THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Doctor of Philosophy (Tropical Agriculture)

DEGREE

Tropical Agriculture	Agriculture
FIELD	**FACULTY**

TITLE: Molecular Diversity of Aromatic Rice Gene in Different Isozyme Groups

NAME: Miss Khin Myo Myint

THIS THESIS HAS BEEN ACCEPTED BY

THESIS ADVISOR

(Associate Professor Apichart Vanavichit, Ph.D.)

THESIS CO-ADVISOR

(Mr. Somvong Tragoonrung, Ph.D.)

THESIS CO-ADVISOR

(Mr. Jean Christophe Glaszmann, Ph.D.)

THESIS CO-ADVISOR

(Miss Brigitte Courtois, Ph.D.)

GRADUATE COMMITTEE CHAIRMAN

(Associate Professor Yingyong Paisooksantivatana, Ph.D.)

APPROVED BY THE GRADUATE SCHOOL ON _____

DEAN

(Associate Professor Gunjana Theeragool, D.Agr.)

THESIS

MOLECULAR DIVERSITY OF AROMATIC RICE GENE IN DIFFERENT ISOZYME GROUPS

KHIN MYO MYINT

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy (Tropical Agriculture),
Graduate School, Kasetsart University
2012

Khin Myo Myint 2012: Molecular Diversity of Aromatic Rice Gene in Different Isozyme Groups. Doctor of Philosophy (Tropical Agriculture), Major Field: Tropical Agriculture, Faculty of Agriculture.
Thesis Advisor: Associate Professor Apichart Vanavichit, Ph.D. 95 pages

Aromatic rice is an important commodity in international trade. The recessive *Os2AP* gene located on chromosome 8 has been reported to be associated with rice aroma. The 8-bp deletion in exon 7 is an aromatic allele present in most aromatic accessions, however, other mutations associated with aroma have been detected, but in less frequent. In this study, we report an aromatic allele, a 3-bp insertion in exon 13 of *Os2AP*, as a majority allele found in aromatic rice varieties from Myanmar. The insertion is in frame and causes an additional tyrosine (Y) in the amino acid sequence. However, the mutation did not affect the expression of the *Os2AP* gene. A functional marker for detecting this allele was developed and tested in an aroma-segregating F_2 population. The aroma phenotypes and genotypes showed perfect co-segregation in this population. The marker was also used for screening a collection of aromatic rice varieties collected from different geographical sites of Myanmar. Twice as many aromatic Myanmar rice varieties containing the 3-bp insertion allele were found as compared to rice varieties containing the 8-bp deletion allele, which suggested that the 3-bp insertion allele originated in regions of Myanmar.

Existing aromatic rice varieties have been known as belonging in Isozyme group 1, 5 and 6. SSR were able to retrieve the well-established classification into Indica (isozyme group 1), Japonica (group 6, comprising temperate and tropical forms) and specific groups from the Himalayan foothills including some Aus varieties (group 2) and some aromatic varieties (group 5). In this study, they revealed a new cluster of accessions close to, but distinct from, non-Myanmar varieties in group 5. With reference to earlier terminology, we propose to distinguish a group "5A" including group 5 varieties from the Indian subcontinent (South and West Asia) and a group "5B" including group 5 varieties from Myanmar. In Myanmar varieties, aroma was distributed in group 1 (Indica) and in group 5B and very less in group 5A. New *Os2AP* variants were found, including a 43 bp deletion in the 3' UTR that was not completely associated with aroma. Some accessions of group 5B displayed a particular *Os2AP* allele with a 3 bp insertion and 100% association with aroma, suggesting that this allele appears to have remained more localized than 8bp deletion allele which is found in Isozyme group 1,5A and 6.

_____ _____ ___ / ___ / ___

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor, Associate Professor Dr. Apichart Vanavichit. I greatly appreciate for his arrangement, constructive discussions and invaluable comments for my study and suggestions on the separate manuscripts and this thesis. I also would like to express my special thanks to Dr. Somvong Tragoonrung for serving in my committee and for his advice and much helpful discussion to my research. I am especially indebted to my co-advisors, Dr. Jean Christophe Glaszmann and Dr. Brigitte Courtois for their kind arrangements in nine-month collaborative research project at CIRAD, Montpellier, France. I wish to express my gratitude to their interesting and valuable suggestion in research work, the good comments in writing manuscript and thesis.

