on SSSIIIb locus.

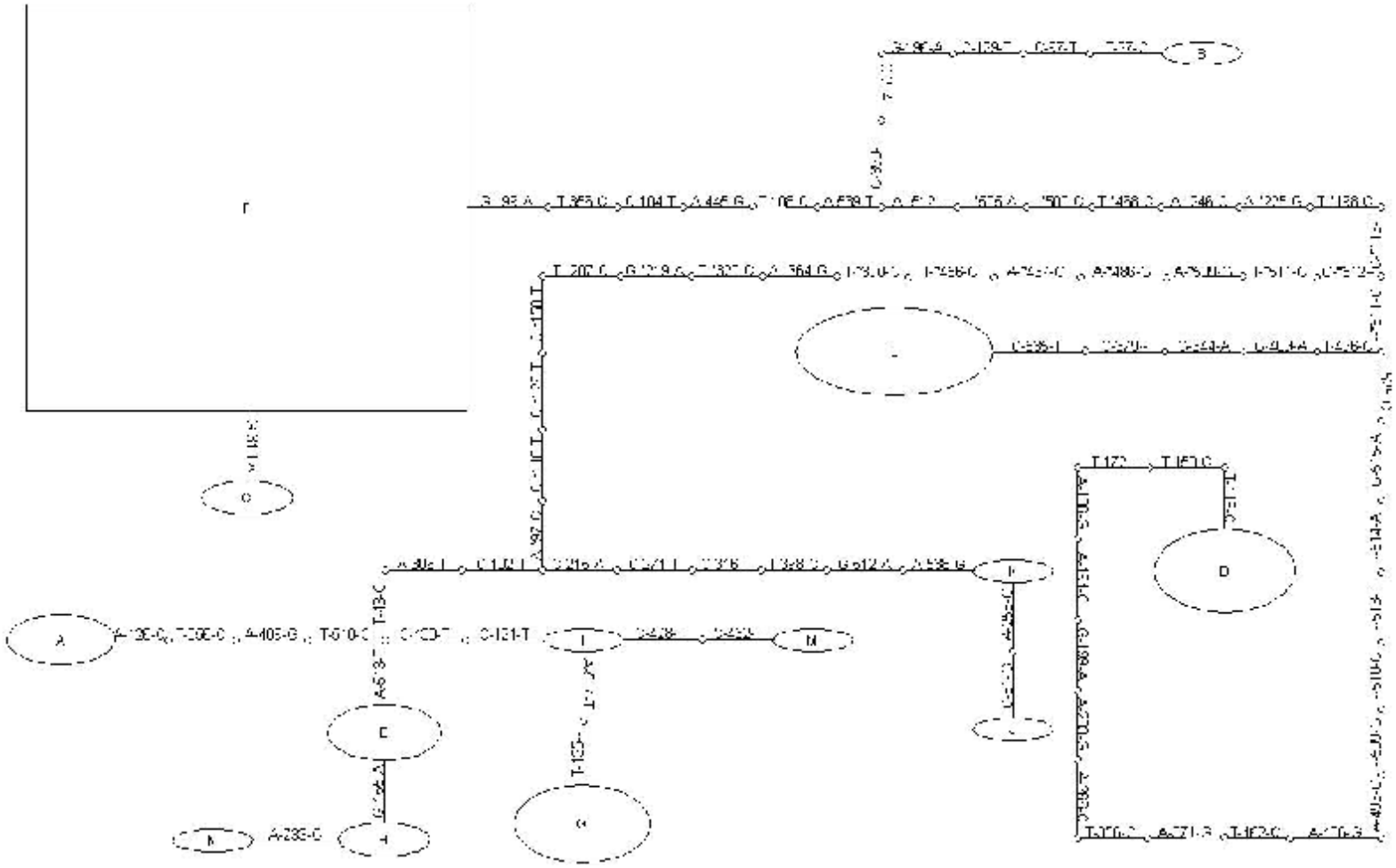


Figure 11 Phylogenetic parsimonious network of *O. sativa* and its wild relatives based on SSSIVa locus.

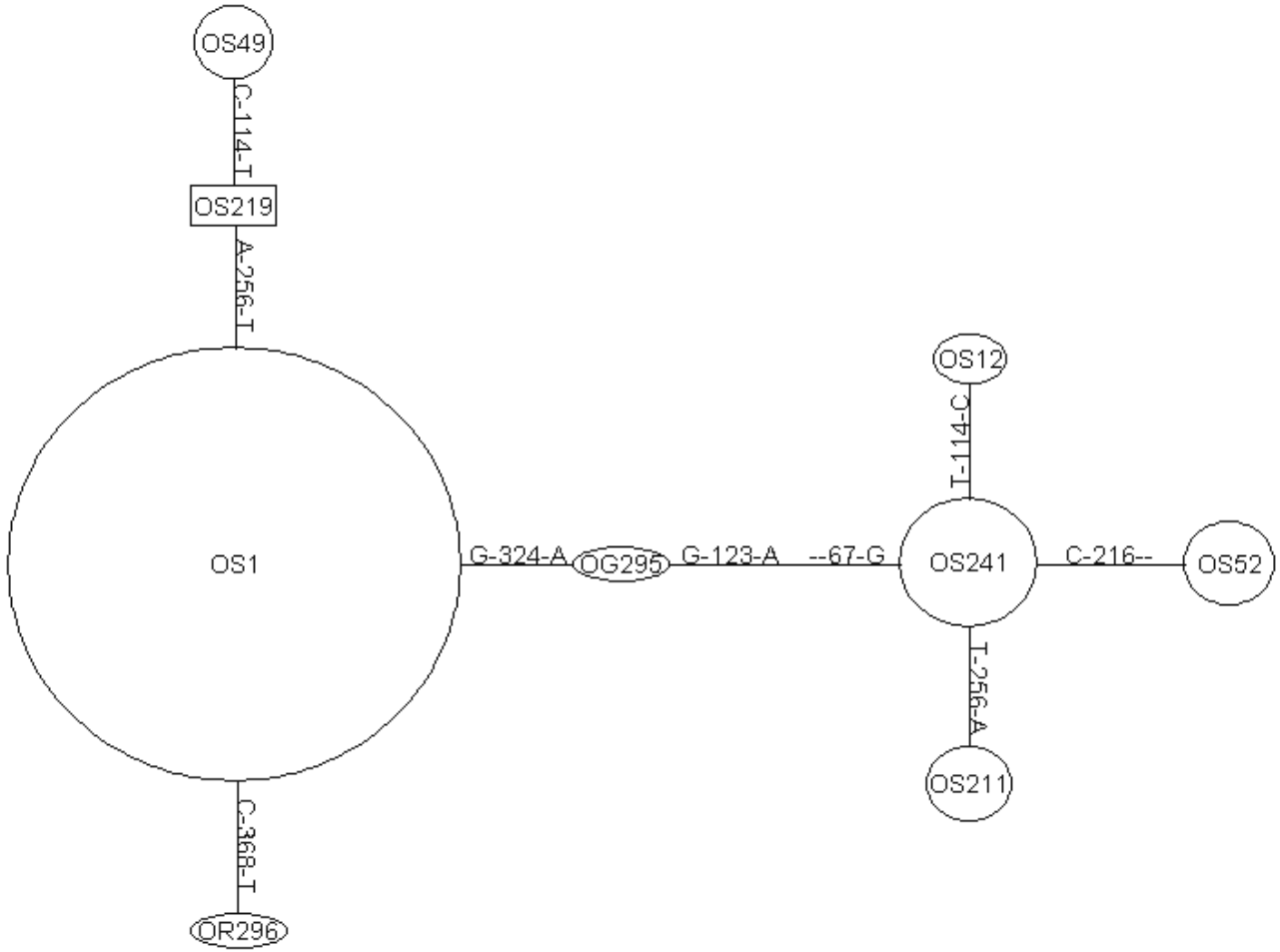


Figure 12 Phylogenetic parsimonious network of *O. sativa* and its wild relatives based on SSSIVb locus

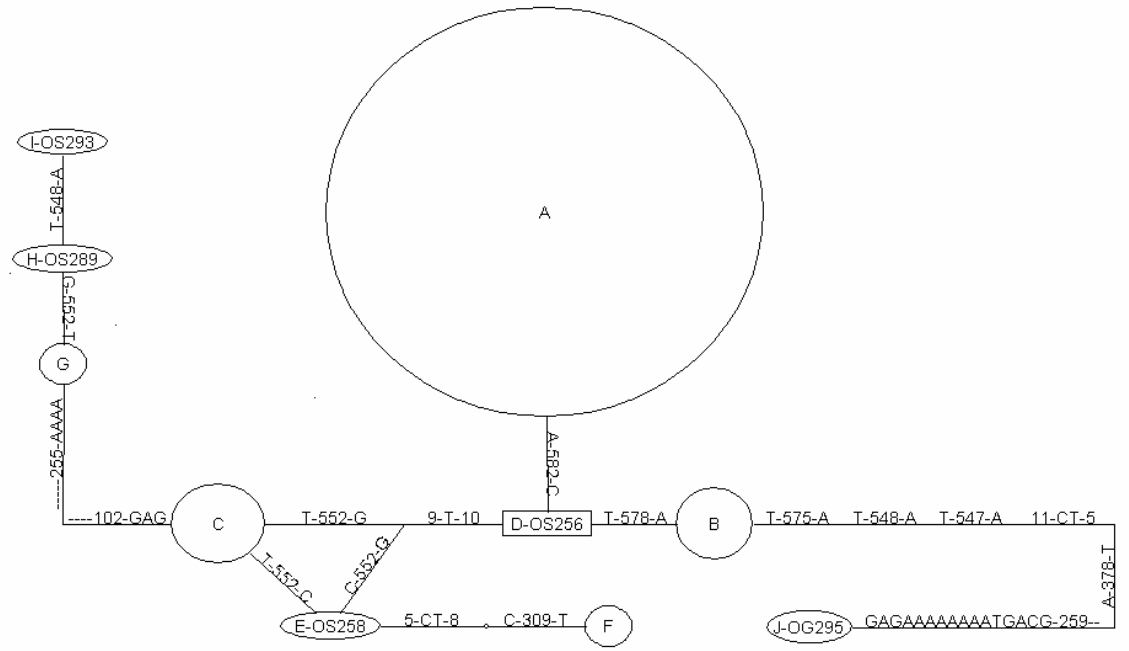


Figure 13 Phylogenetic parsimonious network of *O. sativa* and its wild relatives based on RBE1 locus.

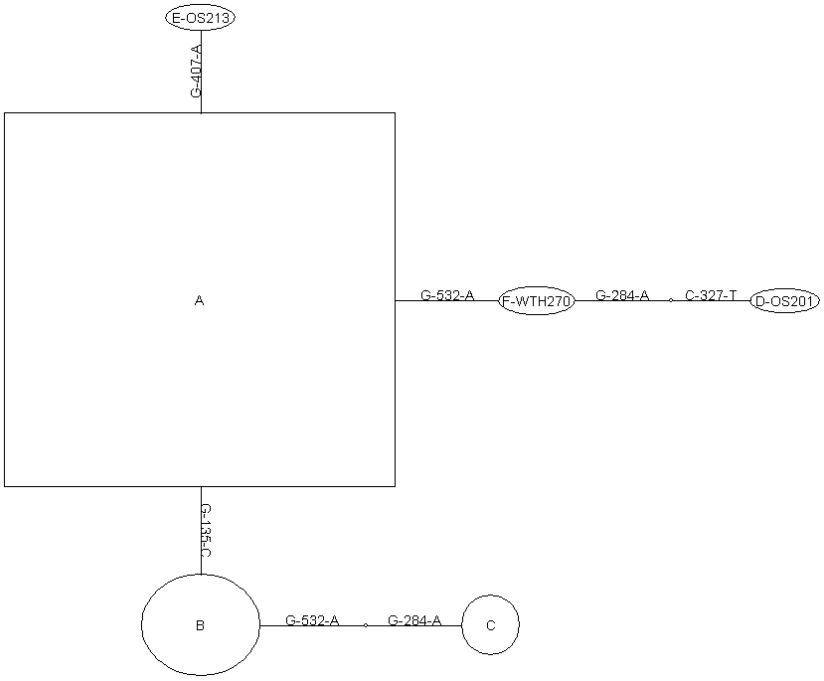


Figure 14 Phylogenetic parsimonious network of *O. sativa* and its wild relatives based on RBE3 locus.

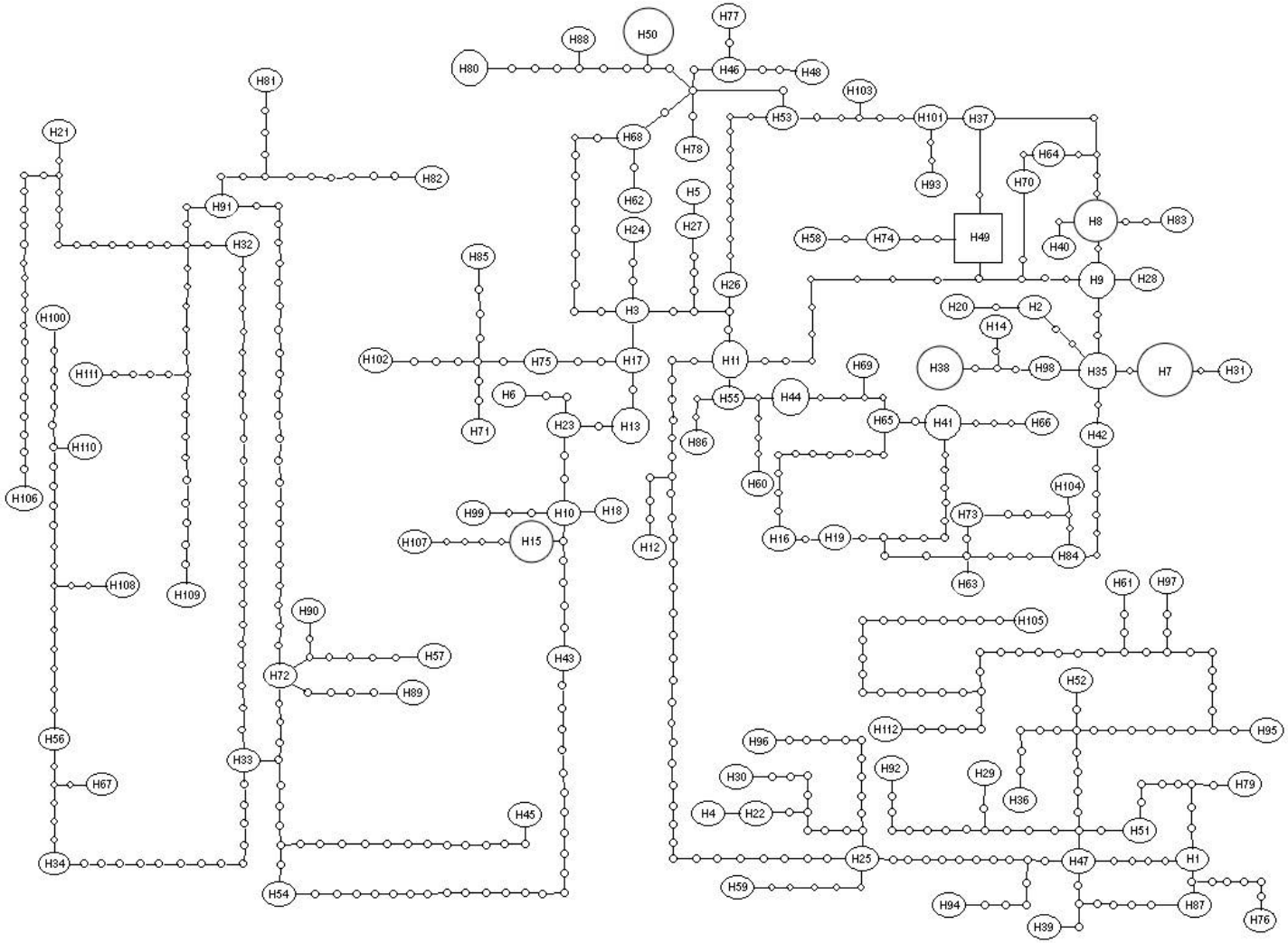


Figure 15 Phylogenetic parsimonious network of *O. sativa* and its wild relatives based on 7 gene loci.

6. Comparison between starch physico-chemical properties and SSCP haplotypes

The accessions were grouped according to the alleles observed at each of the loci and the values of the starch properties were compared across alleles. Some alleles were very rare, represented by only a single or couple of accessions. In such cases, those accessions were grouped with another accession that was nearly identical in DNA sequence with just one mutation different. If no nearly identical allele could be identified, the accessions were deleted from the analysis.

The ANOVA indicates that highly significant differences ($P < 0.01$) for each of the starch properties were present for the alleles of the GBSSI locus. For the other loci, the ANOVA indicated some significant differences in AAC and GC across alleles ($0.01 < P < 0.05$). For AAC, the ANOVA results were significant ($P < 0.05$) at all genes except for SSSIIIb and RBE1 implying that there was at least one allele which showed a means different from other alleles. The ANOVA results of GT were significant ($P < 0.05$) for GBSSI only. For GC, the ANOVA results were significant at the GBSSI, SSSIVa, SSSIVb and RBE3 loci (Table 6). The ANOVA result of AAC and GC were significant ($P < 0.05$) in many starch synthesis gene loci because the ranges of AAC and GC in Thai rice vary widely and the high correlation between the two variables. Tukey's multiple range tests were conducted for comparison between means of each of the cooking quality characters and alleles at each locus. The results are shown in Table 7. Although some significant differences between haplotypes could be detected for several genes, almost all groups that showed significantly different means contained a small number of accessions.

For SSSIIIa, SSSIIIb and RBE3 have the association of haplotypes apparently corresponds more to the AAC than other starch physico-chemical properties whereas the means of GT and GC between haplotypes did not differ significantly. For the SSSIIIa gene the haplotypes B, E, M and K presented high AAC whereas the haplotype L showed low AAC. For SSSIIIb locus, the haplotype E showed the low AAC with significantly. For RBE3 locus, the allele C showed a significantly lower

AAC compared to accessions with other alleles. The SSSIVa, SSSIVb and RBE1 loci did not show different between means of each the starch physico-chemical properties.

Table 6 The ANOVA results of comparison of the mean values of starch physico-chemical properties of the haplotypes at each gene locus.

Genes	AAC[a]		GT[b]		GC[c]	
	F-statistic	*P*[d]	*F*-statistic	*P*[d]	*F*-statistic	*P*[d]
GBSSI	8.9635	0.0000*	3.9973	0.0003*	6.7595	0.0000*
SSSIIIa	2.3588	0.0134*	0.9958	0.4504	1.5914	0.1156
SSSIIIb	2.3248	0.0775	1.1753	0.3214	2.4453	0.0665
SSSIVa	2.7709	0.0141*	1.8778	0.0886	2.1880	0.0475*
SSSIVb	2.8975	0.0243*	0.4163	0.7967	2.9981	0.0207*
RBE1	1.9785	0.142	1.1843	0.3089	2.8901	0.0588
RBE3	4.1968	0.0168*	0.6491	0.5240	3.1248	0.0468*

[a] *AAC* apparent amylose content

[b] *GT* gelatinization temperature

[c] *GC* gel consistency

[d] *P*-value of normality test

* Significant at $P < 0.05$

Table 7 Comparison of the mean values of starch physiochemical properties with the haplotypes at each gene locus.

Genes	Haplotypes[a]	No. of samples	AAC[b]		GT[c]		GC[d]	
			Mean	σ^e	Mean	σ^e	Mean	σ^e
GBSSI	A $(CT)_{17}$	18	8.6099a	5.0478	6.3889b	1.1950	10.1556b	2.9404
	B $(CT)_{11}$	35	15.0575abc	8.4655	6.3429ab	0.7648	7.9571ab	4.0808
	C $(CT)_{18}$	16	18.3557abc	8.2778	5.8125ab	1.0468	7.5313ab	3.5118
	D $(CT)_8$	5	7.6281a	9.2345	6.2000ab	0.8367	10.0200b	1.7427
	E $(CT)_{18}$	5	18.9341abc	7.5132	6.8000b	0.4472	5.2400ab	2.3776
	F $(CT)_9$+I $(CT)_7$	22	11.2275ab	8.9630	6.3182ab	0.7162	9.6227ab	3.4112
	G $(CT)_{17}$+J $(CT)_{19}$	35	22.9791bc	6.1844	5.5714ab	0.9167	4.6771a	2.4860
	H $(CT)_{12}$	4	10.2507a	8.1117	6.7500b	0.5000	9.9250b	3.9953
	K $(CT)_{10}$+L $(CT)_{10}$	4	23.5926c	0.9432	5.0000a	0.0000	7.7750ab	3.4596
	Total	144						
SSSIIIa	A	48	15.7010ab	9.2888	6.1042	0.9048	7.1900	4.0653
	B+E+M	6	23.8014b	3.8476	5.8333	0.9832	5.5800	3.1403
	C	26	12.8286ab	8.9060	6.0385	0.9584	8.7200	3.7445
	D	15	19.9303ab	7.9292	5.8667	0.9904	6.3700	2.9531
	F	3	12.7784ab	8.2152	7.0000	0	7.7000	4.3555
	G	23	13.7196ab	9.6324	6.2609	0.7518	9.5300	3.5898
	H	7	11.9085ab	9.2720	5.7143	1.8898	9.3300	4.5305
	I+N	6	13.5622ab	6.9769	6.3333	1.2111	7.5300	2.6326
	J	4	20.0878ab	4.7349	6.5000	0.5774	7.5300	2.8906
	K	4	25.5750b	1.9186	5.2500	0.5000	6.2500	3.1300
	L	2	5.9473a	1.9551	6.0000	0	11.3000	0.3536
	Total	144						

Table 7 (Continued)

Genes	Haplotypes[a]	No. of samples	AAC[b]		GT[c]		GC[d]	
			Mean	σ[e]	Mean	σ[e]	Mean	σ[e]
SSSIIIb	A	81	16.9830b	8.6450	5.9506	1.0356	7.3300	3.8077
	B	54	15.8143b	9.5442	6.1852	0.8027	7.7400	3.6501
	C	7	13.3564ab	7.9924	5.8571	1.0690	9.8900	3.8247
	E	3	4.0784a	1.7709	6.6667	0.5774	12.1000	1.6258
	Total	145						
SSSIVa	A	6	20.7450	4.5323	6.5000	0.8367	5.0000	3.1887
	C	3	17.9350	9.8699	7.0000	0	4.9300	4.0550
	D	8	17.4830	7.8823	5.7500	1.1650	7.8800	3.6714
	E	7	19.8980	10.4097	5.5714	1.1339	6.4600	4.7689
	F	104	16.5410	8.8408	6.0192	0.9752	7.4600	3.5982
	G	7	19.1530	9.42934	5.7143	0.7559	8.0600	4.4211
	L	14	8.0722	6.8828	6.5000	0.5189	10.3000	3.0417
	Total	149						
SSSIVb	A	121	17.0500	8.8064	6.0413	0.9866	7.3300	3.7073
	B+I	11	14.6800	9.2416	6.0000	1.0000	8.1800	3.4388
	C+F	4	10.5100	7.0485	6.2500	0.5000	8.1300	4.5573
	D	5	5.4420	0.7489	6.4000	0.5477	12.9000	1.3008
	E	4	12.3000	10.6725	6.5000	1.0000	8.9300	2.9477
	Total	145						
RBE1	A	120	15.3800	9.1298	6.0250	0.9912	8.0400	3.8758
	B	19	17.3770	8.4622	6.3684	0.8307	6.3100	3.3454
	C+E	10	20.8970	8.6240	5.9000	0.8756	5.9500	2.5692
	Total	149						

Table 7 (Continued)

Genes	Haplotypes[a]	No. of samples	AAC[b]		GT[c]		GC[d]	
			Mean	σ^e	Mean	σ^e	Mean	σ^e
RBE3	A+E	129	16.7502b	8.9646	6.0388	0.9715	7.3900	3.7785
	B	21	13.9409b	8.6790	6.0952	0.8891	8.7400	3.4150
	C	3	3.0919a	1.1731	6.6667	0.5774	11.8000	1.7898
	Total	153						

Means having a different letter are significantly different ($P < 0.05$).

[a] Some haplotypes that were represented by 1-3 accessions were combined with a closely related one or were excluded from the analysis if no close relative could be identified. A few accessions for which no SSCP pattern could be obtained were not included in the analysis.

[b] AAC apparent amylose content, %.

[c] GT gelatinization temperature, Alkali score 1 to 7.

[d] GC gel consistency, cm.

[e] standard deviation

All the genes examined in this paper encode key enzymes in the synthesis of amylose and amylopectin (Nakamura, 2002). However, other genes such as SSSIIa may play additional roles in influencing the physico-chemical behavior of rice starch. Different alleles of SSSIIa in rice contribute to different fine structure of amylopectin, leading to different gelatinization temperature of starch, as indicated by Umemoto *et al.* (2002). It is not well understood whether different alleles of these genes exist in natural populations and whether these alleles are correlated with starch physico-chemical properties (Nakamura *et al.*, 2005). Therefore, further studies are needed to investigate the roles these genes play in relation to naturally occurring variation in starch properties.

7. Linkage disequilibrium

Linkage disequilibrium (LD) is a measure of non-random association between alleles at different loci in a population. In a statistical sense, LD represents the covariance of polymorphisms exhibited by two molecular markers/genes. One of the major current and future uses of LD in plants would be to study marker-trait association (without the use of a mapping population) followed by marker-assisted selection (MAS). Another important application is the study of genetic diversity in natural populations and germplasm collections and its use in the study of population genetics and in crop improvement programmes respectively.

Consider two loci, *A* and *B*, with alleles A_1/A_2 and B_1/B_2, respectively. Let p_{Ai} stand for the frequency of allele *Ai*, where i = 1, 2, at locus *A*, and similarly for locus *B*. Let p_{AiBj} stand for the frequency of the *AiBj* haplotype. The basic component of all LS statistics is the difference between the observed and expected haplotype frequencies, $D = p_{A1B1} - p_{A1}p_{B1}$. The two most common statistics for measuring LD were r^2 and |D'|. The distinction between these statistics lies in the scaling of this difference. The first of the two measures, r^2, is calculated ad $r^2 = D^2/(p_{A1}p_{A2}p_{B1}p_{B2})$. It is convenient to consider r^2 as the square of the correlation coefficient between the two loci. Statistical significance (P-value) for LD is usually calculated using either Fisher's exact test to compare sites with two alleles at each locus. Alternatively, the LS statistic |D'| is the absolute value of $D = p_{A1B1} - p_{A1}p_{B1}$, normalized to take values between 0 and 1 regardless of the allele frequencies (Nordborg and Tavare, 2002; Flint-Garcia *et al.*, 2003; Gupta *et al.*, 2005; Du *et al.*, 2007; Rakshit *et al.*, 2007).

There are two common ways to visualize the extent of LD between pairs of loci. LD decay plots are used to visualize the rate at which LD declines with genetic distance. Scatter plots of r^2 values versus genetic distances between all pairs of alleles. Alternatively, disequilibrium matrixes are effective for visualizing the linear arrangement of LD between polymorphic sites.

LD measure (r^2) between pairs of polymorphic sites was calculated (upper right), and significance was determined by a Fisher's exact test (lower left). Shading

indicated the magnitude of the linkage disequilibrium and the significance level. Two LD block observed in GBSSI and SSSIVa loci (Figure 16)

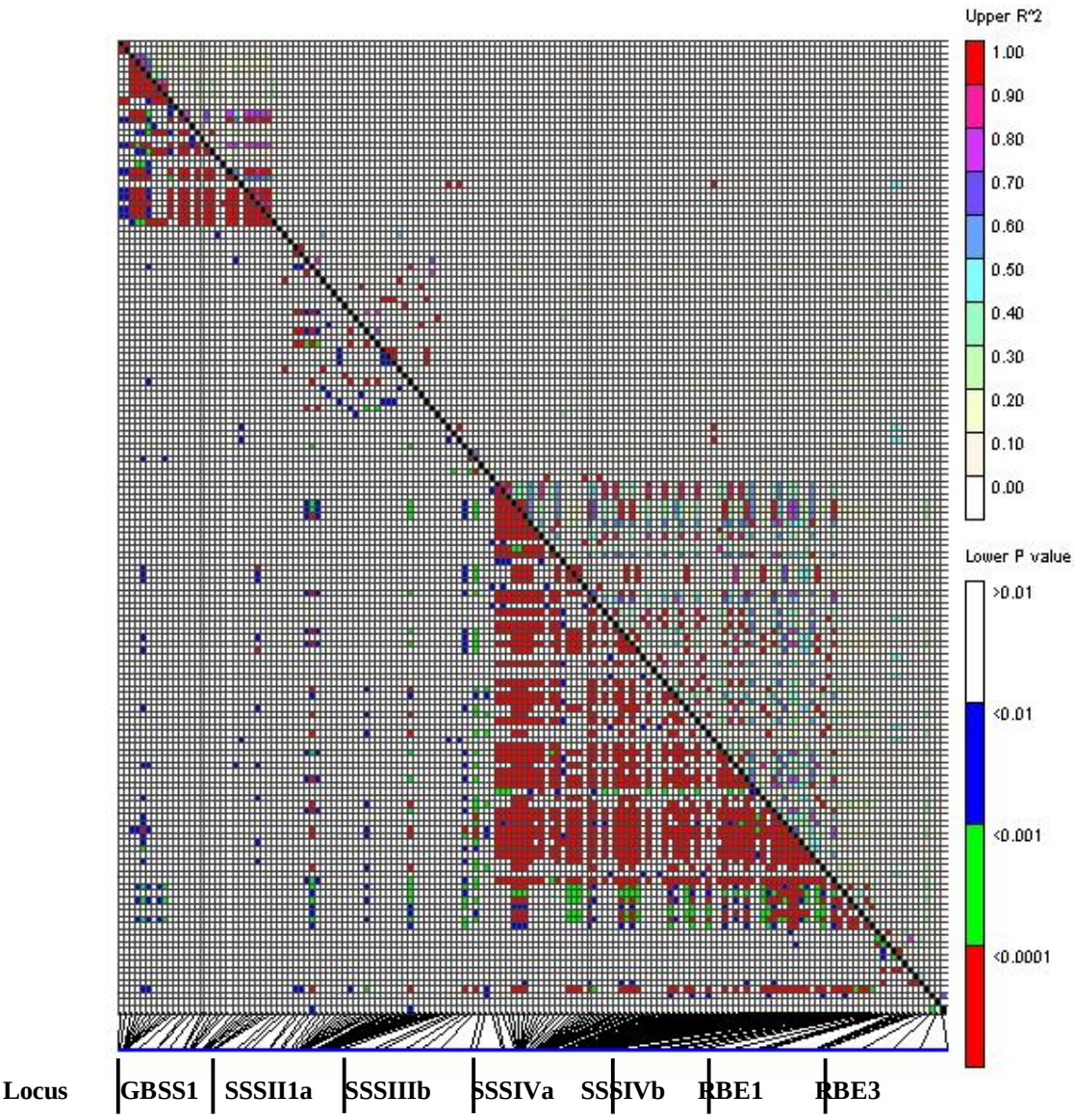


Figure 16 LD measure (r^2, above diagonal line) and probability value (*P*, below diagonal line) for sequence data of 7 gene loci.

Four interloci LD groups were detected at $r^2 > 0.2$ and $P < 0.05$. The first group was relationship between 38 polymorphic sites in five gene loci namely SSSIIIa, SSSIIIb, SSSIVa, SSSIVb, and RBE3. The second group was relationship between twelve polymorphic sites in five gene loci namely GBSSI, SSSIIIa, SSSIIIb, SSSIVa, and RBE1. The third group was relationship between four polymorphic sites

in two gene loci namely SSSIIIa and RBE1. Finally, the last group was relationship between three polymorphic sites in two gene loci namely SSSIIIa and SSSIVa.

For GBSSI locus, 251 out of 435 comparisons (57.70%) were in LD with $P < 0.05$, and 195 comparisons had r^2 higher than 0.05. The values of r^2 ($P < 0.05$) varied in the range of 0.0269-1.0000 with an average of 0.2542. LD extended to distances as long as up to 439 bp with $r^2 > 0.05$, or up to 413 bp with $r^2 > 0.2$. The regression curve and LD matrix indicated that LD declined slowly (Figure 17 and 18).

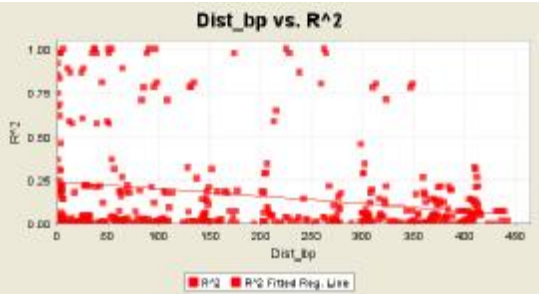


Figure 17 Extent of LD in GBSSI locus.

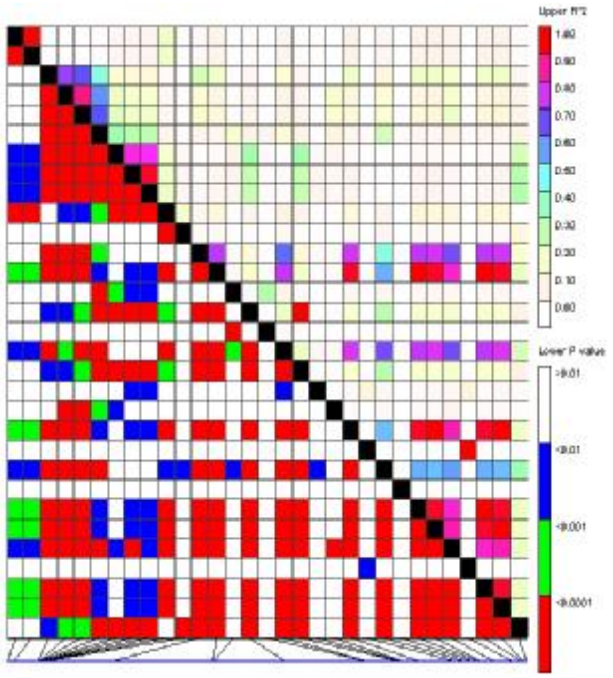


Figure 18 LD measure (r^2, above diagonal line) and probability value (P, below diagonal line) for sequence data of GBSSI locus.

For SSSIII1a locus, LD was assessed for 630 combinations. Significant LD was observed in 86 comparisons (13.65%) at $P < 0.05$, and 58 comparisons had r^2 higher than 0.05. The values of r^2 ($P < 0.05$) varied in the range of 0.0222-1.0000 with an average of 0.3384. LD extended to distances as long as up to 345 bp with $r^2 > 0.05$, or with $r^2 > 0.2$. The regression curve and LD matrix indicated that LD declined slowly (Figure 19 and 20).

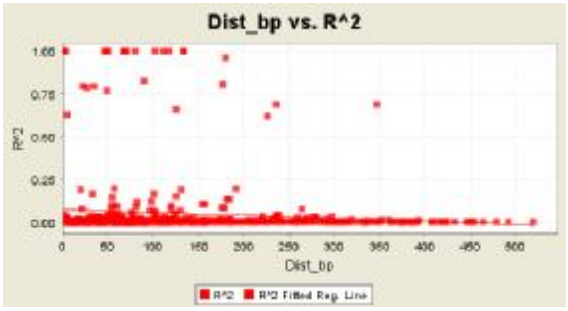


Figure 19 Extent of LD in SSSIIIa locus.

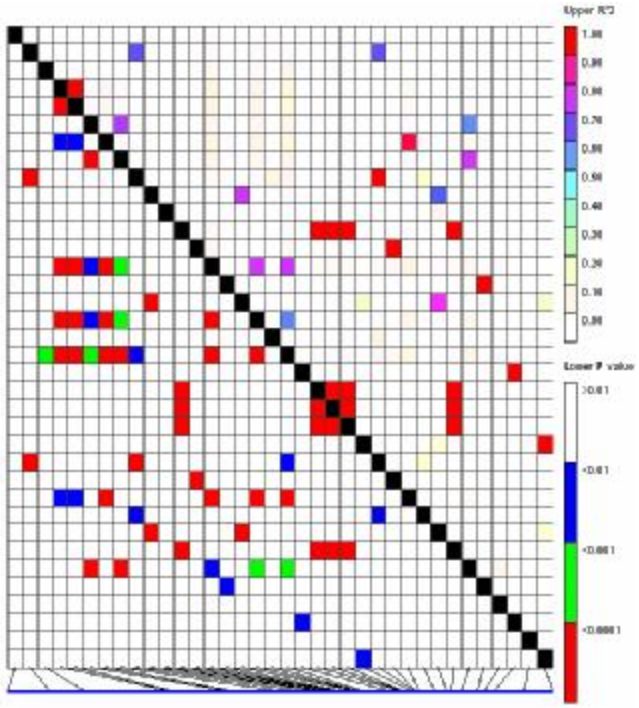


Figure 20 LD measure (r^2, above diagonal line) and probability value (P, below diagonal line) for sequence data of SSSIIIa locus.

For SSSIIIb locus, LD was assessed for 45 combinations. Significant LD was observed in 10 comparisons (22.22%) at $P < 0.05$, and 8 comparisons had r^2 higher than 0.05. The values of r^2 ($P < 0.05$) varied in the range of 0.0272-1.0000 with an average of 0.4194. LD extended to distances as long as up to 478 bp with $r^2 > 0.05$, or up to 81 bp with $r^2 > 0.2$. The regression curve and LD matrix indicated that LD declined slowly (Figure 21 and 22).

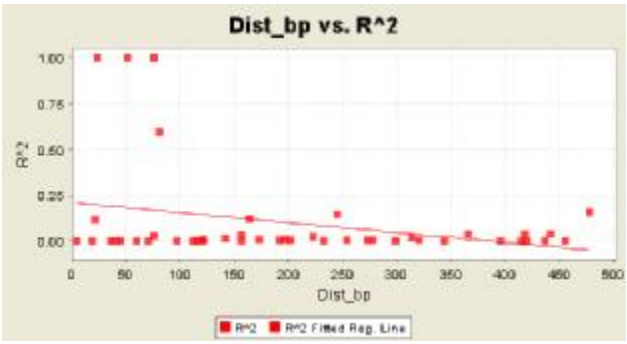


Figure 21 Extent of LD in SSSIIIb locus.

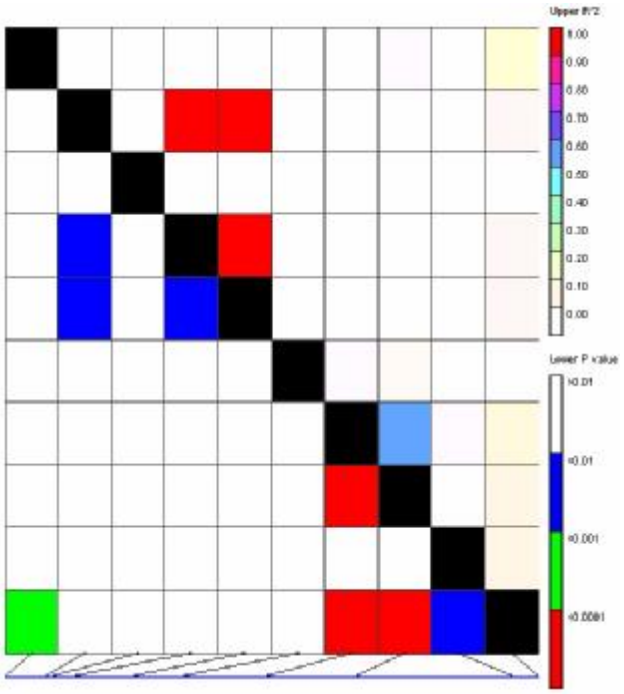


Figure 22 LD measure (r^2, above diagonal line) and probability value (P, below diagonal line) for sequence data of SSSIIIb locus.

For SSSIVa locus, LD was assessed for 2016 combinations. Significant LD was observed in 1050 comparisons (52.08%) at $P < 0.05$, and 922 comparisons had r^2 higher than 0.05. The values of r^2 ($P < 0.05$) varied in the range of 0.0213-1.0000 with an average of 0.4021. LD extended to distances as long as up to 497 bp with $r^2 > 0.05$, or $r^2 > 0.2$. The regression curve and LD matrix indicated that LD did not decline (Figure 23 and 24). This locus should be only one haplotype block on the chromosome.

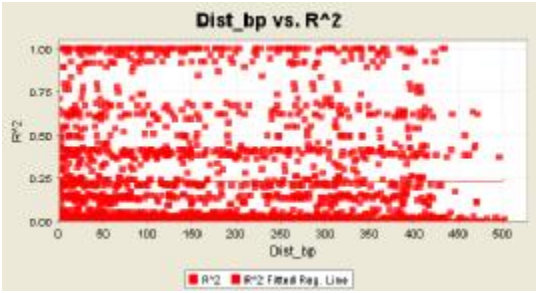


Figure 23 Extent of LD in SSSIVa locus.

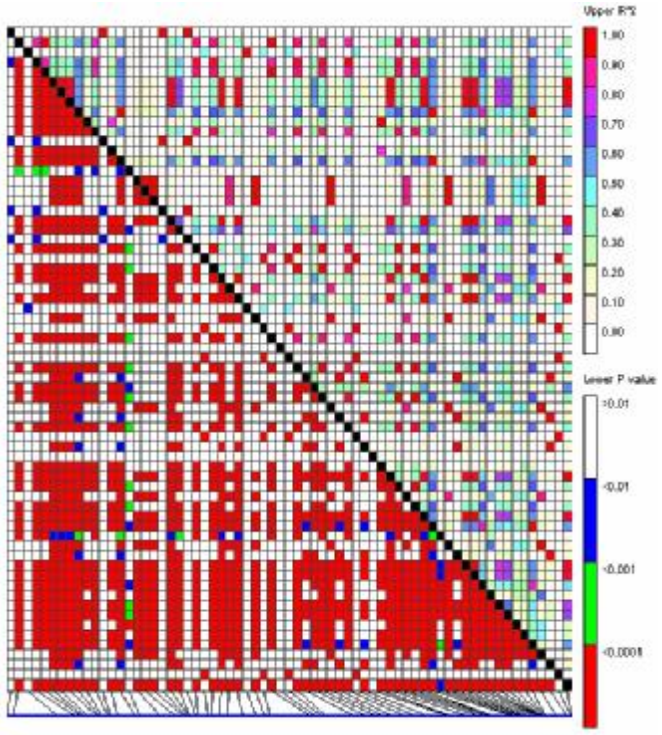


Figure 24 LD measure (r^2, above diagonal line) and probability value (P, below diagonal line) for sequence data of SSSIVa locus.

For SSSIVb locus, LD was assessed for 28 combinations. Significant LD was observed in 15 comparisons (53.57%) at $P < 0.05$, and 12 comparisons had r^2 higher than 0.05. The values of r^2 ($P < 0.05$) varied in the range of 0.0369-1.0000 with an average of 0.3256. LD extended to distances as long as up to 326 bp with $r^2 > 0.05$, or up to 257 bp with $r^2 > 0.2$. The regression curve and LD matrix indicated that LD did not decline (Figure 25 and 26).

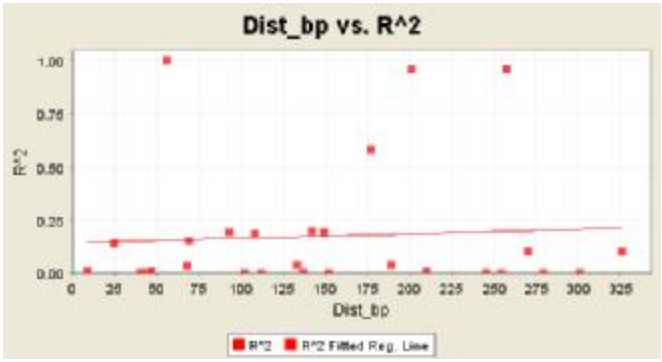


Figure 25 Extent of LD in SSSIVb locus.

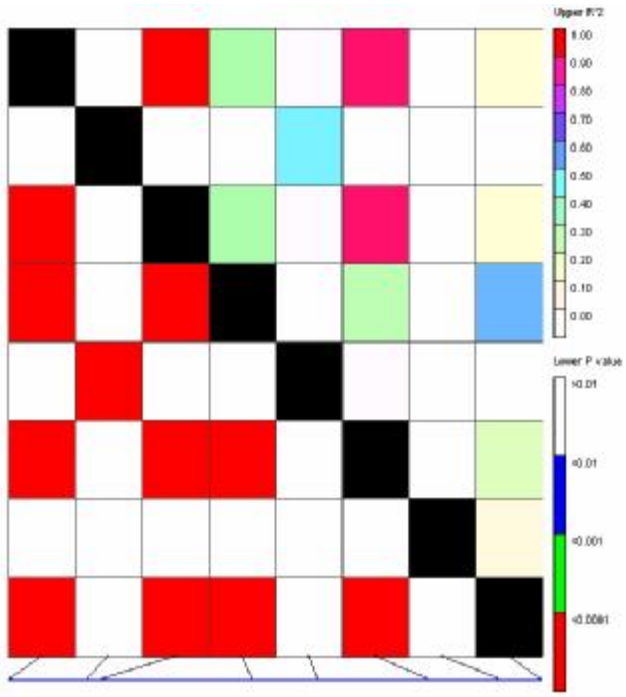


Figure 26 LD measure (r^2, above diagonal line) and probability value (P, below diagonal line) for sequence data of SSSIVb locus.

For RBE1 locus, LD was assessed for 91 combinations. Significant LD was observed in 44 comparisons (48.35%) at $P < 0.05$, and 43 comparisons had r^2 higher than 0.05. The values of r^2 ($P < 0.05$) varied in the range of 0.0497-1.0000 with an average of 0.4760. LD extended to distances as long as up to 433 bp with $r^2 > 0.05$, or up to 313 bp with $r^2 > 0.2$. The regression curve and LD matrix indicated that LD declined slowly (Figure 27 and 28).

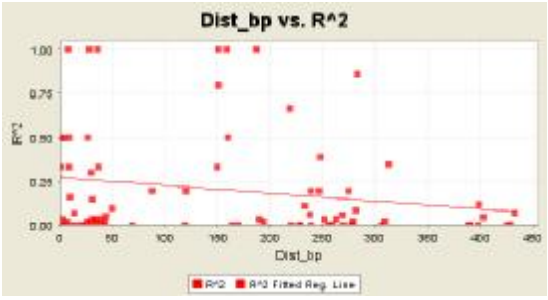


Figure 27 Extent of LD in RBE1 locus.