My special thanks go Thailand International Cooperation Agency (TICA), the Thailand Research Fund (TRF) and the French Embassy in Thailand for fellowship attributed to my Ph.D study in the framework of Thai-French Trilateral Development Cooperation under the Royal Golden Jubilee Programme, Thailand.

I would like to express my special gratitude to all members and my friends in Rice Gene Discovery Unit, Kasetsart University, Kamphaeng Saen campus, Thailand and in CIRAD, Montpellier, France, for their friendship and kind support in my study time. My special thanks go to Dr. Siwaret Arikit for his good training and great support throughout my study and preparing for the manuscript. I also thank in deep to Mr. Anuchar Plabpla and Mr. Siriphat Ruengphayak for sharing their experiences and for their support in conducting laboratory research work and field experiment.

I am grateful to my country and Ministry of Agriculture and Irrigation, Myanmar, for kind permission of my Ph.D study in abroad, providing of the research materials and all supporting for my study. Finally, my heartfelt gratitude goes to my dear parents, my sister and brothers who give me the endless support in my life. I would like to express my special thankfulness and dedicate this degree to them.

Khin Myo Myint
March 2012

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

MOLECULAR DIVERSITY OF AROMATIC RICE GENE IN DIFFERENT ISOZYME GROUPS

INTRODUCTION

Rice is the main cereal consumed by humans. *Oryza sativa,* the main rice species grown in the world, grows in a very wide range of ecological conditions all over the world and encompasses an extremely wide diversity. Wide phenotypic diversity accompanies genetic diversity for all types of traits including grain characteristics. Rice is consumed mostly as cooked grain with little processing, and grain appearance, cooking quality, and taste are the factors that determine grain quality. High-quality rice, as characterized by the physical, chemical, cooking and eating qualities of the grains, is demanded by consumers worldwide.

Grain aroma is one of the most valuable traits judged by all groups of consumers. Aromatic rice is particularly appreciated by consumers due to its flavor and palatability and, for this reason, obtains a premium price in both domestic and international markets. Although most of the trade is from India, Pakistan and Thailand, aromatic rice is cultivated and prized in many other countries of the world. However, the preferences vary in different groups of consumers varying from one ethnic group or geographical region to another, or from country to country. Therefore, advance in the productivity of aromatic rice, achieved in both developed and developing countries have also depended on access to a wide range of diverse types available.

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 becomes a major component of marker assisted selection. So far, rice breeders are interested in the genetic controls of grain aroma and several studies on gene regulations of aroma trait, biosynthetic pathway of aroma compound in aromatic rice plants and gene regulating the trait has been elucidated in some aromatic rice varieties. These studies primarily revealed that a single recessive gene located on

chromosome 8 has been identified as the gene responsible for the aroma trait because it has a key role in the synthesis of aromatic compound, 2-acetyl-1-pyrrolline (2AP). The identified gene has been named differently by different groups as *BAD2* (Bradbury *et al.,* 2005a, 2008), *BADH2* (Niu *et al.,* 2008) and *Os2AP* (Vanavichit *et al.,* 2008). The most important sequence variation on the gene was found in exon 7 as an 8-bp deletion. From the studies of different groups, it was found that several aromatic varieties including world famous aromatic rice accessions; Jasmine or Khao Dawk Mali 105 (Isozyme group I), Basmati (Isozyme group V) and Azucena (Isozyme group VI) were uniquely found to have that same deletion, suggesting that this aroma allele is common by descent in aromatic varieties. From this discovery of aromatic gene, gene specific markers for the aroma characteristic had been developed to access a simple, accurate and inexpensive method for distinguishing between aromatic and non-aromatic rice.

On the other side, it was considered that although the unique genotype (8-bp deletion) determines the aromatic biosynthetic capability in most aromatic rice strains, the aroma content in different aromatic rice varieties does not always correlate with this deletion. The trait may be limited or regulated by other genes or other regions of the known aromatic gene. A previous study has shown that a number of aromatic varieties, primarily from South and Southeast Asia, do not contain the 8-bp deletion and that 2AP is identified in both raw and cooked rice of these varieties (Fitzgerald *et al*., 2008). Studies on diversity of the gene in a large collection of varieties have recently shown that the 8-bp deletion in exon 7 is present in most aromatic accessions but that other less frequent mutations associated with aroma are also detected.