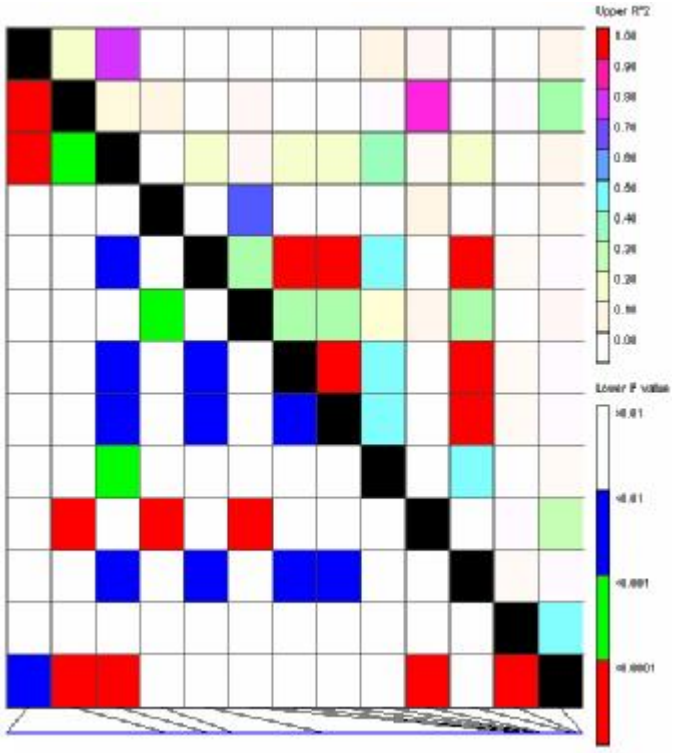


Figure 28 LD measure (r^2, above diagonal line) and probability value (P, below diagonal line) for sequence data of RBE1 locus.

For RBE3 locus, LD was assessed for 10 combinations. Significant LD was observed in 5 comparisons (50%) at $P < 0.05$, and 4 comparisons had r^2 higher than 0.05. The values of r^2 ($P < 0.05$) varied in the range of 0.0388-0.8281 with an average of 0.2558. LD extended to distances as long as up to 248 bp with $r^2 > 0.05$ and $r^2 > 0.2$. The regression curve and LD matrix indicated that LD did not decline (Figure 29 and 30).

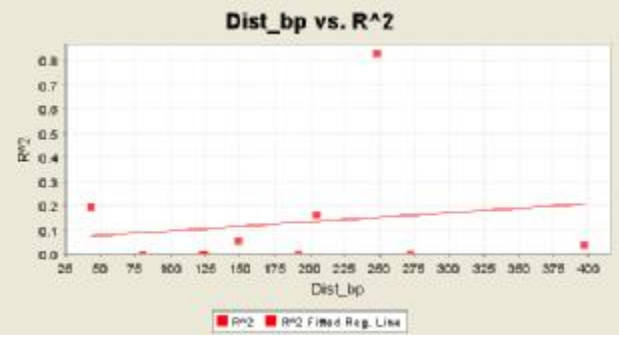


Figure 29 Extent of LD in RBE3 locus.

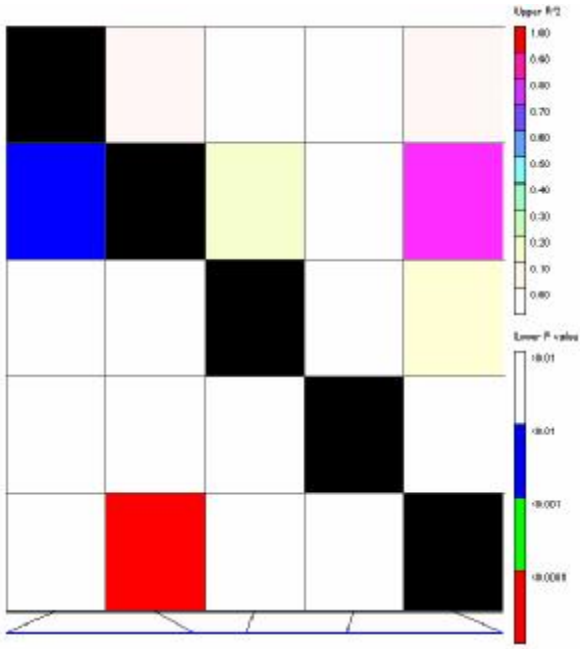


Figure 30 LD measure (r^2, above diagonal line) and probability value (P, below diagonal line) for sequence data of RBE3 locus.

8. Population structure

Population structure was inferred from haplotypes of the seven gene sequences by using the program STRUCTURE Version 2.1. This software applies a Bayesian clustering approach to identify subpopulations. The procedure assigns individuals to these populations while simultaneously estimating the population allele frequencies. STRUCTURE produces a Q-matrix that lists the estimated membership coefficients for each individual in each cluster (Pritchard ***et al.***, 2000; Falush ***et al.***, 2003). The estimated Q-matrices were used in the subsequent association analysis carried out by logistic regression in the TASSEL.

The present study, we treated each accession as a haploid in the model of the Bayesian clustering analysis because rice is a highly selfing species and the accessions were almost completely homozygous. We used a Markov chain Monte Carlo (MCMC) method to estimate the number of subpopulations, the allele frequencies in each subpopulation, and the value of Q for each sampled individual. For each run, MCMC cycles were repeated 1,000,000 times after 10,000 cycles of a burn-in period followed by Iwata ***et al.*** (2007).

Cluster analysis from STRUCTURE allowed the lines from the four pools including cultivated rice-ORF100 deleted type (CD), wild rice-ORF100 deleted type (WD), cultivated rice-ORF100 no deleted type (CN) and wild rice-ORF100 no deleted type (WN). The summary plot of estimates of Q was shown in Figure 30. Each individual in the data set was represented by a single vertical line, which was partitioned into K colored segments that represent that individual's estimated membership fraction each of the K inferred clusters.

From Figure 31 and followed the four steps for the graphical method allowing detection of the true number of groups K using an ad hoc statistic ΔK based on the rate of change in the log probability of data between successive K values (Evanno ***et al.***, 2005), the true value of K cannot be clearly indicated because our data set consisted of just 7 loci because the present study was similar the result of Pritchard ***et***

al. (2000) that analyzed the data set, including just 5 microsatellite loci, actually contains two populations. The result for K is set to 3 is also shown for comparison.

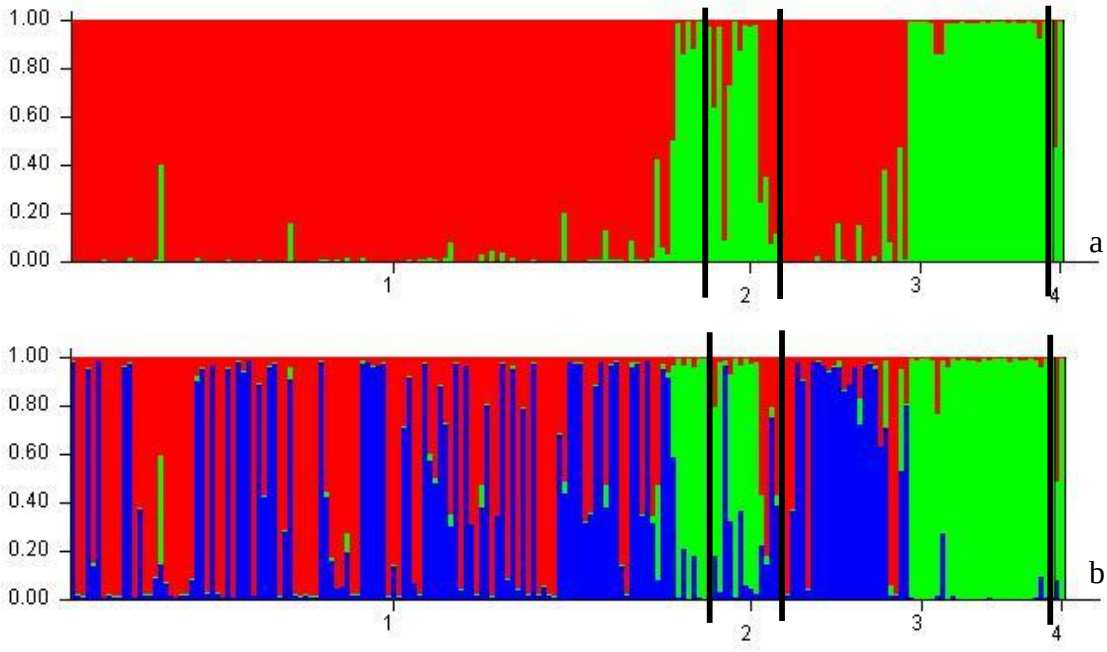


Figure 31 Summary plot of estimates of Q at K = 2 (a) and K = 3(b). The numbers (1-4) correspond to the four pools: 1=CD, 2=WD, 3=CN, and 4=WN.

9. Association analysis

The program TASSEL Version 2.0.1 (http://www.maizegenetics.net/ bioinformatics/tasselindex.htm) applies a logistic regression ration test to calculate the likelihood of the candidate marker allele distribution being associated with population structure and eating-cooking quality. The test statistic the ratio between these two likelihoods indicates associations between individual polymorphisms and traits. Furthermore, the general linear model (GLM) analysis in Tassel was used in the analysis of associations between nucleotide polymorphisms or marker alleles and different eating and cooking quality.

Association analysis by GLM carried out associations between segregating sites and phenotypes, while accounting for population structure. Population covariates, The Q-matrix from STRUCTURE, were included in the model to adjust

for population structure (Bradbury *et al.*, 2006). The AAC, GT and GC of the haplotypes at each locus were compared.

9.1 GBSSI locus

For AAC, thirteen polymorphisms were significantly associated at $P < 0.05$, of which seven SNPs showed a very strong association at $P < 0.0001$. Ten polymorphisms were significantly associated with GT at $P < 0.05$, of which one SNP (C/T at site 490) showed a strong association with GT at $P < 0.01$, and eleven polymorphisms were significantly associated with GC at $P < 0.01$, of which two mutations (indel (G) at site 436 and C/T at site 490) showed a very strong association at $P < 0.0001$ (Table 8).

Table 8 Nucleotide polymorphisms in the GBSSI locus associated with the AAC, GT and GC.

Site	Haplotype													Association (F value/r^2)		
	A[a]	B	C	D	E	F	G	H	I	J	K	L	M	AAC	GT	GC
37	A	A	A	A	C	A	A	A	A	A	A	A	A	4.1175/ 0.0246*	4.4998/ 0.0319*	0.1842/ 0.0014
43	T	T	T	T	A	T	T	T	T	T	T	T	T	4.0484/ 0.0272*	4.4998/ 0.0319*	0.18455/ 0.0118
49-82 $(CT)_n$	17	11	18	8	18	9	17	12	7	19	10	10	17	0.4823/ 0.0033	0.0414/ 0.0003	0.1850/ 0.0014
138	T	G	T	G	G	T	G	G	T	G	G	G	G	10.2973/ 0.0596**	0.4959/ 0.0036	15.4283/ 0.1004**
223	G	A	G	A	G	G	A	A	G	A	A	A	A	16.4522/ 0.101***	5.5547/ 0.0358*	13.5488/ 0.0894**
224-225 TA	+	+	+	+	+	+	+	+	+	+	-	-	+	1.8258/ 0.0125	4.6667/ 0.0303*	0.3303/ 0.0024
269-284 $(AATT)_n$	4	5	4	4	4	4	4	4	4	4	4	4	4	1.0818/ 0.0074	1.1384/ 0.0083	0.7934/ 0.0058
352	C	C	C	C	C	C	C	C	C	C	C	T	C	0.6852/ 0.0047	0.6878/ 0.0050	0.1009/ 0.0001
353	A	G	A	G	A	A	G	G	A	G	A	A	G	11.2768/ 0.0718**	0.5020/ 0.0037	14.765/ 0.0966**
380	C	T	C	C	C	C	C	C	C	C	C	C	C	1.0818/ 0.0074	1.1384/ 0.0083	0.7934/ 0.0058
382	C	C	C	C	C	C	C	T	C	C	C	C	C	1.4809/ 0.0102	0.5632/ 0.0041	0.4060/ 0.0030
384	C	C	C	T	C	C	C	C	C	C	C	C	C	0.2755/ 0.0019	0.6066/ 0.0044	0.0153/ 0.0001

Table 8 (Continued)

Site	Haplotype													Association (F value/r^2)		
	A[a]	B	C	D	E	F	G	H	I	J	K	L	M	AAC	GT	GC
397	A	G	A	G	A	A	G	G	A	G	G	G	G	16.4522/ 0.101***	5.5547/ 0.0358*	13.5488/ 0.0894**
436	G	-	G	G	G	G	-	G	G	-	G	G	-	15.7225/ 0.0970**	1.7928/ 0.0119	17.683/ 0.1132**
448	C	T	C	T	C	C	T	T	C	T	T	T	T	16.4522/ 0.101***	5.5547/ 0.0358*	13.5488/ 0.0894**
451	A	C	A	C	A	A	C	C	A	C	C	C	C	16.4522/ 0.101***	5.5547/ 0.0358*	13.5488/ 0.0894**
461	T	G	T	A	T	T	G	G	T	G	G	G	G	9.0963/ 0.1111**	2.9622/ 0.0383	7.1239/ 0.0943**
485	C	T	C	T	C	C	T	T	C	T	T	T	T	16.4522/ 0.101***	5.5547/ 0.0358*	13.5488/ 0.0894**
487	T	C	T	C	T	T	C	C	T	C	C	C	C	16.4522/ 0.101***	5.5547/ 0.0358*	13.5488/ 0.0894**
490	C	C	C	C	C	C	T	C	C	T	C	C	C	20.4421/ 0.122***	8.5955/ 0.0540**	30.8029/ 0.1798***

*Significant difference at $P < 0.05$ level

**Significant difference at $P < 0.01$ level

***Significant difference at $P < 0.001$ level

[a] The nucleotide site position refers to the sequence order of the EF990806

9.2 SSSIIIa locus

None of the 22 SNPs and 5 Indels were significantly associated with GT. Five polymorphisms were significantly associated with AAC and three SNPs were significantly associated with GC at $P < 0.05$ (Table 9).

Table 9 Nucleotide polymorphisms in the SSSIIIa locus associated with the AAC, GT and GC.

Site	Haplotype														Association (F value/r^2)		
	A	B	C	D	E	F	G	H[a]	I	J	K	L	M	N	AAC	GT	GC
32	T	T	T	T	T	T	T	T	T	T	T	T	T	C	0.1662/ 0.0010	1.2735/ 0.0085	0.6443/ 0.0050
72-73 AG	+	+	+	+	+	+	+	+	+	+	+	+	-	+	0.0024/ 0	1.9439/ 0.0129	2.7256/ 0.0208
100-107 $(AT)_n$	3	3	3	3	3	3	3	4	3	3	3	3	3	3	0.5181/ 0.0032	3.7539/ 0.0245	0.1142/ 0.0009
187	A	A	G	A	A	A	A	A	A	A	A	A	A	A	4.0173/ 0.024*	0.1728/ 0.0012	1.8022/ 0.0139
190	A	A	G	A	A	A	A	A	A	A	A	A	A	A	4.0173/ 0.024*	0.1728/ 0.0012	1.8022/ 0.0139
241	T	T	T	A	T	T	T	T	T	T	T	T	T	T	1.0890/ 0.0067	01356/ 0.0009	0.1760/ 0.0014
246	G	G	G	G	G	G	T	G	G	G	G	G	G	G	0.1702/ 0.0010	0.1104/ 0.0007	2.3359/ 0.0179
290	C	C	C	G	C	C	C	C	C	G	C	C	C	C	0.9950/ 0.0061	0.1395/ 0.0009	0.0369/ 0.0003
310	G	A	G	G	A	G	G	G	G	G	G	G	A	G	1.5833/ 0.0096	0.0954/ 0.0006	3.6739/ 0.0279
327	C	C	C	C	C	T	C	C	C	C	C	C	C	C	3.0985/ 0.0187	0.0764/ 0.0005	5.4408/ 0.0407*
339	-	-	-	-	-	-	-	-	-	-	-	G	-	-	1.3274/ 0.0081	0.4286/ 0.0029	0.3680/ 0.0029
341	C	C	C	C	C	C	C	C	C	C	T	C	C	C	4.9917/ 0.297*	3.0729/ 0.0202	0.0999/ 0.0008
343	A	G	G	G	G	G	G	G	G	G	A	A	G	G	0.0163/ 0.0001	0.3143/ 0.0021	1.1100/ 00086
361	C	C	C	C	C	T	C	C	C	C	C	C	C	C	3.0985/ 00187	0.0764/ 0.0005	5.4408/ 0.0407*
365	A	G	G	G	G	G	G	G	G	G	G	G	G	G	0.2748/ 0.0017	0.0011/ 0	0.9598/ 0.0074
366-368 AAG	+	+	+	+	-	+	+	+	+	+	+	+	+	+	0.0724/ 0.0004	0.1096/ 0.0007	0.3648/ 0.0028
372	C	T	T	T	T	T	T	C	T	T	C	C	T	T	0.0234/ 0.0001	1.5512/ 0.0103	1.3471/ 0.0104
408	G	G	G	G	G	G	G	G	G	G	G	-	G	G	1.3274/ 0.0081	0.4286/ 0.0029	0.3680/ 0.0029
410	T	T	T	T	T	T	T	T	T	T	T	G	T	T	1.3274/ 0.0081	0.4286/ 0.0029	0.3680/ 0.0029
411	T	T	T	T	T	T	T	T	T	T	T	A	T	T	1.3274/ 0.0081	0.4286/ 0.0029	0.3680/ 0.0029
422	T	A	T	T	A	T	T	T	T	T	T	T	A	T	1.5833/ 0.0096	0.0954/ 0.0006	3.6739/ 0.0279
423	T	T	T	T	T	T	T	T	T	T	C	T	T	T	4.9917/ 0.297*	3.0729/ 0.0202	0.0999/ 0.0008
427	C	C	C	C	C	C	T	C	C	C	C	C	C	C	0.1560/ 0.0010	0.1541/ 0.0010	1.8204/ 0.0140
441	-	T	-	-	-	-	-	-	-	-	-	-	-	-	4.8319/ 0.0287*	1.3111/ 0.0087	3.7488/ 0.0284

Table 9 (Continued)

Site	Haplotype														Association (F value/r^2)		
	A	B	C	D	E	F	G	H[a]	I	J	K	L	M	N	AAC	GT	GC
452	C	C	C	C	C	T	C	C	C	C	C	C	C	C	3.0985/ 0.0187	0.0764/ 0.0005	5.4408/ 0.0407*
457	T	T	T	T	T	T	T	T	T	T	T	C	T	T	1.3274/ 0.0081	0.4286/ 0.0029	0.3680/ 0.0029
467	C	C	C	T	C	C	C	C	C	T	C	C	C	C	1.0371/ 0.0063	0.1369/ 0.0009	0.0220/ 0.0002

*Significant difference at $P < 0.05$ level

[a] The nucleotide site position refers to the sequence order of the EF990829

9.3 SSSIIIb locus

None of the six SNPs in this gene were significantly associated with any of the eating and cooking quality parameters (AAC, GT and GC) at $P < 0.05$ (Table 10).

Table 10 Nucleotide polymorphisms in the SSSIIIb locus associated with the AAC, GT and GC.

Site	Haplotype								Association (F value/r^2)		
	A[a]	B	C	D	E	F	G	H	AAC	GT	GC
55	C	C	C	T	C	T	C	C	0.1362/0.0008	0.0373/0.0003	0.2898/0.0023
212	T	T	A	T	T	T	T	T	0.0333/0.0002	1.9540/0.0129	0.1873/0.0015
288	G	A	G	A	A	G	A	G	0.0011/0	1.6967/0.0113	02221/0.0017
369	T	G	G	G	G	G	G	G	0.0189/0.0001	0.2295/0.0015	0.4726/0.0037
511	G	G	G	G	G	G	A	G	0.2457/0.0015	0.0742/0.0005	0.1246/0.0010
533	C	C	C	A	A	A	A	C	0.3667/00023	0.0995/0.0007	0.0955/0.0007

[a] The nucleotide site position refers to the sequence order of the EF990843

9.4 SSSIVa locus

For this locus, five polymorphisms were a strongly associated with AAC at $P < 0.01$. Nine polymorphisms were significantly associated with GT at $P < 0.05$, and none of the 53 polymorphisms were significantly associated with GC at $P < 0.05$ (Table 11). Haplotype block were located between at site 174 – 198 because three polymorphisms were same F value.

Table 11 Nucleotide polymorphisms in the SSSIVa locus associated with the AAC, GT and GC.

Site	Haplotype								Association (F value/i^2)		
	A	B	C	D	E	F	G[a]	L	AAC	GT	GC
37	T	C	T	T	T	T	T	T	3.2141/0.0193	1.4525/.00097	1.3118/0.0101
43	T	C	C	C	T	C	T	C	0.4590/0.0028	0.0071/0	0.6317/0.0049
97	A	T	C	C	A	C	A	C	1.7555/0.0212	0.7366/0.0099	0.9032/0.0140
102	C	T	T	T	C	T	C	T	0.4502/0.0028	0.0071/0	0.6317/0.0048
104	T	T	C	T	T	C	T	T	0.1429/0.0009	0.0600/0.0004	0.2534/0.0020
108	C	C	T	C	C	T	C	C	0.1429/0.0009	0.0600/0.0004	0.2534/0.0020
116	C	T	T	C	C	T	C	C	0.0071/0	0.0069/0	0.6905/0.0054
118	C	T	T	C	C	T	C	T	2.0207/0.0123	0.9959/0.0066	1.7522/0.0135
121	C	C	C	C	C	C	T	C	0/0	0.4457/0.0030	0.0004/0
126	A	T	T	T	C	T	C	T	0.4875/0.0060	0.8167/0.0109	0.3134/0.0049
130	C	T	C	C	C	C	C	C	3.2142/0.0193	1.4525/0.0097	1.3118/0.0101
135-137 TAA	-	-	-	-	-	-	+	-	0.0002/0	0.4457/0.0030	0.0004/0
152	C	T	T	C	C	T	C	T	2.0207/0.0123	0.9959/0.0066	1.7522/0.0135
174-176 TAC	+	+	+	-	+	+	+	+	1.7861/0.0109	4.6279/0.0300*	1.5546/0.0120
190	G	G	G	A	G	G	G	G	1.7861/0.0109	4.6279/0.0300*	1.5546/0.0120
198	C	C	C	A	C	C	C	C	1.7861/0.0109	4.6279/0.0300*	1.5546/0.0120
200	G	A	G	G	G	G	G	G	3.2142/0.0193	1.4525/0.0097	1.3118/0.0101
202	C	T	T	C	C	T	C	C	0.0071/0	0.0069/0	0.6905/0.0054
203	A	A	G	G	A	G	A	A	1.2052/0.0074	1.6471/0.0109	0.0316/0.0002
207-208 CAATG	-	+	-	-	-	-	-	-	3.2141/0.0193	1.4525/.00097	1.3118/0.0101
210	T	C	C	C	T	C	T	C	0.4590/0.0028	0.0071/0	0.6317/0.0049

Table 11 (Continued)

Site	Haplotype								Association (F value/r^2)		
	A	B	C	D	E	F	G[a]	L	AAC	GT	GC
222	G	A	A	A	G	A	G	A	0.7224/0.0050	0.0583/0.0004	00008/0
228	G	A	A	G	G	A	G	G	0.0071/0	0.0069/0	0.6905/0.0054
236	G	G	G	A	G	G	G	G	1.7861/0.0109	4.6279/0.0300*	1.5546/0.0120
249	C	A	A	C	C	A	C	C	0.0071/0	0.0069/0	0.6905/0.0054
269	C	C	C	A	C	C	C	C	1.7861/0.0109	4.6279/0.0300*	1.5546/0.0120
309	A	T	T	T	A	T	A	T	0.4502/0.0028	0.0071/0	0.6317/0.0048
313	G	G	A	G	G	G	G	G	0.8698/0.0053	07314/0.0049	1.0530/0.0082
323	T	C	C	C	T	C	T	C	0.4590/0.0028	0.0071/0	0.6317/0.0049
347	G	G	G	G	G	G	G	A	6.8795/0.0403**	3.3922/0.0222	1.4304/0.0111
359	T	C	T	T	C	T	C	C	0.9946/0.0061	0.6054/0.0041	0.1397/0.0011
367	A	G	G	G	A	G	A	G	0.4590/0.0028	0.0071/0	0.6317/0.0049
374	G	G	G	A	G	G	G	G	1.7861/0.0109	4.6279/0.0300*	1.5546/0.0120
382	C	C	C	C	C	C	C	-	6.8795/0.0403**	3.3922/0.0222	1.4304/0.0111
393	T	T	C	C	T	C	T	C	0.0417/0.8385	01694/0.0011	0.2168/0.0017
412	G	G	G	G	G	G	G	A	6.8795/0.0403**	3.3922/0.0222	1.4304/0.0111
433	C	C	C	C	C	C	T	C	0/0	0.4457/0.0030	0.0004/0
448	G	G	A	G	G	A	G	G	0.1429/0.0009	0.0600/0.0004	0.2534/0.0020
459	T	C	C	C	T	C	T	C	0.4590/0.0028	0.0071/0	0.6317/0.0049
465	C	C	C	T	C	C	C	C	1.7861/0.0109	4.6279/0.0300*	1.5546/0.0120
470	A	G	G	G	A	G	A	G	0.4590/0.0028	0.0071/0	0.6317/0.0049
471	C	T	T	C	C	T	C	C	0.0071/0	0.0069/0	0.6905/0.0054
489	A	G	G	A	A	G	A	G	2.0207/0.0123	0.9959/0.0066	1.7522/0.0135
492	A	G	G	G	G	G	G	G	0.0625/0.0004	2.6081/0.0188	0.1352/0.0011
496	G	G	G	A	G	G	G	G	1.7861/0.0109	4.6279/0.0300*	1.5546/0.0120
499	C	C	C	C	C	C	C	T	6.8795/0.0403**	3.3922/0.0222	1.4304/0.0111
503-509 CATGT GA	+	-	-	+	+	-	+	+	0.0071/0	0.0069/0	0.6905/0.0054
514 AA	+	-	-	+	+	-	+	+	1.2834/0.0156	2.2994/0.0301	1.1008/0.0170
519 Indel 30bp	+	+	+	-	+	+	+	+	0.9567/0.0117	3.0557/0.0395	0.8381/0.0130
549	G	A	A	G	G	A	G	A	2.0207/0.0123	0.9959/0.0066	1.7522/0.0135
559	T	T	T	C	T	T	T	T	1.7861/0.0109	4.6279/0.0300*	1.5546/0.0120

Table 11 (Continued)

Site	Haplotype								Association (F value/r^2)		
	A	B	C	D	E	F	G[a]	L	AAC	GT	GC
569	C	C	C	C	C	C	C	T	6.8795/0.0403**	3.3922/0.0222	1.4304/0.0111
573	T	T	A	T	T	A	T	T	0.1429/0.0009	0.0600/0.0004	0.2534/0.0020

*, ** Significant difference at $P < 0.05$ and $P < 0.01$, respectively

[a] The nucleotide site position refers to the sequence order of the EF990859

9.5 SSSIVb locus

None of the seven polymorphisms were significantly associated with GT at $P < 0.05$. However, one indel of 22-bp at site 220 and one SNP (C/T at site 418) were significantly associated with AAC and GC at $P < 0.05$ (Table 12). The SNP at site 418 was the only SNP in the coding region (non-synonymous).

Table 12 Nucleotide polymorphisms in the SSSIVb locus associated with the AAC, GT and GC.

Site	Haplotype							Association (F value/i^2)		
	A	B[a]	C	D	E	F	I	AAC	GT	GC
52 $(GGGTT)_n$	3	4	3	4	4	3	3	0.0751/0.0005	0.0001/0	0.3622/0.0027
118	T	C	C	T	T	T	T	0.1281/0.0008	0.3251/0.0022	1.4082/0.0102
127	G	A	G	A	A	G	G	0.0751/0.0005	0.0001/0	0.3622/0.0027
220 Indel 22 bp	+	+	+	-	+	+	+	4.4434/0.0265*	0.7938/0.0053	5.7591/0.0404*
281	T	T	A	T	A	A	T	0.6953/0.0043	0.3408/0.0023	2.4626/0.0178
349	G	A	G	A	A	G	A	0.0751/0.0005	0.0001/0	0.3622/0.0027
418	C	C	C	T	C	C	T	4.4434/0.0265*	0.7938/0.0053	5.7591/0.0404*

*Significant difference at $P < 0.05$ level

[a] The nucleotide site position refers to the sequence order of the EF990884

9.6 RBE1 locus

None of the seven polymorphisms was significantly associated with AAC or GC at $P < 0.05$. One SNP (C/A at site 538) was significantly associated with GT at $P < 0.05$ (Table 13).

Table 13 Nucleotide polymorphisms in the RBE1 locus associated with the AAC, GT and GC.

Site	Haplotype					Association (F value/r^2)		
	A[a]	B	C	D	E	AAC	GT	GC
224	T	T	-	T	-	0.2521/0.0016	0.1249/0.0008	0.0185/0.0001
508	G	G	T	G	C	0.4336/0.0053	00755/0.0010	0.0617/0.0010
534	T	A	T	T	T	0.0022/0	1.8281/0.0121	3.7827/0.0287
538	A	C	C	C	C	0.2683/0.0017	4.0922/0.0291*	1.955/0.0131

*Significant difference at $P < 0.05$ level

[a] The nucleotide site position refers to the sequence order of the EF990867

9.7 RBE3 locus

None of the four SNPs was significantly associated with any of the eating and cooking quality parameters (AAC, GT and GC) at $P < 0.05$ (Table 14).

Table 14 Nucleotide polymorphisms in the RBE3 locus associated with the AAC, GT and GC.

Site	Haplotype					Association (F value/r^2)		
	A[a]	B	C	D	E	AAC	GT	GC
135	G	C	C	G	G	0.0102/0.0001	0.1728/0.0012	0.3527/0.0028
284	G	G	C	C	G	0.0361/0.0002	0.0742/0.0005	0.1246/0.0010
407	G	G	G	G	A	0.9331/0.0057	0.1139/0.0008	03636/0.0028
532	G	G	A	A	G	0.0121/0.0001	0.0101/0.0001	0.5930/0.0046

[a] The nucleotide site position refers to the sequence order of the EF990877

Summary, 13, 5, 5 and 1 polymorphisms from GBSSI, SSSIIIa, SSSIVa and SSSIVb, respectively, were significantly associated with AAC, especially, 7 polymorphisms from GBSSI were significant at $P < 0.05$. Whereas, 10, 9, 7 and 1 polymorphisms from GBSSI, SSSIVa SSSIVb and RBE1, respectively, were significantly associated with GT, especially, one polymorphism from GBSSI were significant at $P < 0.01$. Finally, 11, 3 and 1 polymorphisms from GBSSI, SSSIIIa and SSSIVb, respectively, were significantly associated with GC, especially, 2 polymorphisms from GBSSI were significant at $P < 0.0001$. However, r^2 is still low, indicating that a lot of variation is still not explained.

In association mapping, alleles of the candidate gene can be tested for association with a particular phenotype. The variation predicted to have a functional consequence, such as causing changes in amino acid or in the level of gene expression, should be the first choice for the association analysis. It is possible to use LD to identify actual functional polymorphisms within the candidate gene. Recent reports showed four nonsynonymous mutations at the base positions 264, 1,810, 2,209, and 2,340/2,341 in the exons of the *SSIIa* gene by Nakamura *et al.* (2005) and three SNPs in exon 8 by Bao *et al.* (2006) are strongly associated with chain-length distribution of amylopectin and the GT of rice flour.

The association of *GBSSI* SSR and *GBSSI* SNPs with AAC is easy to understand because it is the granule-bound starch synthase encoded by *GBSSI* that controls the synthesis of amylose (Ayres *et al.*, 1997). QTL mapping has consistently confirmed that *GBSSI* is a major locus contributing to the variation of AAC in rice grain (He *et al.*, 1999; Bao *et al*., 2004; Tan *et al.*, 1999). The present study has shown that thirteen polymorphisms in 5'upstram region of *GBSSI* were significantly associated with AAC at $P < 0.05$, of which the six SNPs showed a very strong association at $P < 0.0001$, so the *GBSSI* was the most important in controlling the genetic basis of AAC, agreeing with those results from QTL mapping.

SNPs and InDels are very useful genetic markers for LD analysis and association studies, and also for marker-assisted selection (MAS) (Rafalski 2002). Some of the SNPs discovered in genes and expressed sequence tags (EST) are

functional markers that may causally affect phenotypic trait variation (quantitative trait nucleotide). These functional DNA markers are superior to other markers owing to their complete linkages with the alleles at the trait loci (Andersen and Lubberstedt 2003). Such markers can be directly used in MAS without prior mapping if a relationship between the marker alleles and phenotypic variation has been established. For example, Bao *et al.* (2006) found the relationship between the GC/TT marker alleles and GT holds in 90% (27/30 sequenced rices) to 94% (479/509) genotyped rice samples. This functional GC/TT SNP marker can therefore be used in diagnostic analysis to predict whether the rice's starch GT is high or low. It can also be used for MAS in rice breeding programs if the parents have different GC/TT alleles as well as divergent GTs.

10. Gene expression analysis

mRNA was extracted from endosperm and leaves and cDNA fragments were PCR-amplified using the specific primers. Details of all primers are presented in Table 15.

GBSSI gene should be expressed only in endosperm, but bands corresponding to the expected mRNA appeared in both seed and leaf RNA as did an additional band that was the same size as genomic DNA. Similarly, for *SSSIVb* a band of the same size as genomic DNA was observed in RNA from leaf and seed (Figure 32). Maybe, there were alternative splicing because many matured mRNAs from higher eukaryote genes can be spliced in alternative ways and generate two or more different mRNAs. In this case, introns might have been retained and the splicing regulation that generated different forms of mRNA at different times, under different condition, or in different cell or tissue type. Since the RNA was extracted from mixed immature seeds may be generated at different times.

DNA fragments from *SSSIIIa*, *SSSIIIb*, *SSSIVa* and *RBE1* showed the correct size (Figure 33, 34). For the RBE3 gene, DNA fragments could not be amplified from genomic DNA templates from both samples but a fragment could be amplified from mRNA templates of seed and had the correct size (Figure 35).

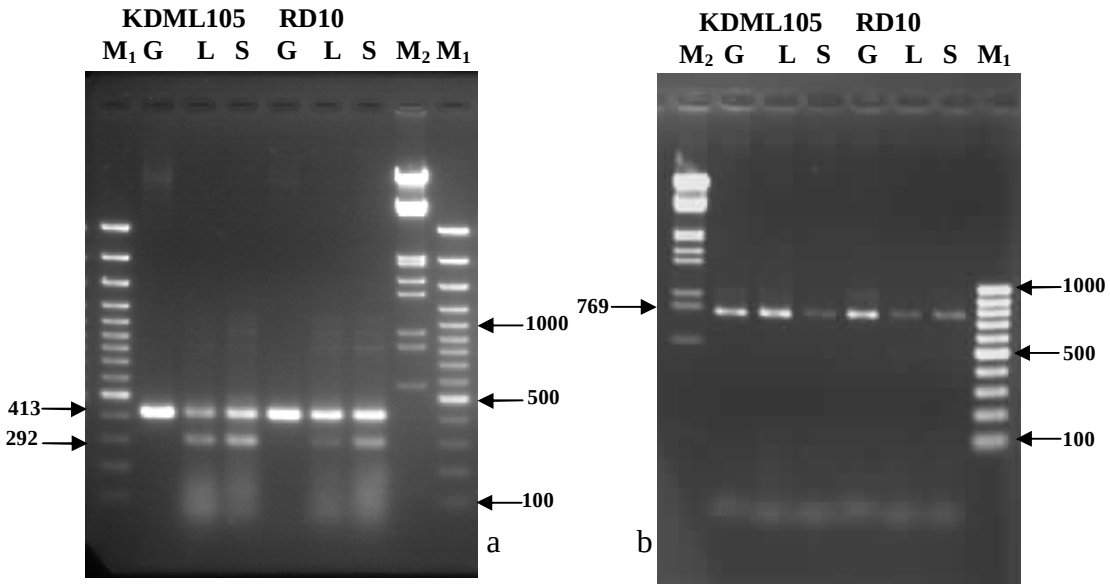


Figure 32 PCR products amplified from genomic DNA and mRNA of leaf and seed with GBSSI (a) and SSSIVb (b) primers. M_1 = size standard of 100 bp ladder, M_2 = size standard of λ/*Hin*dIII+*Eco*RI, G = genomic DNA, L = mRNA from leaf, S = mRNA from seed.

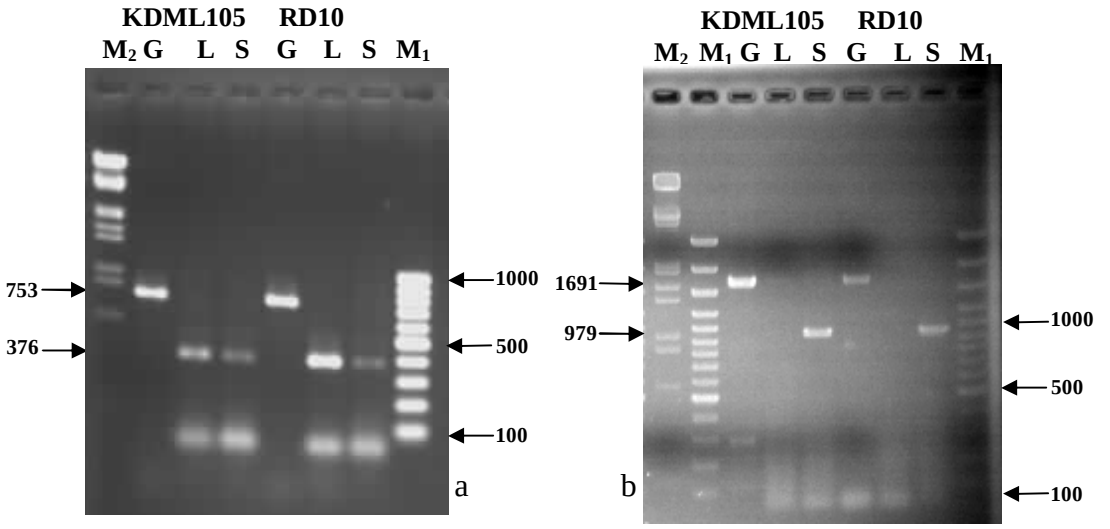


Figure 33 PCR products amplified from genomic DNA and mRNA of leaf and seed with SSSIIIa (a) and SSSIIIb (b) primers. M_1 = size standard of 100 bp ladder, M_2 = size standard of λ/*Hin*dIII+*Eco*RI, G = genomic DNA, L = mRNA from leaf, S = mRNA from seed.

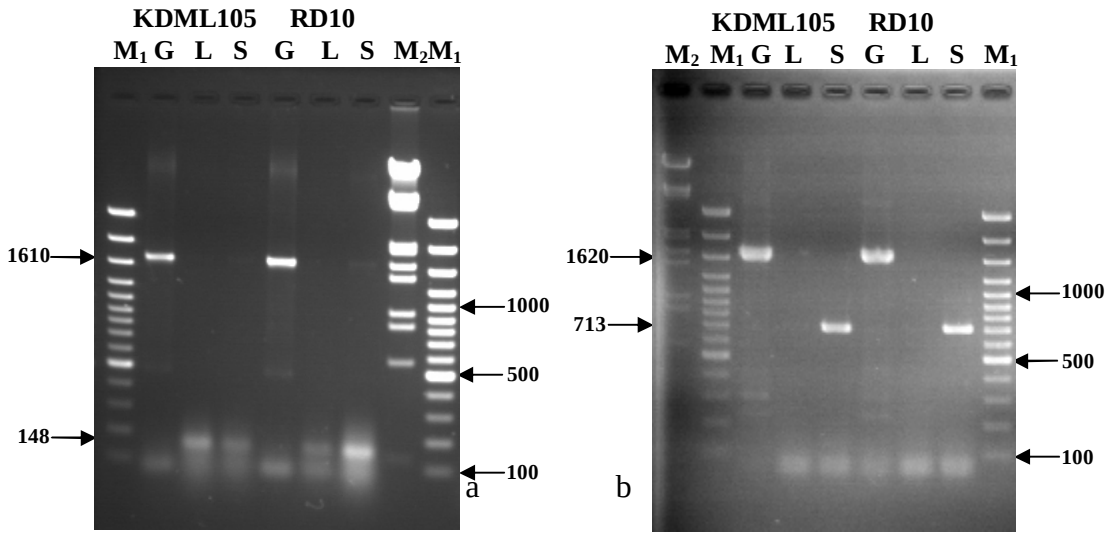


Figure 34 PCR products amplified from genomic DNA and mRNA of leaf and seed with SSSIVa (a) and RBE1 (b) primers. M_1 = size standard of 100 bp ladder, M_2 = size standard of λ/*Hin*dIII+*Eco*RI, G = genomic DNA, L = mRNA from leaf, S = mRNA from seed.

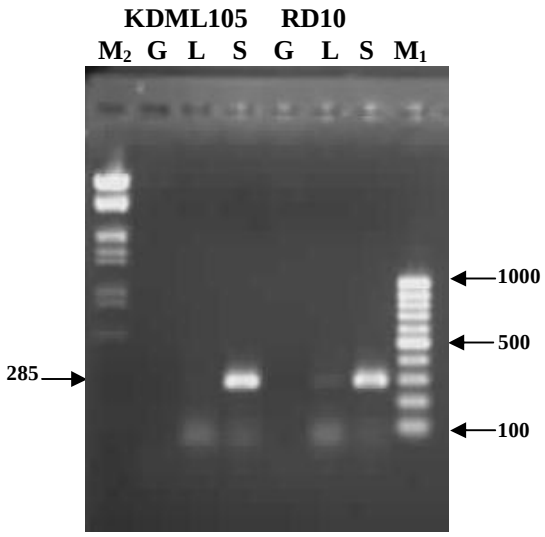


Figure 35 PCR products amplified from genomic DNA and mRNA of leaf and seed with RBE3 primers. M_1 = size standard of 100 bp ladder, M_2 = size standard of λ/*Hin*dIII+*Eco*RI, G = genomic DNA, L = mRNA from leaf, S = mRNA from seed.

Table 15 Gene-specific PCR primers used for gene expression analysis.

Gene	Accession Number	Exon		Size (bp)		Forward Primer	Reverse Primer
		For.	Rev.	gDNA	mRNA		
GBSSI	AB002542	7	8	413	292	TTCTgCATCCACAACATCTC	TgACgTCCATgCCgTTgAC
SSS IIIa	AF432915	10	12	753	376	g ggTCAAgTgCTCCAgTTgCTTg	ACAACAggggTATCAgTTTgCTg
SSS IIIb	AP005441	1	3	1691	979	CACCgAAAgACTgCACAAgA	AACCAAACCTCTggCTTTCC
SSS IVa	AP003292	5	6	1610	148	CACATTgTCCATATAgCAgCAgAgATg	ggTCAAgTTgCATACAgTCgTATTTgg
SSS IVb	AC121365	7	10	769	471	ACAAAATATggACTgggACTgTTgAAg	CTgATCgCACCTCTAgTgCATATg
RBE1	AP004685	6	9	1620	713	TTCAACTATgCyAAyTgggA	CTgCATTTATCATAgCTCCA
RBE3	AP004879	4	8	1287	285	TTTTCTCgCggTTATgAgAAgTTTgg	CACAgAgTACTTgATCCAggCAgg

CONCLUSION AND RECOMMENDATION

Conclusion

From the experimental results and discussion of this study, the following conclusions can be drawn:

1. This is the first report of comparison between starch physico-chemical properties and haplotypes of starch synthesis genes in Thai rice. The sensitivity of the SSCP method to detect single nucleotide substitution mutations or insertion/deletion mutations is demonstrated here by the large number of alleles that could be detected at each locus. We found the ANOVA results were significant ($P < 0.05$) at all loci for at least one of the physico-chemical properties implying that there was at least one haplotype which showed a means different from other haplotypes. The present information can be used to develop molecular marker for marker-assisted selection in breeding program.

2. One hundred and ninety-four accessions of *Oryza* were assayed using PCR-SSCP and alleles at seven starch synthesis genes loci (GBSSI, SSSIIIa, SSSIIIb, SSSIVa, SSSIVb, RBE1 and RBE3) were identified. The nucleotide sequences corresponded to each of the observed SSCP patterns. The frequency of nucleotide polymorphism in this study is about one SNP per 29 bp and thus the *O. sativa* from Thailand maintains a very high nucleotide diversity, higher than in another recent study.

3. At the (CT)n region in the GBSSI locus, four new alleles were observed $(CT)_7$, $(CT)_9$, $(CT)_{10}$ and $(CT)_{12}$. Allele $(CT)_7$ was present only in Thai rice and the Indian rice accession (waxy rice). The result of relationship between repeat number polymorphism and waxy-rice were not conclusive because there was conflict between results from present and previous studies and showed that this SSR are not useful in marker-assisted breeding for the improvement of starch quality of rice.

4. The haplotype networks did not show the main mutation corresponding with eating and cooking quality, geography, or cultivation. However, *O. glaberrima* in all networks were placed far out from other haplotype excepted in the SSSIVb gene that *O. glaberrima* separated other haplotype into two groups.

5. We found that LD in 7 genes loci generally decayed rapidly with distance between sites within loci. Two LD block observed in GBSSI and SSIV-1 loci and four LD groups, including polymorphism from different loci, were detected at $r^2 > 0.2$ and $P < 0.05$ indicating relationship between haplotype of seven gene loci.

6. Association tests were performed to investigate the relationships between the SNPs found in five loci and AAC, GT and GC. Twenty-four polymorphisms from four loci were significantly associated with AAC, twenty-seven polymorphisms from four loci were significant with GT and fifteen polymorphisms from three loci were significantly associated with GC. The SSSIIIB and RBE3 did not have polymorphism that significantly associated with the eating and cooking quality parameters (AAC, GT and GC)

Recommendation

Out of six SNPs showing significant relationships and two SNPs (indel (G) at site 436 and C/T at site 490) in GBSSI locus showed a very strong association with AAC and GC, so these could be used for marker-assisted selection (MAS) in rice breeding programs.

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.

Andersen, J.R. and T. Lübberstedt. 2003. Function markers in plants. **Trends Plant Sci.** 8: 554-560.

_____, T. Schrag, A.E. Melchinger, I. Zein and T. Lübberstedt. 2005. Validation of Dwarf8polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). **Theor. Appl. Genet.** 111: 206-217.

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.

Ball, S.G., H.P. Guan, M. James, A. Myers, P. Keeling, G. Mouille, A. Buleon, P. Colonna and J. Preiss. 1996. From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. **Cell** 86: 349-352.

_____, M.H.B.J. van de Wal and R.G.F. Visser. 1998. Progress in understanding the biosynthesis of amylose. **Trends Plant Sci.** 3: 462-467.

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

_____, M. Sun, L. Zhu and Harold Corke. 2004. Analysis of quantitative trait loci for some starch properties of rice (*Oryza sativa* L.): thermal properties, gel texture and swelling volume. **J. Cereal Sci.** 39: 379-385.

_____, 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.

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

_____, _____ and P.M. Gresshoff. 1991. Fast and Sensitive Silver Staining of DNA in Polyacrylamide Gels. **Analytical Biochem.** 196: 80-83.

Bradbury, L.M.T., R.J. Henry, Q. Jin, R.F. Reinke and D.L.E. Waters. 2005. A perfect marker for fragrance genotyping in rice. **Mol. Breeding** 16: 279-283.

Bundock, P.C. and R.J. Henry. 2004. Single nucleotide polymorphism, haplotype diversity and recombination in the Isa gene of barley. **Theor. Appl. Genet.** 109: 543-551.