Although aromatic varieties can be found in isozymic groups 1, 5 and 6, their frequency is much higher in group 5 (Singh *et al*., 2000). Glaszmann (1987, 1988) reported that a high proportion of Myanmar varieties belonged to group 5 and that some of these were peculiar in terms of allelic combination, notably at isozyme locus *Amp3* on chromosome 6, as well as in terms of grain quality. Numerous aromatic varieties belonging to Group 5 are now grown under different names. This is interesting to note that the eastern most limits of rice of Group 5 is Myanmar. Rice

belonging to this group have not been found in Southeast and East Asia (Singh *et al*., 2000). Myanmar occupies a key geographic position at the interface between South and South-East Asia. Myanmar is geographically diverse with four recognizable regions (the Ayeyarwady delta, a coastal strip, the central dry plain and mountain ranges). Rice is an important crop for Myanmar, which has the highest per capita consumption of rice in the world: more than 210 kg per person per year. Rice cultivation covers eight million hectares (http://faostat.fao.org , 2009), mainly rainfed lowland rice (54% of the area), but also deepwater rice (24%) irrigated rice (18%), and upland rice (6%) (IRRI, 2002). Because of the interface position and ecological diversity of Myanmar, the genetic diversity of Myanmar rice varieties is expected to be high. The rice preferred by Myanmar people varies depending on locations within the country. However, the majority of Myanmar people prefer intermediate amylose rice (20-25%), which has a soft texture and is aromatic. Good-quality aromatic rice is grown in different regions of Myanmar to fulfill the high local demands. The price of these varieties in local markets is three times the price of non-aromatic rice varieties. In Myanmar, different places require a specific quality of rice for their cooking quality traits. Some of the preferred rice varieties include the popular aromatic varieties such as Paw San Hmwe and Nga Kywe that have a soft texture, high aroma and good kernel elongation, and other rice varieties are locally adapted aromatic rice varieties, which are preferred by local consumers. Locally adapted varieties vary not only in terms of aroma degree, grain shape and grain size but also vary with regard to morpho-agronomic characteristics.

Effective determination of genetic identification on these diverse varieties is essential for efficient utilization of plant genetic resources in crop improvement. Studies on the diversity of sets of varieties from Myanmar have already been conducted using isozymes (Khush *et al*., 2003; Tun 2006) and with molecular markers (Yamanaka *et al*., 2011). However, molecular information regarding the aromatic gene present in these diverse types of aromatic rice varieties is still limited. Moreover, accessions from Myanmar were not widely represented in the worldwide genetic studies conducted so far. Therefore, screening the diverse aromatic and non-aromatic rice varieties in Myanmar will allow for additional information regarding aromatic genes to be gathered for the further studies on the wide range of aromatic rice varieties.

OBJECTIVES

The objective of this study is divided into two main objectives:

Main objective 1

1 To identify the aroma gene and aromatic allele specific to Myanmar aromatic rice varieties.

2. To develop the functional marker for breeding programs and genetic resource conservation programs.

Main objective 2

1. To characterize the genetic diversity of a collection of Myanmar varieties.

2. To relate the diversity in varietal groups with the diversity observed in the *Os2AP* gene.

LITERATURE REVIEW

1. Classification of *Oryza sativa*

Rice (*Oryza sativa* L.) is one of the most important food crops of the world and is the stable food of approximately one-half of the world's population. The organization of *O. sativa* diversity has been of major interest for rice scientists since the early 20th century. The initial and very old distinction of "Hsien" and "Keng" types in China prefigures the main bipolar scheme featuring the current indica and japonica types, well documented in the pioneering work of Kato *et al*., (1928), Matsuo (1952) and Oka (1958). The development of genetic markers, first enzymatic then DNA-based, enabled further refinement of the original classification (Glaszmann 1987; Garris *et al*., 2005). According to generally accepted knowledge, *O. sativa* accessions can be classified in two main groups, the indica group (isozyme group 1), which includes most tropical irrigated and rainfed lowland varieties, and the japonica group (isozyme group 6), a group with a wider range of agro-ecological adaptations. From a genetic viewpoint, the japonica group can be divided into a temperate component comprising irrigated varieties, and a tropical component comprising upland and high elevation varieties, although morphological studies showed that the two components represented a continuum (Glaszmann and Arraudeau 1986). In addition to these two major groups, two smaller groups were identified: the "isozyme group 2" that comprises the aus ecotype from India and Bangladesh as well as other varieties from South Asia and is slightly differentiated from group 1, and the "isozyme group 5" that comprises irrigated varieties from all along the Himalayan border from Iran to Myanmar, including the world famous basmati rices from India and Pakistan and sadri from Iran and is slightly differentiated from group 6 (Glaszmann 1987). Additional groups encompassing peculiar types such as deepwater rices from India and Bangladesh (enzymatic group 3) and rayada floating rices from Bangladesh (isozyme group 4) were identified in isozyme studies (Glaszmann 1987) but not confirmed with other markers. One possible reason for this discrepancy could be their under-representation in the samples analyzed, which in turn, may be explained by the small number of such accessions and the difficulties involved in

obtaining seeds of these highly photoperiodic varieties. Wide phenotypic diversity accompanies genetic diversity for all types of traits including grain characteristics.

The use of the terms indica and japonica is now well established; they should be taken as “predominant in India” and “predominant in Japan”, respectively, and not specific to or originating from India and Japan. Other names with a geographical connotation have been used to designate varietal groups but have been more a source of confusion than of clarification; “sinica” used by Nakagahra (1978) for indica varieties from China and by Chang (1976) for replacing “japonica” given the Chinese origin of this group; “javanica”, used by Chang (1976) to designate varieties from Java (the “bulu” ecotype”, that he described as a derivative of indica whereas they are tropical japonica forms.

More recently, isozyme group 2 tends to be referred to as “Aus” whereas it includes most varieties of the aus ecotype but also many other varieties from South and West Asia, and isozyme group 5 tends to be referred to as “Aromatic”, mainly because it includes the aromatic Basmati varieties, while there are many aromatic rices outside group 5 and many varieties of group 5 are not aromatic.

In our work we use indica and japonica to refer to the well established classification, and isozyme group 2 and group 5 to refer to the groups revealed by molecular markers. Our work clearly addresses the simplistic group5-aromatic assimilation.

2. Aromatic rice

Rice is consumed mostly as cooked grain with little processing, and grain appearance, cooking quality, and taste are the factors that determine grain quality. Aromatic or fragrant rice is rice with natural chemical compounds which give it a distinctive scent. It can be used just like conventional rice for cooking, but adds a new dimension of flavor and aroma to meals. The demand for aromatic rice has increased markedly in recent years to the extent that consumers are willing to pay a premium

price for aromatic rice. The Jasmine type rice of Thailand and the Basmati rice of India and Pakistan are the aromatic cultivars commonly sold in world trade. This rice is highly valued throughout Asia (Baishya *et al*., 2000) and also have wider acceptance in Europe (Berner and Hoff, 1986), Australia (Reinke *et al*., 1991), USA and the Middle East (Shobha Rani *et al*., 2006). Although most of the trade is from Thailand, India and Pakistan, aromatic rice is cultivated and prized in many other countries of the world.

3. Classification of aromatic rice varieties

Genetic variation at isozyme loci was examined to make a precise classification of rice species (Nakagahra, 1978, Glaszmann, 1986). Traditional aromatic rice varieties are classified into three isozyme groups, namely Group 1, Group 5 and Group 6. (Khush *et al*., 2000). Group 1 are Jasmine and include several cultivars from Thailand, Cambodia, Vietnam and Southern China. Among the aromatic rice varieties belonging in Isozyme group 1, Jasmine rice or Khao Dawk Mali 105 (KDML105), due to its sweet aroma and good cooking quality, is appreciated all over the world. Group 6 varieties include Azucena and encompass several cultivars from Indonesia and the Philippines. Most of the cultivars belonging to Group 5 are aromatic and include world famous high quality Basmati and comprise several cultivars from India, Pakistan, Iran, Afghanistan, Bangladesh and China and locally renowned Pawsan Hmwe and Nga Kywe varieties (Glaszmann, 1987; Khush, 2000) and Nama Tha Lay (Singh *et al*., 2000) of Myanmar. Many of rice belonging to this group have excellent lengthwise elongation.