Cagampang, G.B., C.M. Perez and B.O. Juliano. 1973. A gel consistency test of eating quality of rice. **J. Sci. Food Agr.** 24: 1589-1594.

Caicedo, A.L. and B.A. Schaal. 2004. Population structure and phylogeography of *Solanum pimpinellifolium* inferred from a nuclear gene. **Mol. Ecol.** 13: 1871-1882.

Causse, M.A., T.M. Fulton, Y.G. Cho, S.N. Ahn, J. Chunwonges, K.S. Wu, J.H. Xiao, Z.H. Yu, P.C. Ronald, S.E. Harrington, G. Second, S.R. McCouch and S.D. Tanksley. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. **Genetics** 138: 1251-1274.

Chuayjaeng, S. 2005. **Chloroplast diversity and phylogeny in wild and cultivated rice (*Oryza* spp.)**. M.S. Thesis, Kasetsart University.

Clement, M., D. Posada and K.A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. **Mol. Ecol.** 9: 1657-1659.

Dey, P.M. and J.B. Harborne. 1997. **Plant Biochem.** Academic Press, California.

Dian, W., H. Jiang and P. Wu. 2005. Evolution and expression analysis of starch synthase III and IV in rice. **J. Experment. Bot.** 56: 623-632.

Du, F.X., A.C. Clutter and M.M. Lohuis. 2007. Characterizing linkage disequilibrium in pig populations. **Inter. J. Biol. Sci.** 3: 166-178.

Evanno, G., S. Regnaut and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Mol. Ecol.** 14: 2611-2620.

Falush, D., M. Stephens and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. **Genetics** 164: 1567-1587.

Flint-Garcia S.A., J.M. Thornsberry and E.S. 4th Buckler. 2003. Structure of linkage disequilibrium in plants. **Annu. Rev. Plant Biol.** 54: 357-374.

Fukuoka, S., N.V. Alpatyeva, K. Ebana, N.T. Luu and T. Nagamine. 2003. Analysis of Vietnamese rice germplasm provides and insight into Japonica rice differentiation. **Plant Breeding** 122: 497-502.

Glaszmann, J.C. 1987. Isozymes and classification of rice varieties. **Theor. Appl. Genet.** 74: 21-30.

Guan, H. and J. Preiss. 1993. Differentiation of the properties of the branching isozymes from maize (*Zea mays*). **Plant Physiol.** 102: 1269-1273.

Gupta, P.K., S. Rustgi and P.L. Kulwal. 2005. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant Mol. Biol.** 57: 461-485.

Hagenblad, J., C. Tang, J. Molitor, J. Werner, K. Zhao, H. Zheng, P. Marjoram, D. Weigel and M. Nordborg. 2004. Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. **Genetics** 168: 1627-1638.

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 systematics.** 2nd ed. Sinauer Associates, Inc., Massachusetts.

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

Isshiki, M., K. Morino, M. Nakajima, R.J. Okagaki, S.R. Wessler, T. Izawa and K. Shimamoto. 1998. A naturally occurring functional allele of the rice *waxy* locus has a GT to TT mutation at the 5' splice site if the first intron. **Plant J.** 15: 133-138.

Iwata, H., Y. Uga, Y. Yoshioka, K. Ebana and Takeshi Hayashi. 2007. Bayesian association mapping of multiple quantitative trait loci and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms. **Theor. Appl. Genet.** 114: 1437-1449.

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

Juliano, B.O. 1985. **Rice chemistry and technology**. 2nd ed. American Association of Cereal Chemists, Incorporated Saint Paul, Minnesota, USA.

Kanno, A. and A. Hirai. 1993. A transcription map of the chloroplast genome from rice (*Oryza sativa*). **Curr. Genet.** 23: 166-174.

Li, Z., X. Chu, G. Mouille, L. Yan, B. Kosar-Hashemi, S. Hey, J. Napier, P. Shewry, B. Clarke, R. Appels, M.K. Morell and S. Rahman. 1999. The localization and expression of the class II starch synthases of wheat. **Plant Physiol.** 120: 1147-1156.

Little, R.R., G.B. Hilder and E.H. Dawson. 1958. Differential effect of dilute alkalia in 25 varieties of milled white rice. **Cereal Chem.** 35: 111-126.

Mackill, D.J. 1999b. Genome analysis and breeding, pp. 17-41. In K. Shimamoto, ed. **Molecular Biology of Rice.** Springer-Verlag, Tokyo. Cited N. Kurata, Y. Nakamura, K. Yamamoto, Y. Harushima, N. Sue, J. Wu, B.A. Antonio, A. Shomura, T. Shimizu, S.Y. Lin, T. Inoue, A. Fukuda, T. Shimano, Y. Kuboki, T. Toyama, Y. Miyamoto, T. Kirihara, K. Hayasaka, A. Miyao, L. Monna, H.S. Zhong, Y. Tamura, Z.X. Wang, T. Momma, Y. Umehara, M. Yano, T. Sasaki and Y. Minobe. 1994. A 300 kilobase interval genetic map of rice including 883 expressed sequences. **Nat. Genet.** 8: 365-372.

Martin, C., A. Juliano, H.J. Newbury, B.R. Lu, M.T. Jackson and B.V. Ford-Lloyd. 1997. The use of RAPD markers to facilitate the identification of *Oryza* species within a germplasm collection. **Genet. Resources Crop Evol.** 44: 175-183.

_____ and A.M. Smith. 1995. Starch biosynthesis. **Plant Cell** 7: 971-985.

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.

Morishima, H. 1997. Isozyme and storage protein in relation to genome constitution, pp. 54-68. In T. Matsuo, Y. Futsuhara, F. Kikuchi and H. Yamaguhi, eds. **Science of the rice plant**, vol. 3. Genetics, Food and Agriculture Policy Center, Tokyo.

Myers, A.M., M.K. Morell, M.G. James and S.G. Ball. 2000. Recent progress toward understanding biosynthesis of the amylopectin crystal. **Plant Physiol.** 122: 989–997.

Nakamura, I., H. Urairong, N. Kameya, Y. Fukuta, S. Chitarkon and Y.I. Sato. 1998. Six different plastid subtypes were found in *O. sativa-O. rufipogon* complex. **Rice Genet. Newsl.** 15: 80-82.

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.

_____, P.B. Francisco Jr., Y. Hosaka, A. Sato, T. Sawada, A. Kubo and N. Fujita. 2005. Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between japonica and indica rice varieties. **Plant Mol. Biol.** 58: 213-227.

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.

Nordborg, M. and S. Tavaré. Linkage disequilibrium: what history has to tell us. **Trends Genet.** 18: 83-90.

Olsen, K.M. and M.D. Purugganan. 2002. Molecular evidence on the origin and evolution of glutinous rice. **Genetics** 162: 941-950.

_____ and B.A. Schaal. 1999. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. **Proc. Nalt. Acad. Sci. U.S.A.** 96: 5586-5591.

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.

Prashanth, S.R., M. Parani, B.P. Mohanty, V. Talame, R. Tuberosa, A. Parida. 2002. Genetic diversity in cultivars and landraces of Oryza sativa subsp. indica as revealed by AFLP markers. **Genome** 45: 451-459.

Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. **Genetics** 155: 945-959.

Rafalski, A. 2002. Applications of single mucleotide polymorphism in crop genetics. **Curr. Opin. Plant Biol.** 5: 94-100.

Rakshit, S., A. Rakshit, H. Matsumura, Y. Takahashi, Y. Hasegawa, A. Ito, T. Ishii, N.T. Miyashita and R. Terauchi. 2007. Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. **Theor. Appl. Genet.** 114: 731–743.

Rahman, S., A. Ragina, Z. Li, Y. Mukai, M. Yamamoto, B. Kosar-Hashemi, S. Abrahams and M.K. Morell. 2001. Comparison of starch-branching enzyme genes reveals evolutionary relationships among isoforms characterization of a

gene for starch-branching enzyme IIa from the wheat genome donor *Aegilops tauschii*. **Plant Physiol.** 125: 1314-1324.

Rozas, J., J.C. Sánchez-DelBarrio, X. Messeguer and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. **Bioinformatics** 19: 2496-2497.

Salvo-Garrido, H., D.A. Laurie, B. Jaffé and J.W. Snape. 2001. An RFLP map of diploid *Hordeum bulbosum* L. and comparison with maps of barley (*H. vulgare* L.) and wheat (*Triticum aestivum* L.). **Theor. Appl. Genet.** 103: 869–880.

Sanjur, O.I., D.R. Piperno, T.C. Andres and L. Wessel-Beaver. 2002. Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: implications for crop plant evolution and areas of origin. **Proc. Natl. Acad. Sci. U.S.A.** 99: 535-540.

Shewmaker, C.K. and D.M. Stalker. 1992. Modifying starch biosynthesis with transgenes in potatoes. **Plant Physiol.** 100: 1083-1086.

Simko, I., K.G. Haynes, E.F. Ewing, S. Costanzo, B.J. Christ and R.W. Jones. 2004. Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. **Mol. Genet. Genomics** 271: 522-531.

Sivak, M.N. and J. Preiss. 1995. Starch synthesis in seeds, pp. 139-168. In J. Kigel and G. Galili, eds. **Seed Development and Germination.** Marcel Dekker, Inc., New York.

Slattery, C.J., I.H. Kavakli and T.W. Okita. 2000. Engineering starch for increased quantity and quality. **Trends Plant Sci.** 5: 291-298.

Smith, A.M., K. Denyer and C. Martin. 1997. The synthesis of the starch granule. **Annu. Rev. Plant Physiol. Plant mol. Biol.** 48: 67-87.

Sun, C.Q., X.K. Wang, Z.C. Li and A. Yoshimura. 2001. Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. Sativa* L.) using RFLP markers. **Theor. Appl. Genet.** 102: 157-162.

_____, _____, A. Yoshimura and K. Doi. 2002. Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. Sativa* L.). **Theor. Appl. Genet.** 104: 1335-1345.

Szalma, S.J., E.S. 4th Buckler, M.E. Snook and M.D. McMullen. 2005. Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. **Theor. Appl. Genet.** 110: 1324-1333.

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

Tan, Y.F., J.X. Li, S.B. Yu, Y.Z. Xing, C.G. Xu and Q. Zhang. 1999. The three important traits for cooking and eating quality of rice grains are controlled by a single locus in an elite rice hybrid, Shanyou 63. **Theor. Appl. Genet.** 99: 642-648.

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 Physiol.** 135: 412-420.

Templeton, A.R., K.A. Crandall and C.F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction

endonuclease mapping and DNA sequence data. III. Cladogram estimation. **Genetics** 132: 619-633.

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen and E.S. 4th Buckler. 2001. Dwarf8 polymorphisms associate with variation in flowering time. **Nat. Genet.** 28: 286-289.

Umemoto, T., M. Yano, H. Satoh, A. Shomura and Y. Nakamura. 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.

Vaughan, D.A., H. Morishima and K. Kadowaki. 2003. Diversity in the *Oryza* genus. **Curr. Opin. Plant Biol.** 6: 139-146.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. **Nucl. Acids Res.** 23: 4407-4414.

Wang, Z.Y., G. Second and S.D. Tanksley. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. **Theor. Appl. Genet.** 83: 565-581.

_____, F. Zheng, G. Shen, J. Gao, D.P. Snustad, M. Li, J. Zhang and M. Hong. 1995. The amylose content in rice endosperm is related to the *waxy* gene. **Plant J.** 7: 613-622.

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

Williams, J.G.K., A.R. Kubelik, K.I. Livak, J.A. Pafalski and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucl. Acids Res.** 18: 6231-6235.

Yamanaka, S., I. Nakamura, K.N. Watanabe and Y. Sato. 2004. Identification of SNPs in the *waxy* gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. **Theor. Appl. Genet.** 108: 1200-1204.

Yang, G.P., M.A. Maroof, C.G. Xu, Q. Zhang and R.M. Biyashev. 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. **Mol. Gen. Genet.** 245: 187-194.

Zhang, Q.F., M.A. Maroof, T.Y. Lu and B.Z. Shen. 1992. Genetic diversity and differentiation of indica and japonica rice detected by RFLP analysis. **Theor. Appl. Genet.** 83: 495-499.

Zhou, P.H., Y.F. Tan, Y.Q. He, C.G. Xu and Q. Zhang. 2003. Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. **Theor. Appl. Genet.** 106: 326-331.

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.

APPENDIX

Appendix Table 1 Plant Material and Eating-cooking Quality.

Accession No.	IRRI*/GSFF No.	Country/Province	cultivation	Sticky or non	ORF-100 deletion type	Eating-cooking Quality			Haplotype patterns							
						AAC	GT	GC	GBSSI	SSIIIa	SSIIIb	SSIVa	SSIVb	RBE1	RBE3	7 gene loci
1	243	UN	I	N	IND	14.17	7	10.1	A	A	A	A	A	A	A	1
2	5984	UN	I	N	IND	21.13	5	4.2	B	-	-	F	A	A	A	-
3	235	UN	I	N	IND	25.03	7	3.5	B	A	B	F	-	A	A	-
4	1491	UN	I	N	JAP	28.05	5	5.2	B	B	A	F	A	A	A	2
5	64	UN	I	S	IND	3.19	6	12	A	C	A	F	A	A	A	3
6	158	UN	I	N	IND	25.48	6	11.2	C	A	B	G	A	A	A	4
7	6941	UN	U	N	IND	14.81	7	7.8	A	D	A	F	A	B	A	5
8	2879	UN	U	N	IND	20.58	6	11.5	C	E	B	F	A	A	A	6
9	7508	UN	U	N	IND	21.42	7	3.6	B	C	A	F	A	A	A	7
10	3084	UN	U	N	IND	22.31	7	2.9	B	A	B	F	A	A	A	8
11	5294	UN	U	N	JAP	23.21	5	2.4	B	A	A	F	A	A	A	9
12	2915	UN	U	N	JAP	12.58	7	6.9	A	F	C	L	B	-	B	-
13	1804	UN	U	N	JAP	19.69	7	2.7	B	A	-	A	A	A	A	-
14	5967	UN	U	S	IND	1.95	7	7.8	C	A	B	F	A	A	A	10
15	1382	UN	U	S	IND	3.83	7	14.7	A	A	A	F	A	A	A	11
16	9105	UN	U	S	IND	5.34	7	10.8	A	A	A	B	A	A	A	12
17	5484	LN	F	N	IND	17.72	4	7.3	C	C	B	F	A	A	A	13
18	6305	LN	F	N	IND	23.34	5	3.2	B	D	A	F	A	B	A	14
19	5446	LN	F	N	IND	24.06	5	4	C	A	A	F	-	A	A	-
20	6361	LN	F	N	IND	25.80	5	5.4	C	A	A	F	A	A	A	15
21	19111	LN	F	S	IND	3.15	6	8.9	D	A	A	-	A	A	A	-
22	3416	LN	F	S	IND	3.72	7	11.5	D	G	-	-	-	A	A	-
23	6376	LN	F	S	IND	4.35	6	11	D	C	B	F	A	A	A	16
24	6371	LN	I	N	IND	22.18	7	2.6	C	C	A	F	A	A	A	17
25	4934	LN	I	N	IND	23.34	7	4.7	C	A	B	C	A	A	A	18
26	6370	LN	I	N	IND	24.21	5	3.9	C	-	A	F	A	A	A	-
27	3615	LN	I	N	IND	29.66	5	5.4	C	A	A	F	A	A	A	15
28	2729	LN	I	S	IND	2.80	7	11.2	D	C	A	F	A	A	A	19
29	3407	LN	I	S	IND	4.66	7	13.1	A	H	D	E	A	A	-	-
30	4936	LN	I	S	IND	5.68	7	11.6	A	C	A	F	A	A	A	3
31	8984	LN	U	N	IND	18.72	7	6.6	C	-	A	F	A	-	-	-
32	8981	LN	U	N	IND	15.60	7	6.5	A	-	A	F	A	A	A	-
33	6355	LN	U	N	IND	25.40	5	4.5	B	-	A		A	A	A	-
34	1862	LN	U	N	IND	27.57	7	2.7	B	B	B	F	A	A	A	20
35	1783	LN	U	N	JAP	24.11	5	7.5	D	G	C	D	B	A	B	21
36	5114	LN	U	S	IND	5.05	6	13.2	A	A	B	G	A	A	A	22
37	6251	LN	U	S	IND	6.42	7	9.6	C	I	B	F	A	A	A	23
38	5112	LN	U	S	IND	7.60	7	7.2	A	C	A	F	A	C	A	24
39	6421	LN	U	S	JAP	5.82	6	13	A	A	A	E	A	A	A	25
40	5752	UNE	F	N	IND	14.34	2	5.8	A	H	A	F	A	A	A	26
41	4487	UNE	I	N	IND	19.54	7	6.6	A	J	A	F	A	B	A	27
42	3340	UNE	I	N	IND	23.92	7	1	B	A	A	C	A	A	A	28
43	1640	UNE	I	N	IND	26.70	6	3.6	E	-	B	F	A	C	A	-
44	4495	UNE	I	N	IND	28.33	5	6.8	B	K	A	E	A	A	A	29

Appendix Table 1 (Continued)

Accession No.	IRRI*/GSFF No.	Country/Province	cultivation	Sticky or non	ORF-100 deletion type	Eating-cooking Quality			Haplotype patterns							
						AAC	GT	GC	GBSSI	SSIIIa	SSIIIb	SSIVa	SSIVb	RBE1	RBE3	7 gene loci
45	3211	UNE	I	S	IND	5.08	7	13.8	A	A	A	F	A	A	A	11
46	4503	UNE	I	S	IND	5.79	6	7.2	A	C	A	G	A	A	A	30
47	3201	UNE	I	S	IND	7.22	6	14.8	C	C	B	F	A	A	A	13
48	3199	UNE	I	S	IND	7.51	7	13.5	-	C	A	F	A	A	A	-
49	3200	UNE	I	S	IND	7.33	6	11	A	L	B	F	C	-	A	-
50	3888	UNE	U	S	IND	6.55	6	12.3	C	A	A	F	A	A	A	15
51	4009	UNE	U	S	IND	6.69	7	8.5	B	C	B	F	A	A	A	31
52	23107	UNE	U	S	JAP	4.83	7	14.2	F	G	C	D	D	A	B	32
53	21697	UNE	U	S	JAP	5.72	6	13.9	F	G	C	L	D	A	B	33
54	23271	UNE	U	S	JAP	6.43	6	13.3	B	G	C	L	D	A	B	34
55	22207	LNE	F	S	IND	2.59	6	12.6	B	C	A	F	-	A	A	-
56	5824	LNE	I	N	IND	14.58	7	8.8	B	I	A	F	A	A	A	35
57	5358	LNE	I	N	IND	18.66	7	4.8	B	A	B	F	A	A	A	8
58	611	LNE	I	N	IND	25.80	7	3.3	G	C	A	G	A	B	A	36
59	605	LNE	I	N	JAP	21.03	7	4.9	G	A	B	F	A	A	A	37
60	1663	LNE	I	N	JAP	21.64	6	5.4	B	D	B	F	A	A	A	38
61	5759	LNE	I	N	JAP	25.00	5	2	G	A	B	E	A	B	A	39
62	5773	LNE	I	N	IND	22.32	6	5.9	H	-	B	F	A	B	A	-
63	1636	LNE	I	S	IND	4.22	7	10.4	B	A	B	F	A	B	A	40
64	4090	LNE	I	S	IND	5.38	7	10.3	I	A	-	F	A	A	A	-
65	3333	LNE	I	S	IND	6.43	6	12	I	C	A	F	A	A	A	41
66	7975	LNE	U	N	IND	18.24	7	8.5	B	G	A	F	A	A	A	42
67	22718	LNE	U	N	JAP	18.84	4	8.5	C	A	C	D	I	A	B	43
68	2593	LNE	U	S	IND	3.99	6	7.2	I	A	B	F	A	A	A	44
69	1867	LNE	U	S	IND	7.25	6	12.1	B	A	A	F	-	A	A	-
70	1892	LNE	U	S	JAP	4.56	6	11.5	A	L	A	L	D	B	B	45
71	9230	W	F	N	IND	27.65	5	4.4	G	G	A	F	A	A	A	46
72	5326	W	F	N	IND	30.07	5	4.3	G	A	A	E	A	A	A	47
73	11056	W	F	N	JAP	26.20	6	3.7	G	G	A	F	A	C	A	48
74	5302	W	I	N	IND	20.64	5	4.1	G	A	A	F	A	A	A	49
75	9245	W	I	N	IND	22.23	5	2.1	G	A	A	F	A	A	A	49
76	5303	W	I	N	IND	25.83	7	4	G	D	A	F	A	A	A	50
77	5306	W	I	N	JAP	24.15	5	3.4	G	A	A	F	A	A	A	49
78	21428	W	I	S	IND	6.24	6	9.3	B	A	B	F	A	A	A	8
79	9250	W	I	S	IND	6.54	6	13	B	C	B	F	C	-	A	-
80	11110	W	I	S	IND	7.14	6	12.4	B	A	B	F	A	A	A	8
81	9243	W	U	N	IND	24.03	4	2.3	G	I	A	E	A	A	A	51
82	23080	W	U	N	IND	24.47	5	2.4	G	K	A	G	A	A	A	52
83	5319	W	U	N	JAP	20.65	5	3.7	G	H	A	F	A	A	A	53
84	21405	W	U	N	JAP	20.98	6	4.9	C	G	C	L	I	A	B	54
85	5323	W	U	N	JAP	23.35	5	5.2	G	A	A	F	A	A	A	49
86	11566	W	U	S	IND	3.24	6	11	I	A	A	F	A	A	A	55
87	11066	W	U	S	IND	7.12	7	15	B	H	B	-	A	A	A	-
88	21833	W	U	S	JAP	2.16	7	10.8	G	G	E	L	-	A	C	-

Appendix Table 1 (Continued)

Accession No.	IRRI*/GSFF No.	Country/Province	cultivation	Sticky or non	ORF-100 deletion type	Eating-cooking Quality			Haplotype patterns							
						AAC	GT	GC	GBSSI	SSIIIa	SSIIIb	SSIVa	SSIVb	RBE1	RBE3	7 gene loci
89	21407	W	U	S	JAP	4.41	6	13.9	B	G	E	L	I	A	C	56
90	21830	W	U	S	JAP	5.66	7	11.5	I	G	E	L	D	B	B	57
91	8018	C	F	N	IND	21.07	7	4.6	J	A	A	F	C	B	A	58
92	2092	C	F	N	IND	21.37	7	3.7	I	A	F	E	A	B	A	59
93	5695	C	F	N	IND	25.74	5	6.9	I	A	F	F	A	B	A	60
94	1688	C	F	N	JAP	22.87	5	12.8	K	G	A	G	A	A	A	61
95	6562	C	F	N	JAP	24.61	5	6.3	-	-	A	G	A	A	A	-
96	798	C	I	N	IND	23.62	5	4.3	G	C	B	F	A	A	A	62
201	8024	C	I	N	IND	16.66	5	6.4	-	D	A	F	A	A	D	-
202	7584	C	I	N	IND	22.98	5	6.9	L	C	A	F	A	A	A	63
203	6221	C	I	N	IND	26.05	5	3	G	D	A	F	A	A	A	50
204	788	C	I	N	JAP	27.94	5	6.2	G	C	B	F	A	A	A	62
205	8013	C	I	S	IND	5.49	7	12.6	B	D	B	F	A	A	A	38
206	7538	C	I	S	IND	5.64	7	13.5	G	A	A	F	-	A	A	-
207	6721	C	I	S	IND	5.64	7	7.1	H	A	B	F	A	A	A	64
208	6827	C	I	S	IND	4.99	7	12.6	I	C	B	F	A	A	A	65
209	3656	C	U	N	IND	17.98	7	13.8	I	G	A	F	A	A	A	66
210	3706	C	U	N	IND	18.93	7	4.5	E	-	A	F	A	A	A	-
211	8132	C	U	N	JAP	18.56	7	7.1	B	G	G	L	E	A	B	67
212	8967	C	U	N	JAP	22.59	4	9	G	C	A	F	A	A	A	68
213	8932	C	U	S	IND	5.05	6	8.1	I	I	B	F	A	A	E	69
214	6720	C	U	S	IND	5.38	7	13.7	H	A	A	F	A	A	A	70
215	1682	C	U	N	JAP	22.04	7	3.2	E	M	A	F	A	A	A	71
216	8100	C	U	S	JAP	1.11	7	9.8	I	G	B	L	I	A	B	72
217	7214	E	F	N	IND	24.92	5	4.9	K	C	A	F	A	B	A	73
218	9426	E	F	N	IND	26.94	5	3.4	-	A	B	F	-	A	A	-
219	14175	E	F	S	IND	7.11	6	3.9	G	A	A	F	F	B	A	74
220	12399	E	F	S	IND	6.88	7	11.8	B	D	B	F	A	A	A	38
221	9499	E	F	N	JAP	22.32	5	2.1	G	A	A	F	A	A	A	49
222	3825	E	I	N	IND	20.46	7	5.8	E	C	A	F	A	A	A	75
223	3822	E	I	N	IND	22.27	5	4.5	I	A	B	F	A	A	A	44
224	3861	E	I	N	IND	30.01	5	5.4	G	-	B	F	A	C	A	-
225	9537	E	I	S	IND	6.09	6	11.6	B	C	A	F	A	A	A	7
226	3868	E	I	S	IND	6.20	6	11.1	I	C	A	F	A	A	A	41
227	7226	E	I	S	IND	8.72	6	12.3	B	A	A	F	A	A	A	9
228	7274	E	I	N	JAP	22.89	5	4.1	G	D	A	F	A	A	A	50
229	12438	E	U	N	IND	16.21	7	8.5	B	I	A	F	A	A	A	35
230	21284	E	U	N	IND	19.05	7	7.5	B	A	A	A	A	C	A	76
231	12333	E	U	N	JAP	26.29	6	5.6	G	G	B	F	A	A	A	77
232	6173	US	I	N	IND	15.08	7	7.9	M	N	A	F	A	A	A	78
233	6199	US	I	N	IND	20.25	5	2.8	G	C	A	A	A	A	A	79
234	4079	US	I	N	IND	28.05	5	6.6	G	D	A	F	A	A	A	50
235	8069	US	I	N	JAP	19.76	6	5.3	B	G	B	L	-	B	B	-
236	3787	US	I	N	IND	25.03	5	6.7	G	D	A	F	A	A	A	50

Appendix Table 1 (Continued)

Accession No.	IRRI*/GSFF No.	Country/Province	cultivation	Sticky or non	ORF-100 deletion type	Eating-cooking Quality			Haplotype patterns							
						AAC	GT	GC	GBSSI	SSIIIa	SSIIIb	SSIVa	SSIVb	RBE1	RBE3	7 gene loci
237	4024	US	I	N	IND	23.76	5	6.2	-	D	B	F	-	A	A	-
238	8212	US	I	S	IND	7.66	7	13	H	-	A	-	A	A	A	-
239	4017	US	U	N	IND	25.44	6	2.7	J	-	B	F	A	A	A	-
240	4043	US	U	N	IND	28.19	7	3.2	J	D	B	F	A	C	A	80
241	8120	US	U	N	JAP	21.09	7	3.8	I	F	B	D	I	B	B	81
242	4016	US	U	N	JAP	24.09	5	5.8	I	K	B	D	E	C	B	82
243	8121	US	U	S	IND	4.66	7	12.4	-	F	A	F	A	A	A	-
244	8099	US	U	S	JAP	2.70	7	10.8	-	G	G	L	E	B	C	-
245	1958	LS	I	N	IND	13.60	7	11.6	-	J	B	F	A	B	A	-
246	1641	LS	I	N	IND	23.79	6	4.8	J	J	B	A	A	-	A	-
247	1553	LS	I	N	IND	27.62	5	4.6	-	H	B	F	A	A	A	-
248	1896	LS	I	N	JAP	25.74	5	5.2	-	M	H	-	I	-	B	-
249	9975	LS	I	S	IND	5.12	7	11.1	B	H	B	F	A	A	A	83
250	9922	LS	I	N	IND	23.59	5	6.5	K	G	A	F	A	A	A	84
251	4304	LS	I	S	JAP	6.54	7	9.1	E	D	A	C	A	A	A	85
252	12676	LS	U	N	IND	23.78	5	5.4	J	D	B	F	A	C	A	80
253	1975	LS	I	N	IND	25.41	6	10	I	K	A	F	A	A	A	86
254	1989	LN	I	N	IND	27.53	7	2.1	G	A	A	A	A	B	A	87
255	1903	LS	U	N	JAP	18.84	5	5.7	-	M	H	D	-	C	B	-
256	1925	LS	U	N	JAP	23.42	6	7.1	G	J	B	F	A	D	B	88
257	19477	LS	U	S	JAP	3.85	7	12	I	O	B	L	E	A	B	89
258	19257	LS	U	S	JAP	4.52	6	12	I	G	B	L	I	E	B	90
259	1909	LS	U	S	JAP	5.46	7	12.4	I	G	B	D	I	A	B	91
260	9920	LS	U	N	JAP	22.60	6	5.1	I	G	B	D	I	A	B	91
261	5499	THA[b]	-	N	JAP	21.43	5	7.5	B	O	H	H	A	A	B	92
262	5501	THA[b]	-	N	IND	21.77	5	3.8	-	G	E	F	A	A	B	-
263	5503	THA[c]	-	N	JAP	-	4	7.3	K	C	H	G	A	-	B	-
264	8428	THA[c]	-	N	IND	21.31	5	6.1	G	A	E	F	A	C	A	93
265	15150	THA[c]	-	N	IND	-	2	12.6	-	L	H	I	A	C	A	-
266	16150	THA[b]	-	N	IND	-	5	5.2	-	P	B	J	A	A	A	-
267	16186	THA[b]	-	N	JAP	-	5	4.8	D	A	H	H	A	A	A	94
268	17662	THA[a]	-	N	JAP	22.94	5	13.7	-	A	E	F	A	A	A	-
269	18220	THA[b]	-	N	IND	-	5	5.4	H	Q	B	G	A	A	A	95
270	18223	THA[b]	-	N	IND	-	5	8.3	A	G	H	H	A	C	F	96
271	18234	THA[b]	-	N	IND	-	7	2.6	-	-	H	F	A	A	A	-
272	18847	THA[a]	-	N	JAP	-	5	9.6	L	-	E	G	A	A	B	-
273	18851	THA[a]	-	N	JAP	-	5	5.3	L	R	E	G	A	A	B	97
274	-	THA[e]	I	N	IND	-	6	-	B	R	A	F	A	A	A	98
275	-	THA[f]	I	S	IND	-	7	-	C	A	D	F	F	A	A	99
276	-	THA[g]	I	S	JAP	-	7	-	B	C	A	F	A	A	A	7
277	-	THA[h]	I	N	JAP	-	6	-	H	M	E	L	I	A	B	100
278	21302*	IND	-	S	IND	0.81	-	-	I	H	-	F	A	A	B	-
279	24473*	THA	-	N	IND	21.89	-	-	G	A	E	F	A	A	A	101
280	33200*	MMR	-	S	IND	0.00	-	-	E	G	A	F	A	B	A	102

Appendix Table 1 (Continued)

Accession No.	IRRI*/GSFF No.	Country/Province	cultivation	Sticky or non	ORF-100 deletion type	Eating-cooking Quality			Haplotype patterns							
						AAC	GT	GC	*GBSSI*	SSIIIa	SSIIIb	SSIVa	SSIVb	*RBE1*	*RBE3*	7 gene loci
281	30336*	PHL	-	N	IND	21.89	-	-	G	H	E	F	A	A	B	103
282	51070*	LKA	-	N	IND	22.70	-	-	K	G	A	F	A	B	B	104
283	38699*	PAK	-	N	IND	23.51	-	-	H	M	H	K	A	F	B	105
284	33745*	MMR	-	N	IND	16.22	-	-	-	Q	E	D	-	B	B	-
285	27588*	BGD	-	S	IND	0.00	-	-	J	M	H	-	I	A	A	-
286	32382*	BTN	-	N	IND	22.70	-	-	G	T	E	D	A	G	A	106
287	23705*	THA	-	N	JAP	16.22	-	-	E	A	A	F	C	A	A	107
288	23729*	THA	-	S	IND	0.00	-	-	-	G	G	L	D	G	B	-
289	23710*	THA	-	N	IND	12.16	-	-	G	G	E	L	I	H	C	108
290	328*	PHL	-	N	IND	19.46	-	-	J	F	H	-	-	B	B	-
291	48864*	IDN	-	S	IND	4.05	-	-	E	A	H	D	I	B	B	109
292	55506*	KOR	-	S	IND	8.11	-	-	G	M	E	L	I	B	B	110
293	31524*	LKA	-	N	IND	16.22	-	-	N	G	H	D	E	I	B	111
294	106148*	LAO[b]	-	N	IND	-	-	-	O	S	I	M	F	F	A	112
295	104549*	LBR[d]	-	N	IND	-	-	-	P	T	J	-	G	J	-	-
296	105491*	MYS[c]	-	N	JAP	-	-	-	K	U	H	N	H	A	-	-

[a] Wild rice (*Oryza* spp.)

[b] *Oryza nivara*

[c] *Oryza rufipogon*

[d] *Oryza glaberrima*

[e] KDML105

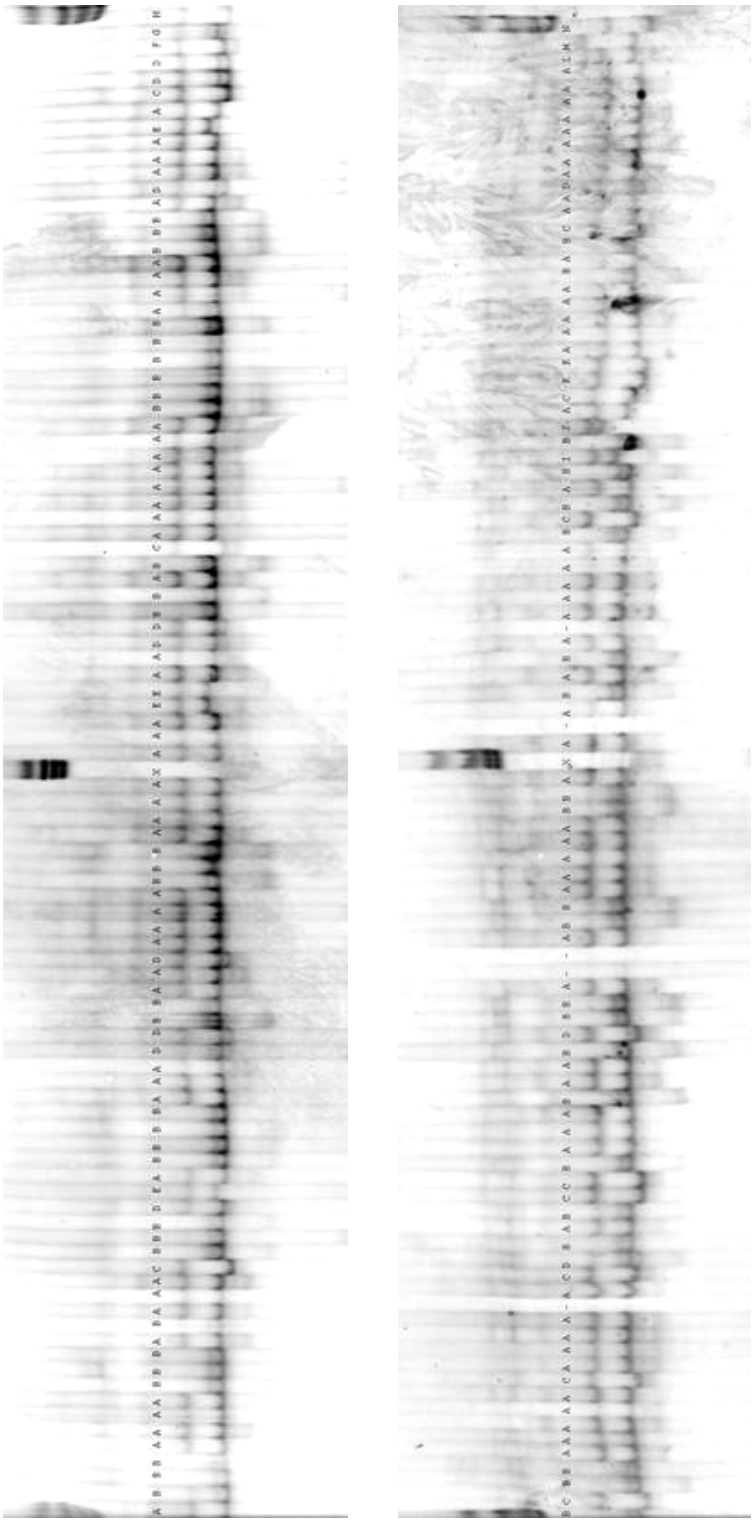
[f] RD6

[g] RD10

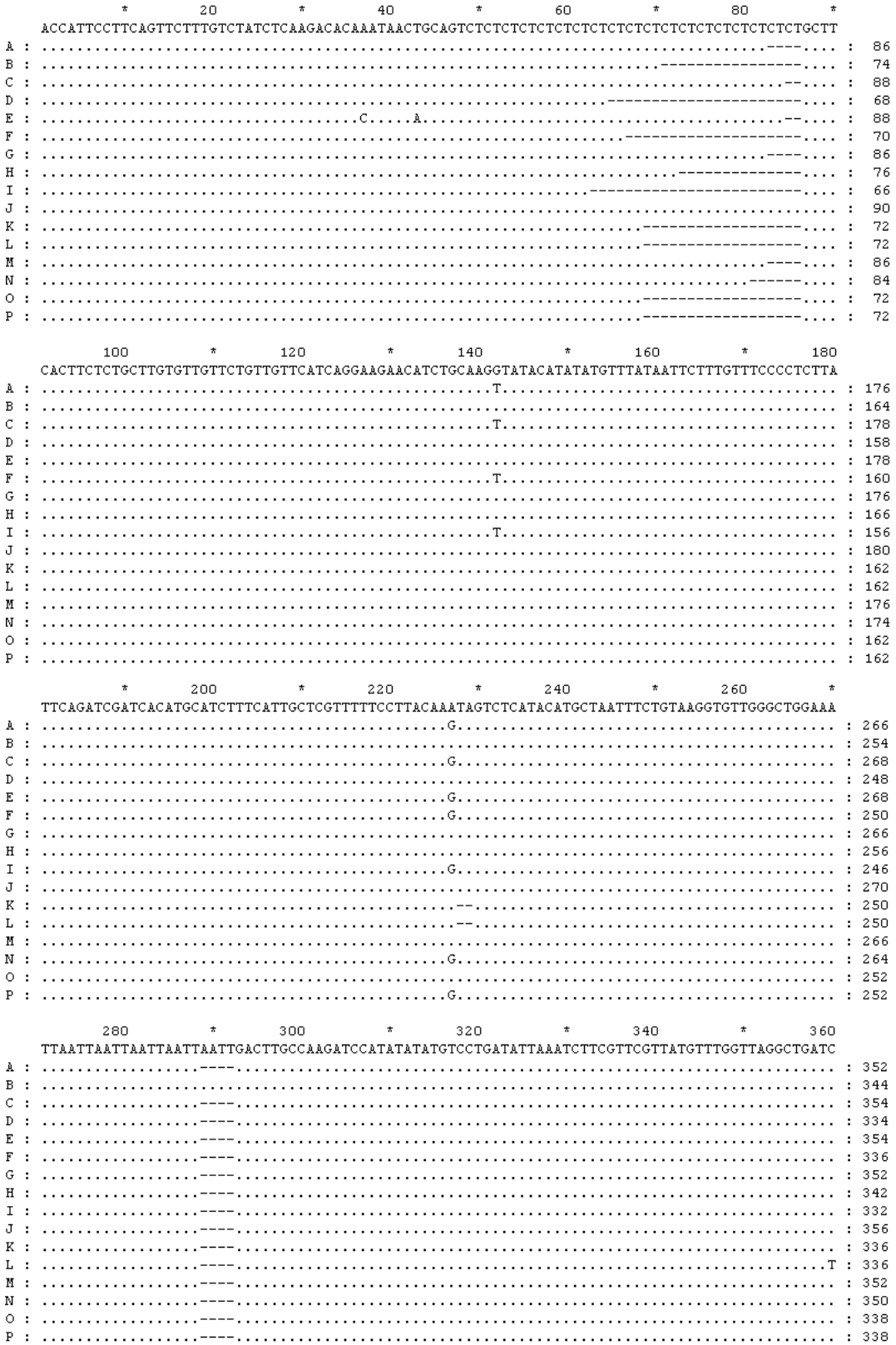
[h] nipponbare

Country/Province : UN-Upper Northern Thailand, (list the provinces included in each of the regions, LN-Lower northern Thailand, UNE-Upper Northeastern Thailand, LNE-Lower Northeastern Thailand, W-Western Thailand, C-Central Thailand, E-Eastern Thailand, US-Upper Southern Thailand, LS-Lower southern Thailand, THA-Thailand, IND-India, MMR-Myanmar, PHL-Philippines, LKA-Sri Lanka, PAK-Pakistan, BGD-Bangladesh, BTN-Bhutan, IDN-Indonesia, KOR-Korea, LAO-Lao, LBR-Liberia, MYS-Malaysia

Planting: I-irrigated rice, U-Upland rice, F-Floating rice



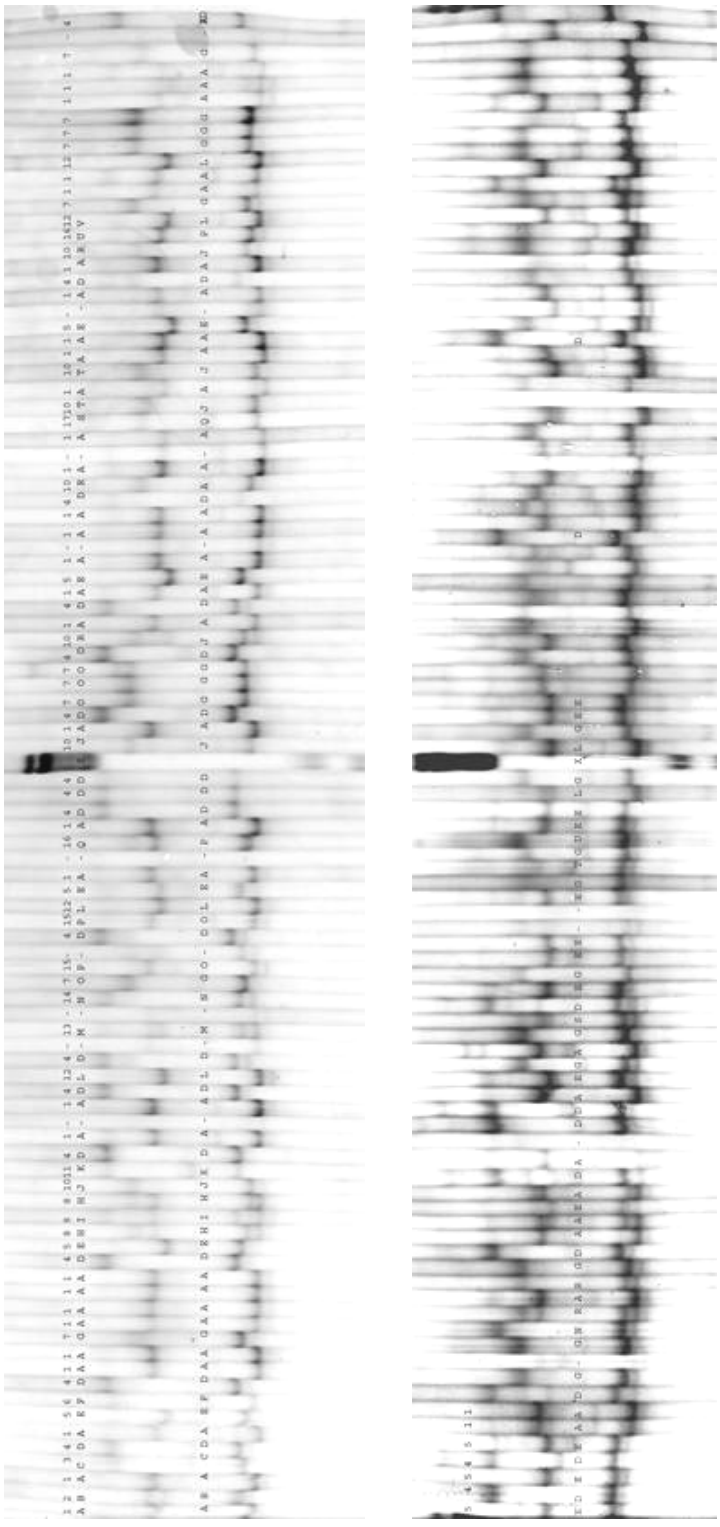
Appendix Figure 1 SSCP analysis of DNA fragments corresponding to the GBSSI locus separated by polyacylamide gel. Different types of banding pattern were identified (pattern A to P).



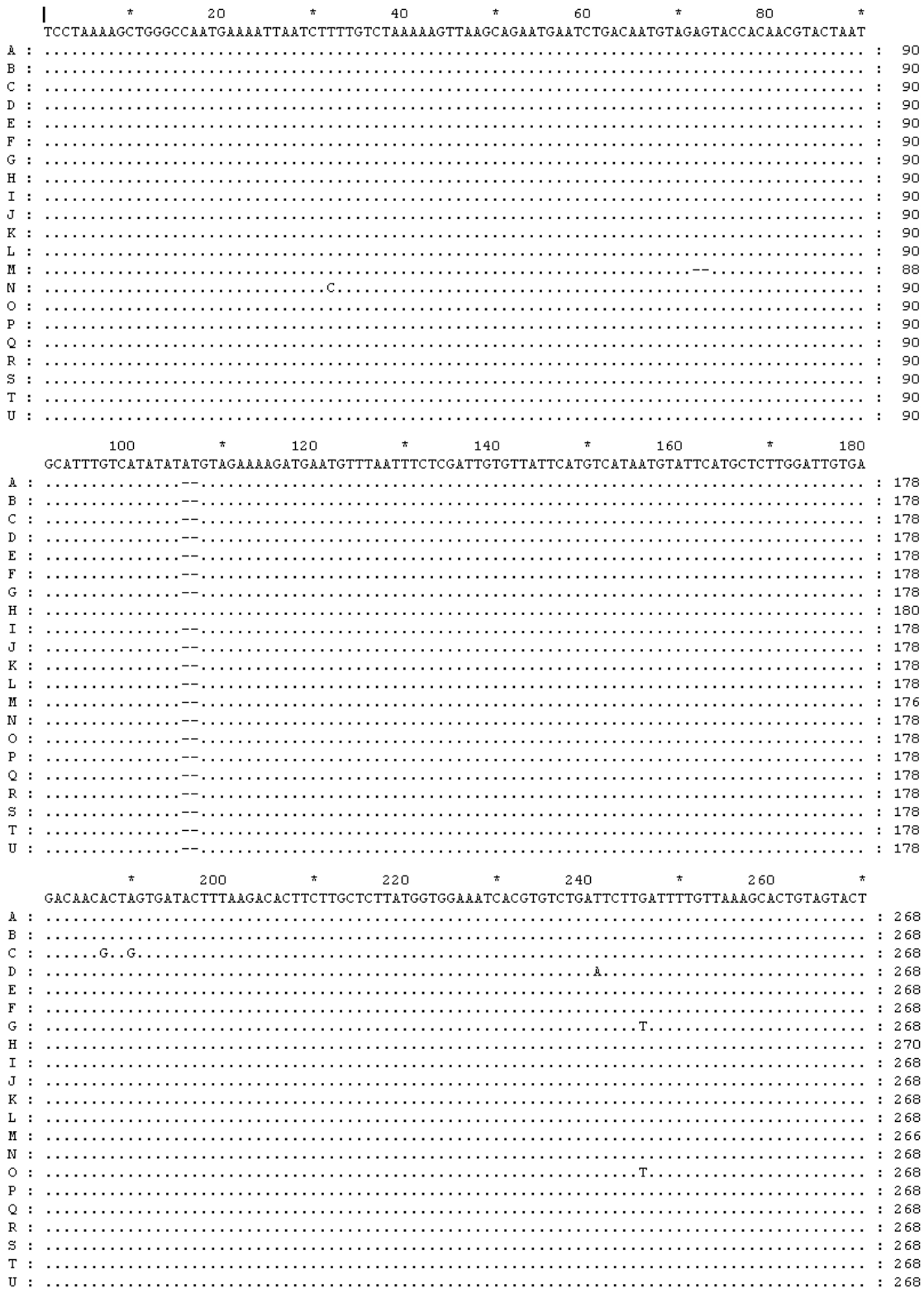
Appendix Figure 2 DNA sequence alignment of the haplotypes at the GBSSI locus. Sixteen haplotypes are shown.

```
          *       380         *       400         *       420         *       440         *
    AATGTTATTCTAGAGTCTAGAGAAACACACCCAGGGGTTTTCCAGCTAGCTCCACAAGATGGTGGGCTAGCTGACCTAGATTTGAAGTCT
A : .........................................A.......................................... : 442
B : G.......................T.....................................................-...... : 433
C : .........................................A.......................................... : 444
D : G............................T..................................................... : 424
E : .........................................A.......................................... : 444
F : .........................................A.......................................... : 426
G : G.....................................................................-...... : 441
H : G.........................T..................................................... : 432
I : .........................................A.......................................... : 422
J : G.....................................................................-...... : 445
K : ..................................................................................... : 426
L : ..................................................................................... : 426
M : G....................................................................-...... : 441
N : .........................................A.......................................... : 440
O : G..................................................................................... : 428
P : G.........................T.....................................T.............. : 428

          460         *       480         *       500         *       520         *       540
    CACTCTTTCTAATTATTTGATATTAGATCATTTTCTAATATTTGCGTCTTTTTTTATTCTAGAGTCTAGATCTTGTGTTCAACTCTCGTT
A : .....C..A.........T.....................C.T.......................................... : 532
B : ..................................................................................... : 523
C : .....C..A.........T.....................C.T.......................................... : 534
D : ..................A.......................................................................... : 514
E : .....C..A.........T.....................C.T.......................................... : 534
F : .....C..A.........T.....................C.T.......................................... : 516
G : ..........................................T.......................................... : 531
H : ..................................................................................... : 522
I : .....C..A.........T.....................C.T.......................................... : 512
J : ..........................................T.......................................... : 535
K : ..................................................................................... : 516
L : ..................................................................................... : 516
M : ..................................................................................... : 531
N : .....C..A.........T.....................C.T.......................................... : 530
O : T.................................................................................... : 518
P : .....C.........................T.......C.......................................... : 518

          *       560
    AAATCATGTCTCTCGCCACTGGAGAAAC
A : ............................ : 560
B : ............................ : 551
C : ............................ : 562
D : ............................ : 542
E : ............................ : 562
F : ............................ : 544
G : ............................ : 559
H : ............................ : 550
I : ............................ : 540
J : ............................ : 563
K : ............................ : 544
L : ............................ : 544
M : ............................ : 559
N : ............................ : 558
O : ............................ : 546
P : ............................ : 546
```

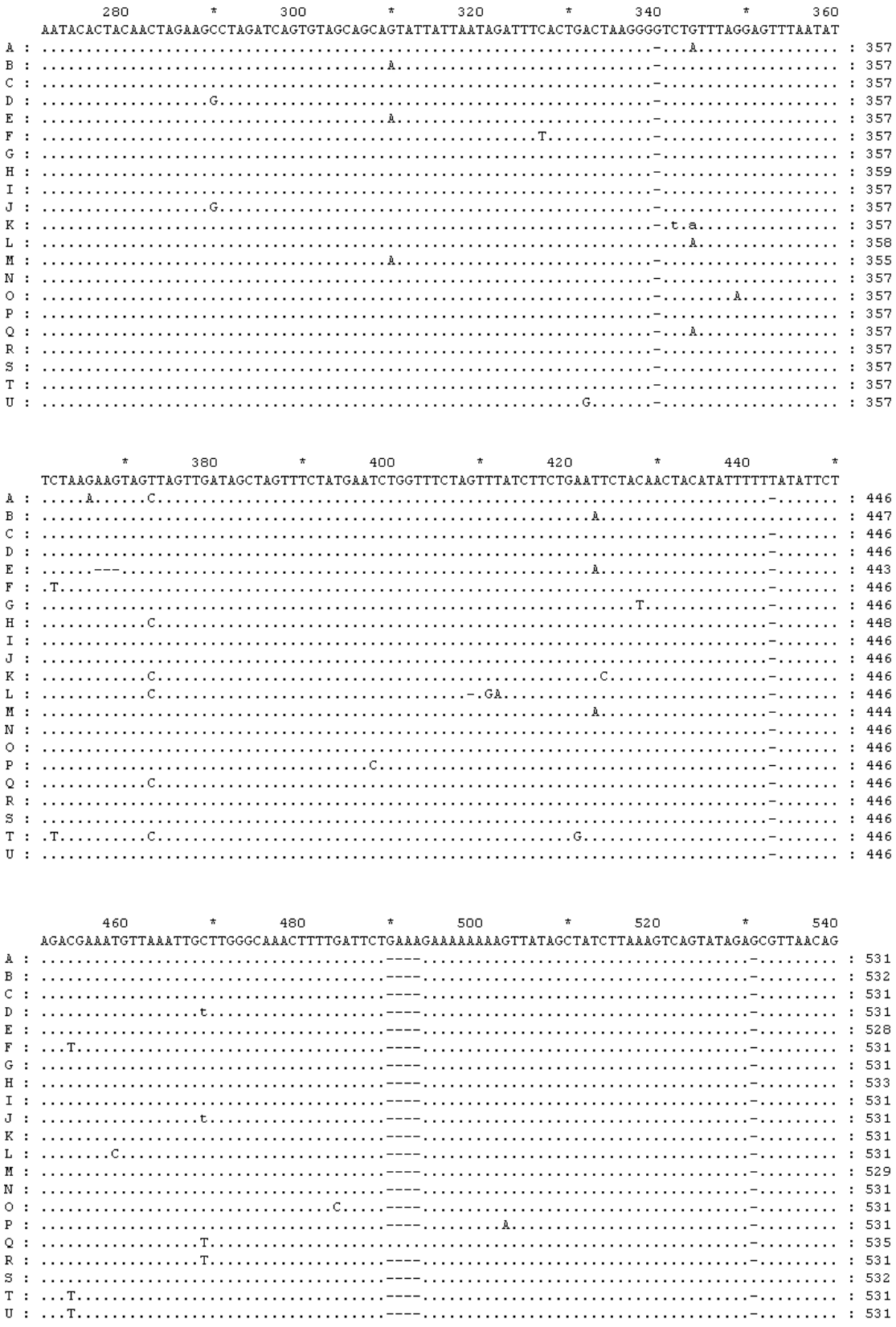
Appendix Figure 2 (Continued)



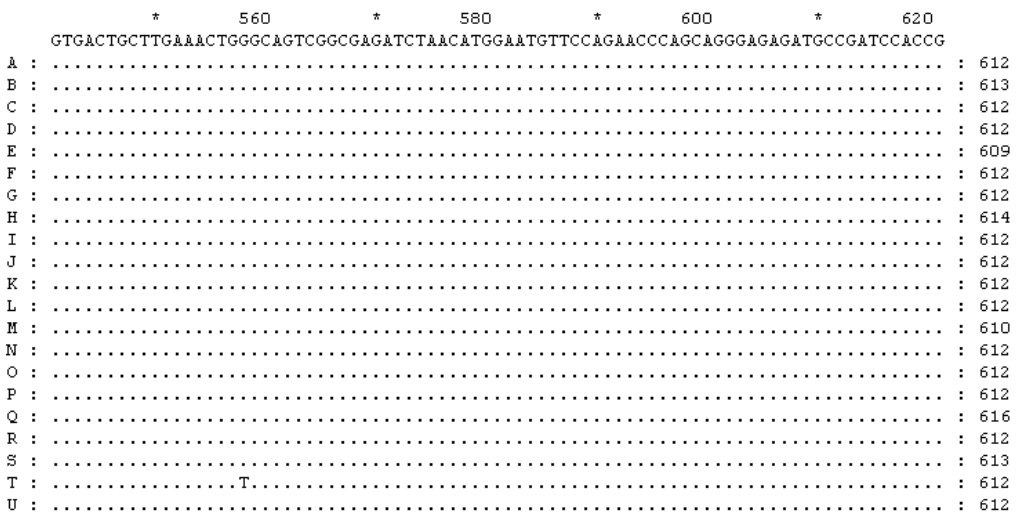
Appendix Figure 3 SSCP analysis of DNA fragments corresponding to the SSSIIIa locus separated by polyacylamide gel. Different types of banding pattern were identified (pattern A to U).



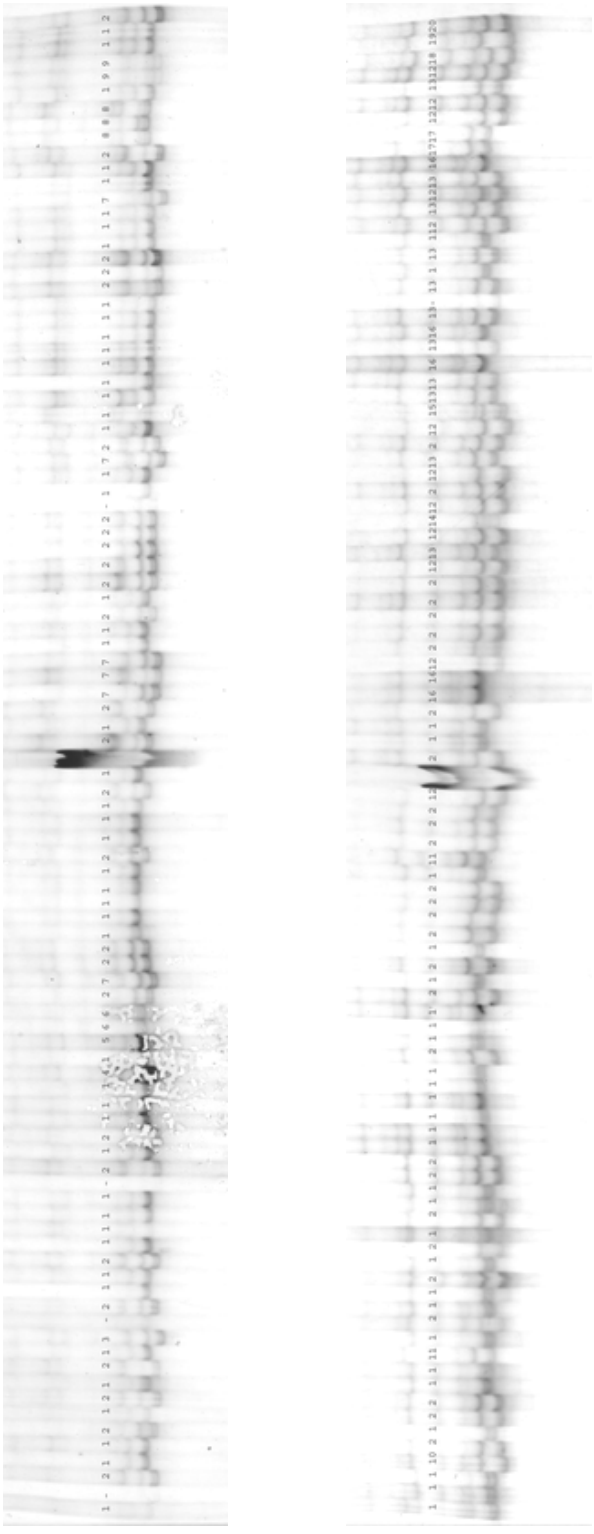
Appendix Figure 4 DNA sequence alignment of the haplotypes at the SSSIIIa locus. Twenty-one haplotypes are shown.



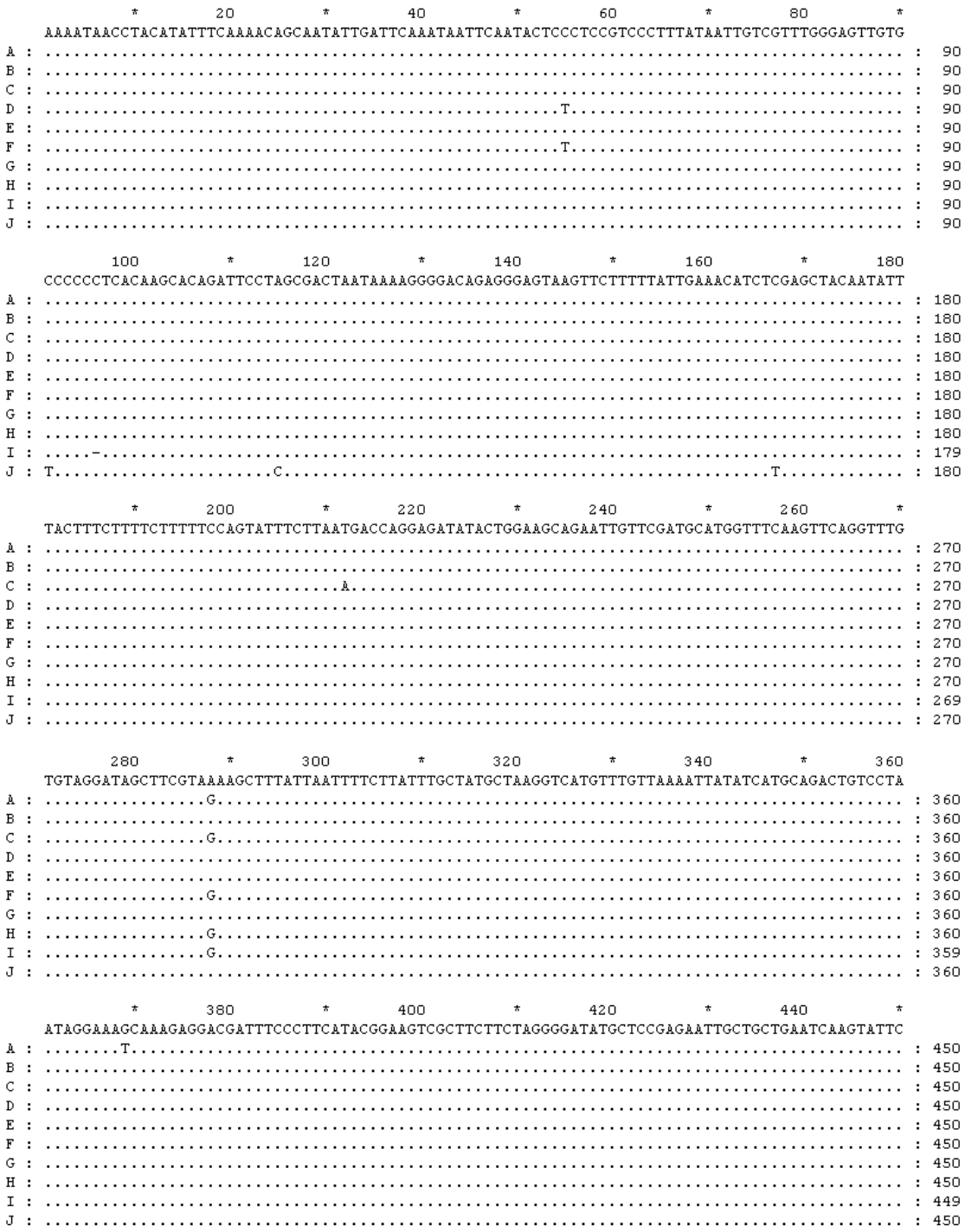
Appendix Figure 4 (Continued)



Appendix Figure 4 (Continued)



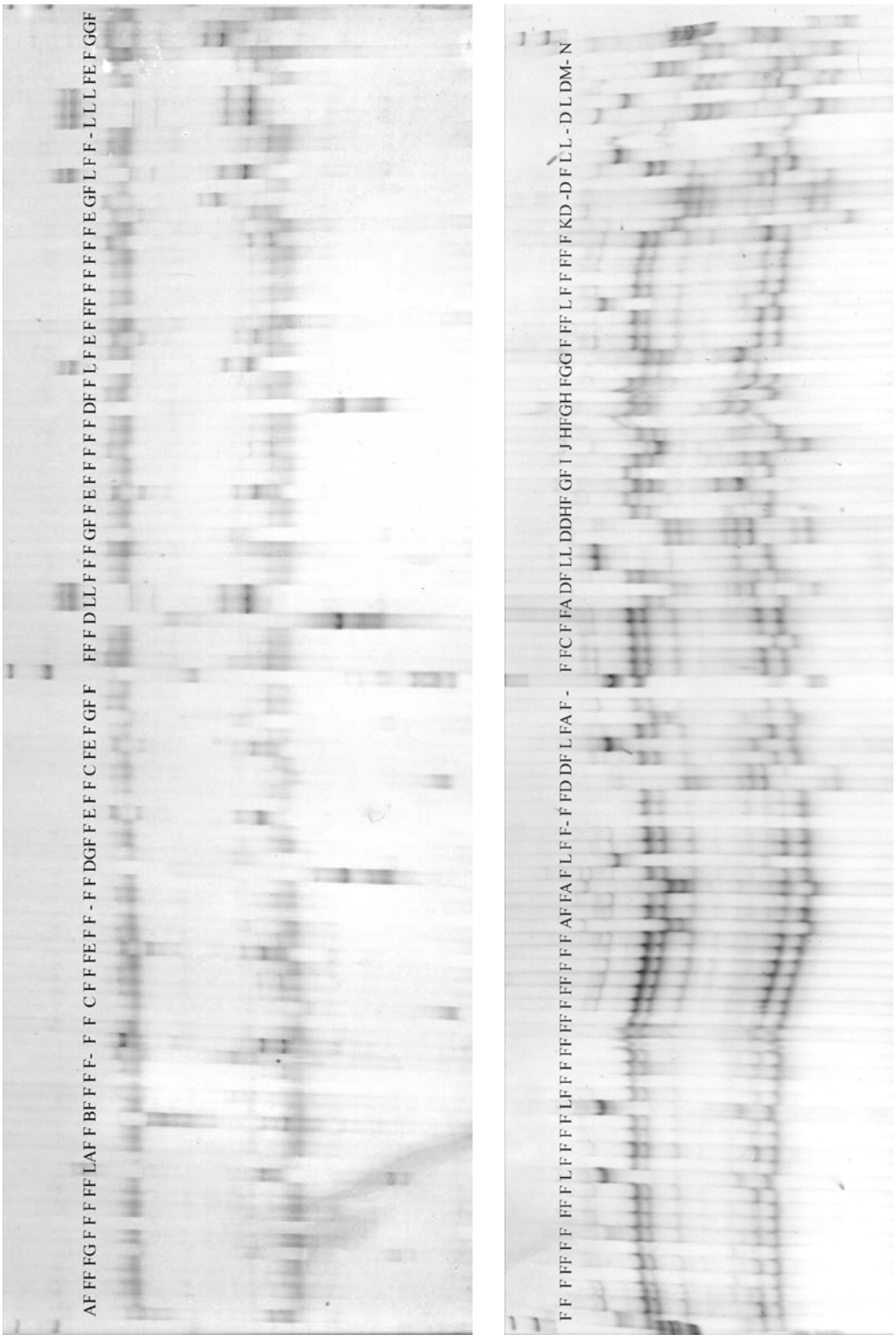
Appendix Figure 5 SSCP analysis of DNA fragments corresponding to the SSSIIIb locus separated by polyacylamide gel. Different types of banding pattern were identified (pattern A to J).



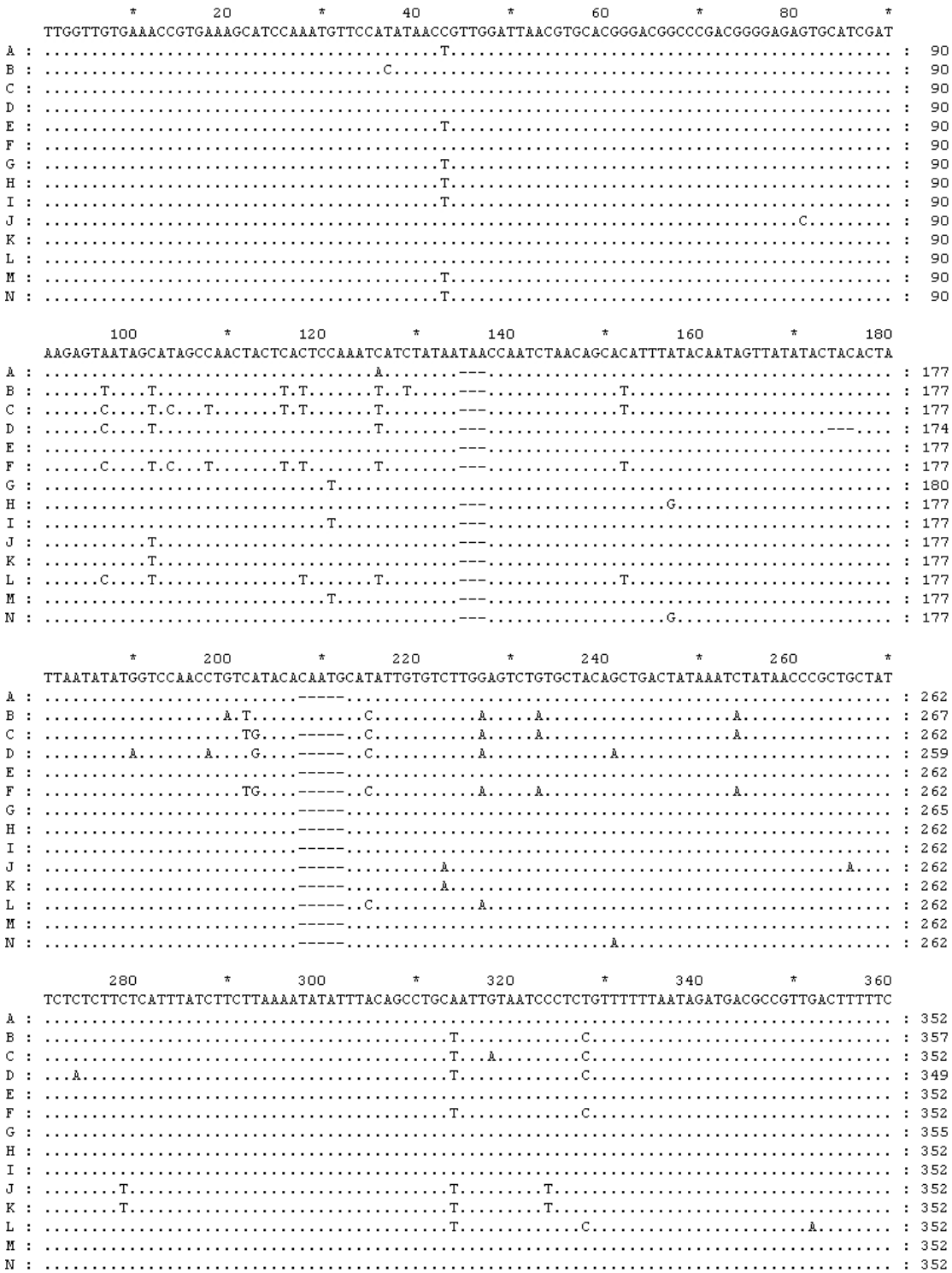
Appendix Figure 6 DNA sequence alignment of the haplotypes at the SSSIIIb locus. Ten haplotypes are shown.

```
          460         *       480         *       500         *       520         *       540
    AGGAGCGAGAACATATTAATAGTGATGAAGAAACATTTGATACGTACAACAGATTACTACGTAATGAGTCAACAGAGTGGAAAAAGTTAG
A : ..................................................................................C....... : 540
B : ..................................................................................C....... : 540
C : ..................................................................................C....... : 540
D : ..................................................................................... : 540
E : ..................................................................................... : 540
F : ..................................................................................... : 540
G : ...........................................................A......................... : 540
H : ..................................................................................C....... : 540
I : ..................................................................................C....... : 539
J : ..................................................................................... : 540

          *       560         *       580         *       600         *       620
    ATACTACTGAAGTGGATTTGTCACAAGATGTTTCAAGCAGTTCAATGAGGAAAGTGGATGCGACGGATGAAGCTAAGCTA
A : ..................................................................... : 620
B : ..................................................................... : 620
C : ..................................................................... : 620
D : ..................................................................... : 620
E : ..................................................................... : 620
F : ..................................................................... : 620
G : ..................................................................... : 620
H : ..................................................................... : 620
I : ..................................................................... : 619
J : ..................................................................... : 620
```

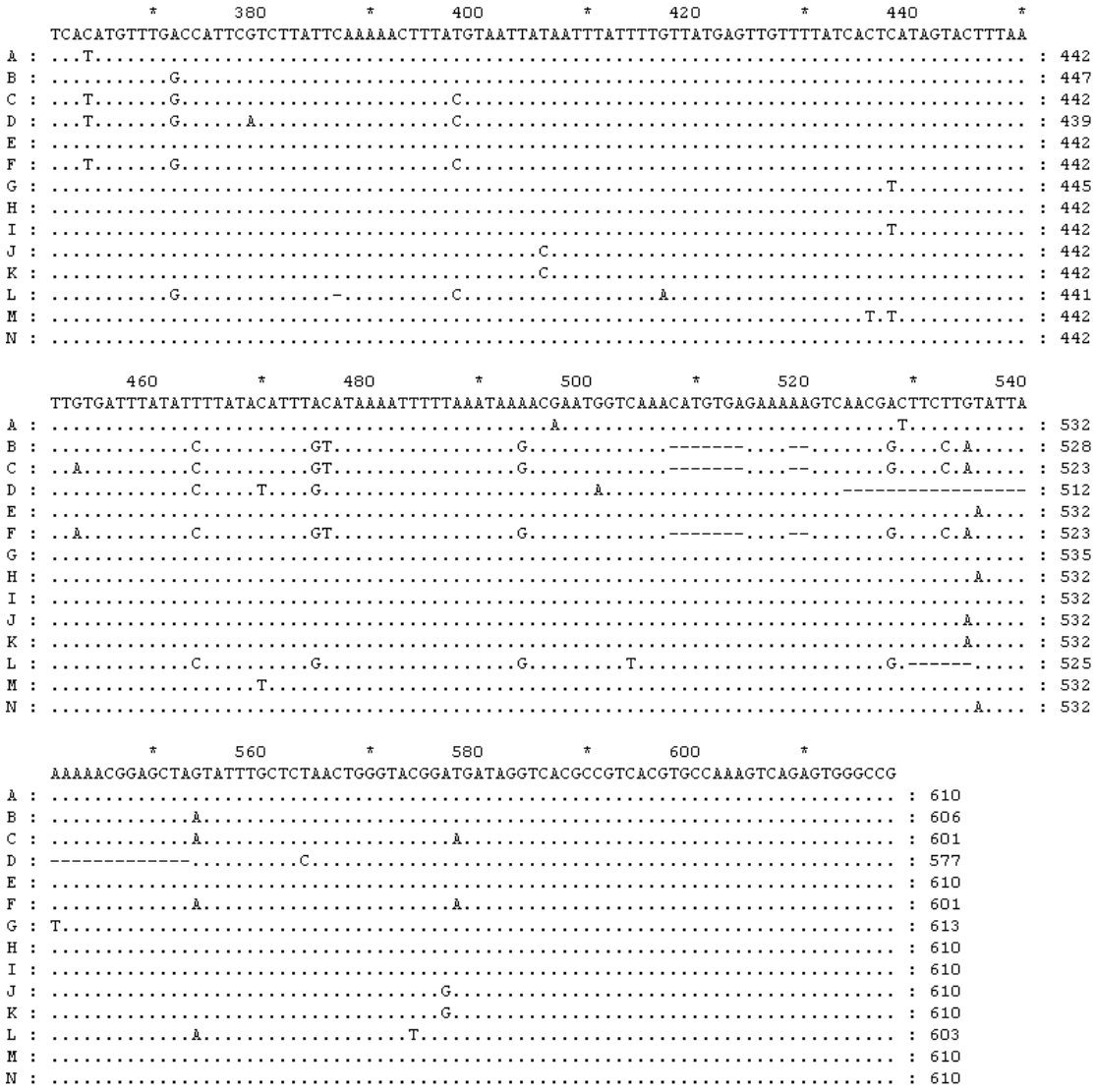
Appendix Figure 6 (Continued)



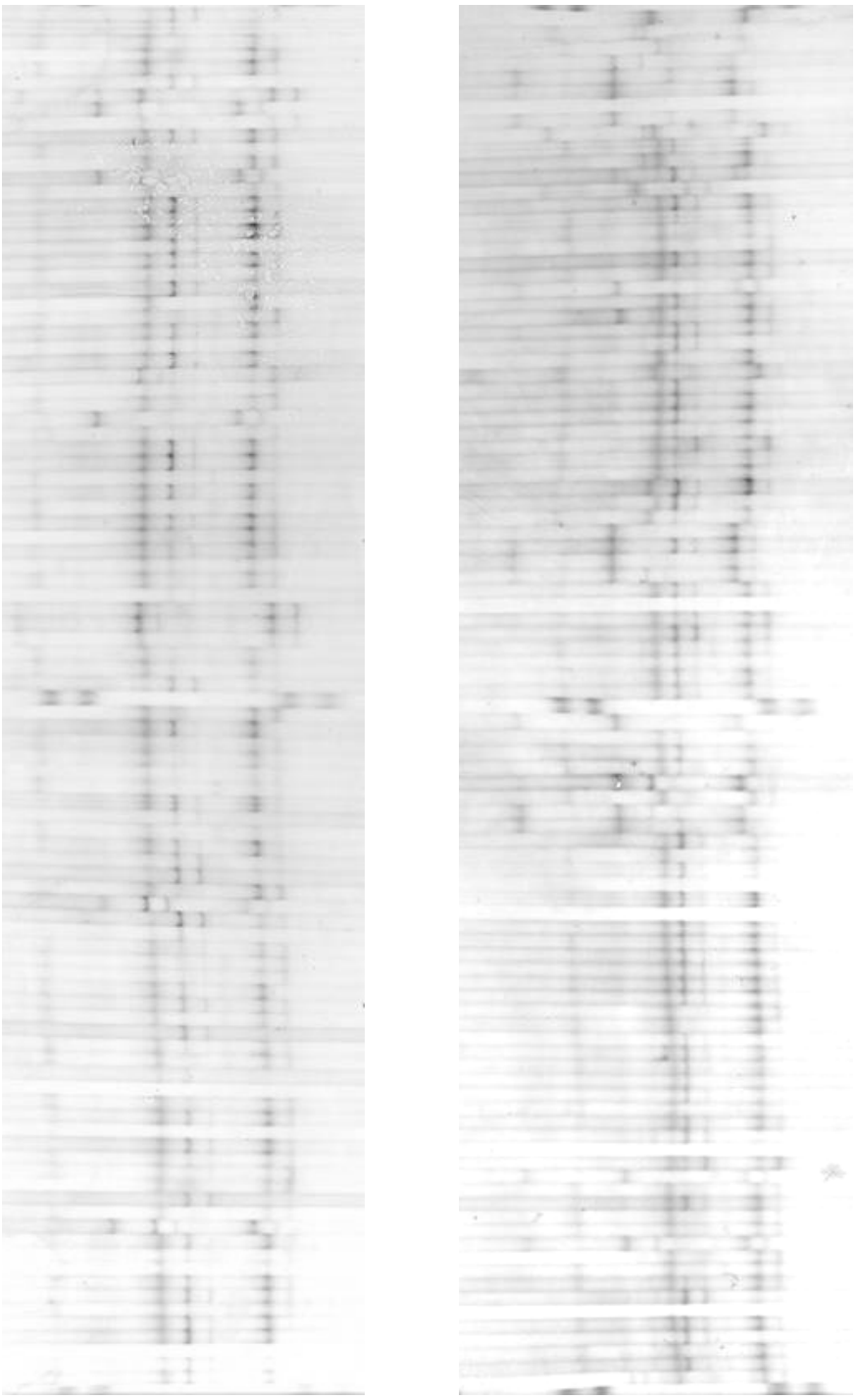
Appendix Figure 7 SSCP analysis of DNA fragments corresponding to the SSSIVa locus separated by polyacylamide gel. Different types of banding pattern were identified (pattern A to N).



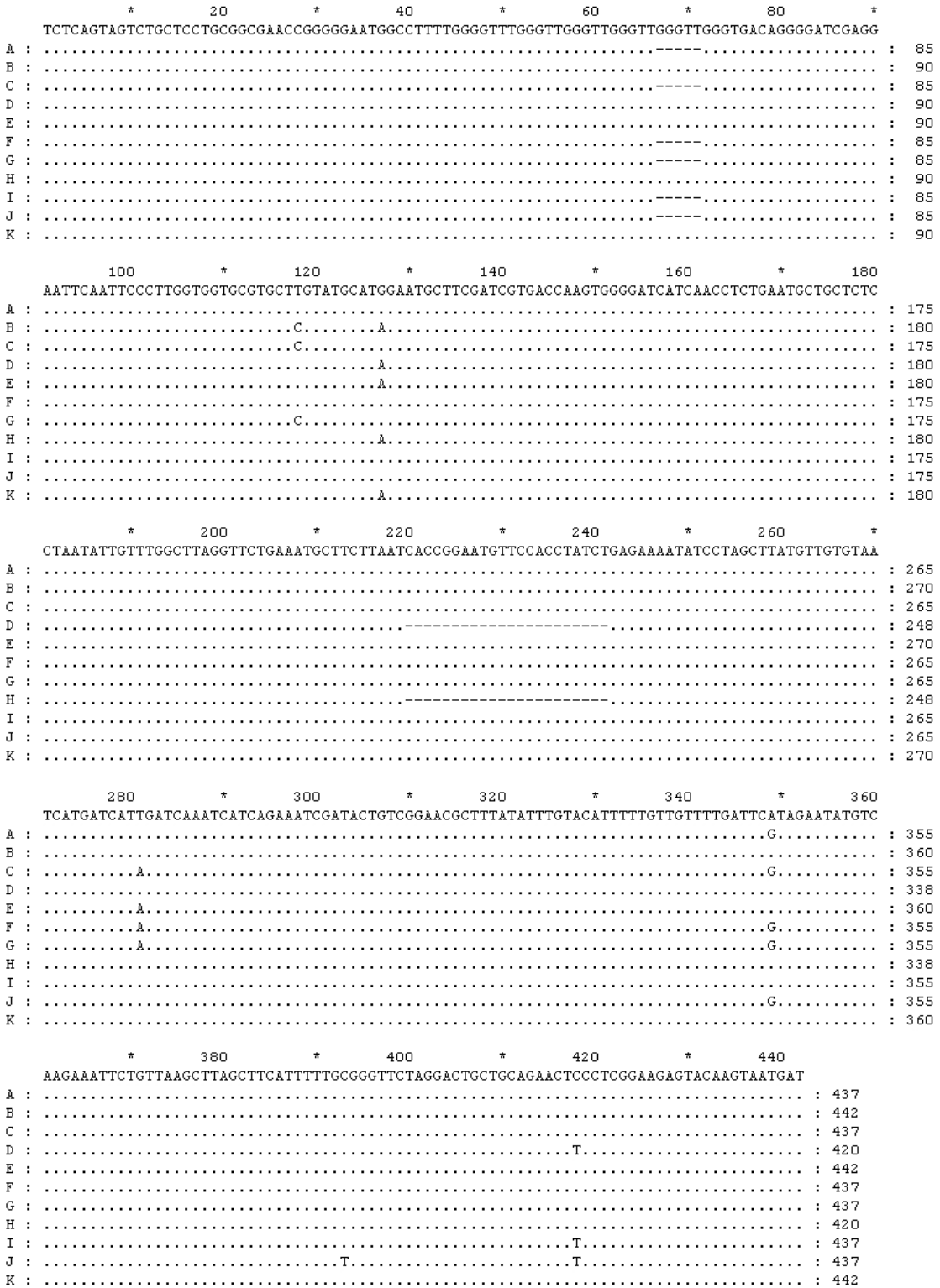
Appendix Figure 8 DNA sequence alignment of the haplotypes at the SSSIVa locus. Fourteen haplotypes are shown.



Appendix Figure 8 (Continued)

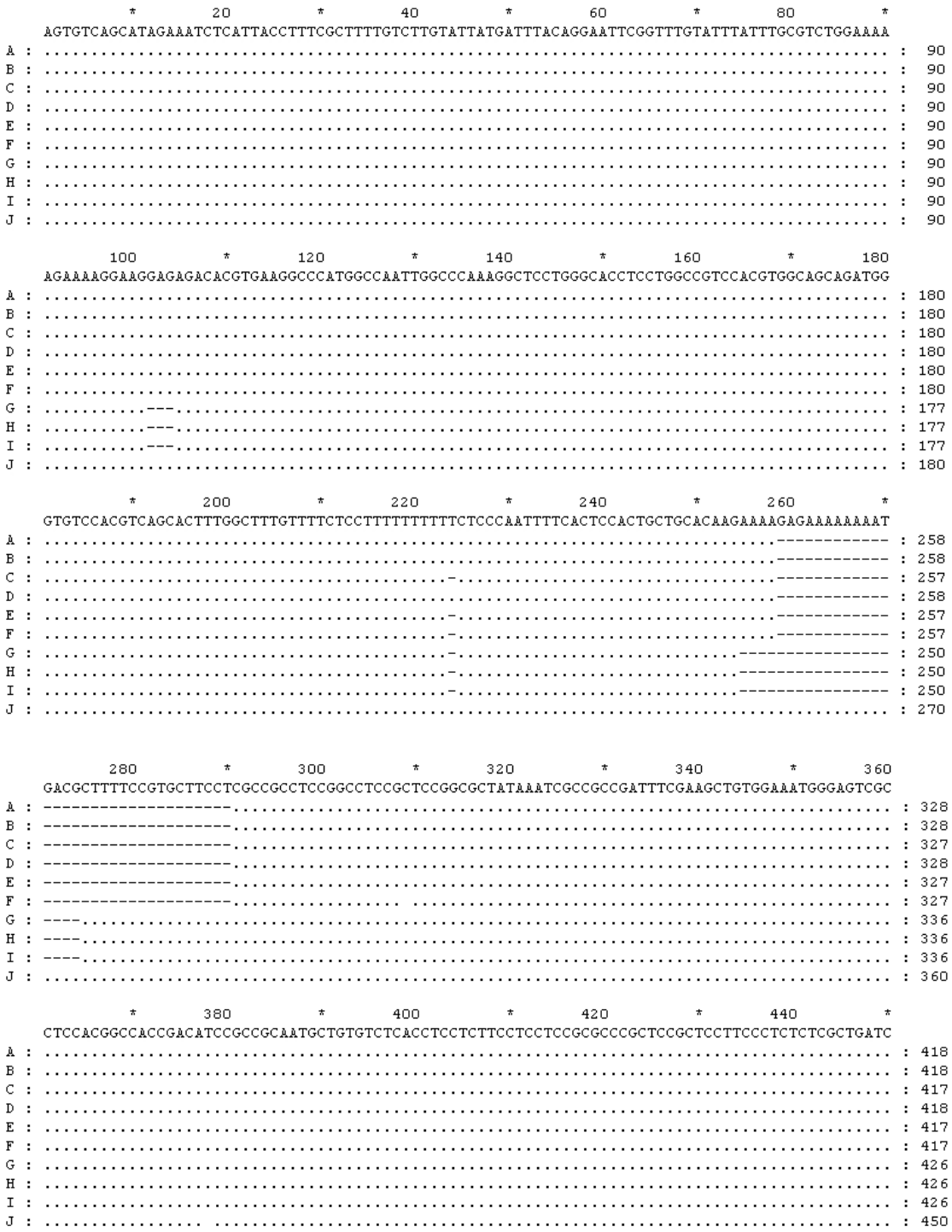


Appendix Figure 9 SSCP analysis of DNA fragments corresponding to the . SSSIVb locus separated by polyacylamide gel. Different types of banding pattern were identified (pattern A to K).

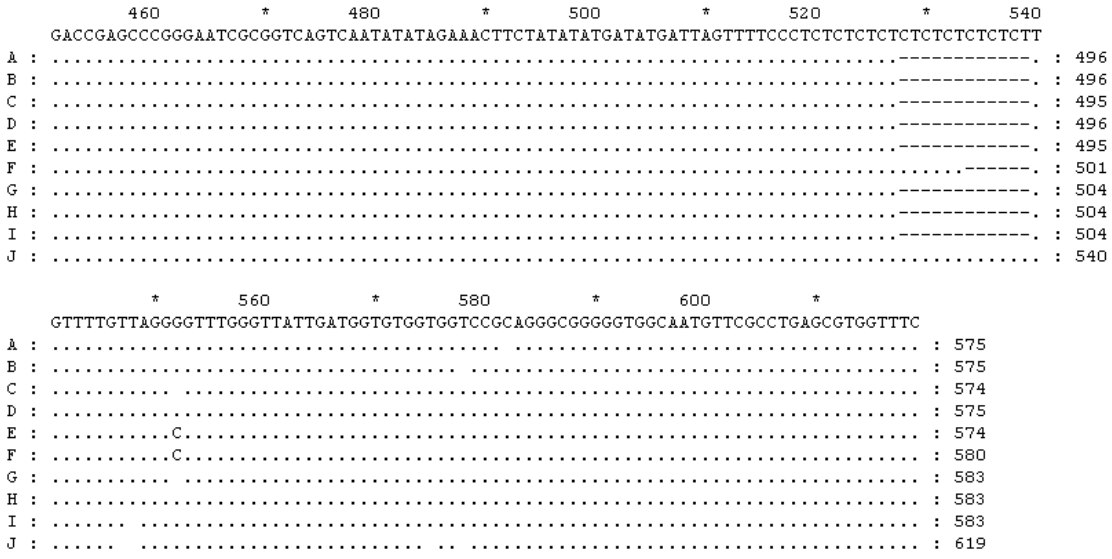


Appendix Figure 10 DNA sequence alignment of the haplotypes at the SSSIVb locus. Eleven haplotypes are shown.

Appendix Figure 11 SSCP analysis of DNA fragments corresponding to the RBE1 locus separated by polyacylamide gel. Different types of banding pattern were identified (pattern A to J).

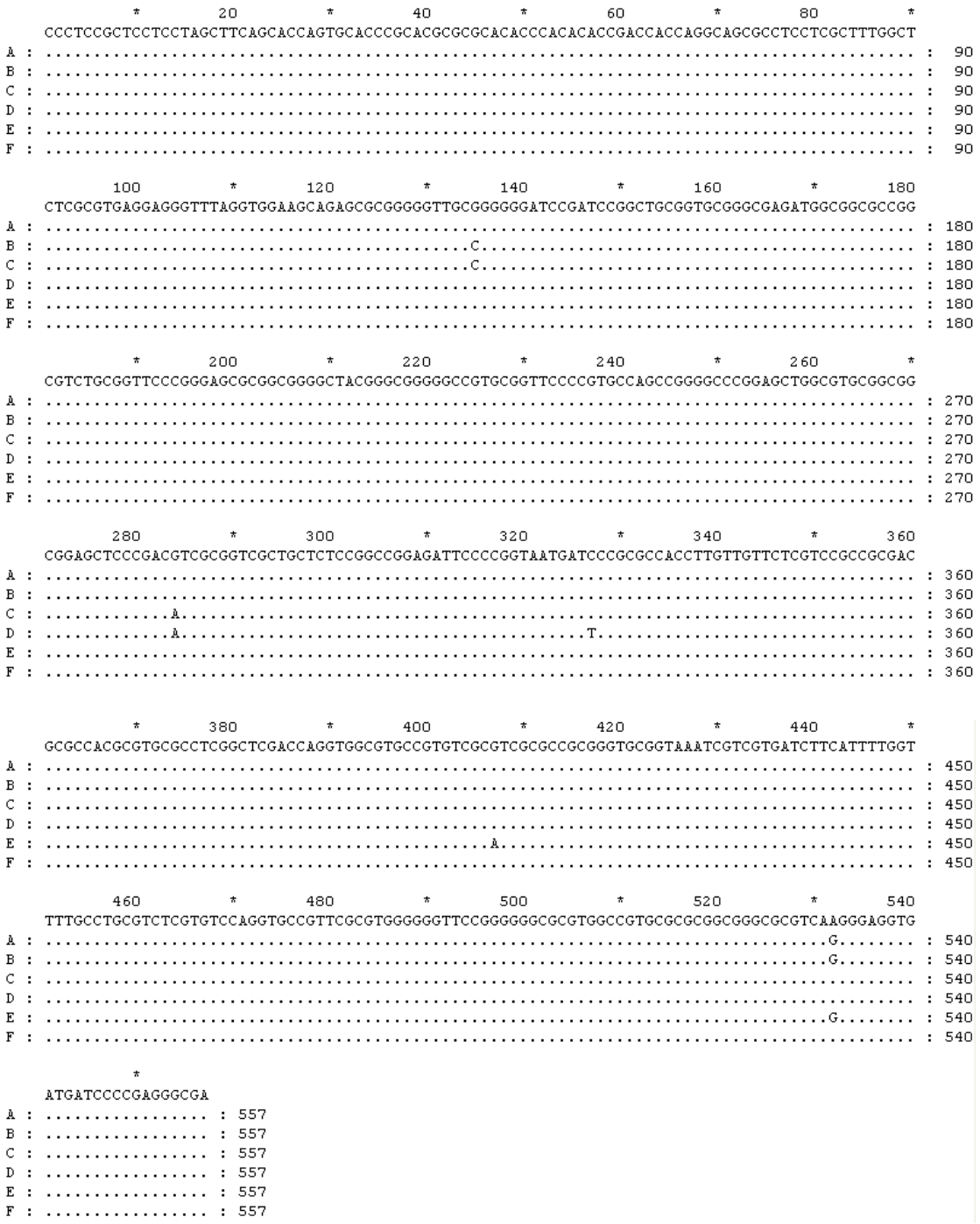


Appendix Figure 12 DNA sequence alignment of the haplotypes at the RBE1 locus. Ten haplotypes are shown.



Appendix Figure 12 (Continued)

Appendix Figure 13 SSCP analysis of DNA fragments corresponding to the RBE3 locus separated by polyacylamide gel. Different types of banding pattern were identified (pattern A to F).



Appendix Figure 14 DNA sequence alignment of the haplotypes at the RBE3 locus. Six haplotypes are shown.