The centre of diversity of aromatic rice of Group 5 are the foothills of Himalayas in the Indian states of Uttar Pradesh (UP) and Bihar, and Tarai region of Nepal. Many aromatic cultivars are still growing in this centre of diversity. From here aromatic rice spread northwestward to Punjab in India and Pakistan, Afghanistan, Iran and Irag, northeast ward to Bangladesh and Myanmar and the Indian states of Orissa, Bengal, Assam and Manipur. The westward distribution occurred to other states of India such as Rajastan, Madhya Pradesh, Maharrastra and Gujarat. Numerous

aromatic varieties belonging to Group 5 are now grown under different names. This is interesting to note that the eastern most limits of rice of Group V is Myanmar. Rice belonging to this group have not been found in Southeast and East Asia. (Singh *et al*., 2000). Employing isozyme analysis method, Khush *et al.* (2003) analyzed various rice varieties from Asian countries. The 82 rice varieties from Asia have been allocated in isozyme group V and the majority (64%) of the varieties (52 rice varieties) was from Myanmar. The total of 1354 accessions from Myanmar stored in the IRRI genebank were used in this study and found that 1115 (82.4%) belonging to group 1 (indica), 17 (1.3%) to group 2, 52 (3.8%) to group 5, 30 (2.2%) to group 6 (japonica) and 140 (10.3%) to group 0 (intermediates).

4. Aroma compound

Chemical analysis of a wide range of rice varieties has revealed many compounds that differ in concentration between fragrant and non-fragrant rice varieties. Using a combination of sensory panels and gas chromatography technique. Buttery *et al.* (1983) determined 2-acetyl-1-pyrroline (2AP), although only present in fragrant rice at low concentrations, was the primary chemical responsible for the characteristic aroma of Jasmine and Basmati rice. 2AP is also present in non-fragrant rice varieties but at a concentration in the range of 10 to 100 times lower than that of fragrant rices (Buttery *et al*., 1983; Buttery *et al*., 1985; Widjaja *et al*., 1996; Wilkie and Wootton, 2004). The threshold concentration at which 2AP can be detected by the human nose is around 0.1ppb when diluted in water (Buttery *et al*., 1983), but is probably somewhat higher in the complex rice grain. A wide range of 2AP concentrations have been observed in both fragrant and non-fragrant varieties in different studies. These differences may be due to the different rice varieties studied, differences in extraction procedure or quantification of 2AP, environmental influences on the level of fragrance such as temperature and salt and drought stress (Itani *et al*., 2004; Yoshihashi *et al*., 2004), harvest time or storage conditions of the rice (Bhattacharjee *et al*., 2002; Itani *et al*., 2004; Yoshihashi *et al*., 2005), whether the rice was milled or unmilled (Buttery *et al*., 1983; Philpot *et al*., 2005) or timing/level of nitrogenous fertilizer application to

the growing plants (Wilkie and Wootton, 2004). 2AP can also be found in other plants, for example, Pandan leaves (*Pandanus amaryllifolius* Roxb.) that contain 2AP at concentrations 10 times higher than the fragrant rice varieties (Buttery *et al.*,1983b), bread flower (*Vallaris glabra*) (Wongpornchai *et al*., 2003) and soybean (Masuda, 1989). Moreover, some microorganisms have also been reported to have abilities to produce 2AP (Romanczyk *et al*., 1995).

5. Aroma evaluation methods for different parts of rice

The pleasant aroma associated with aromatic is not only released in cooked rice but is also often emitted by these varieties in the field at the time of flowering (Weber *et al.*, 2000; Widjaja *et al.*, 1996). It was reported that the 2AP aromatic compound is present in various parts of the rice plant, such as the stems, leaves and grains, but not in the roots (Lorieux *et al*., 1996; Yoshihashi 2002). However, other researches detected 2AP at low levels from rice root and culture media (Vanavichit *et al*., 2005). The following sensory methods have been applied to determine the aroma in rice: chewing several seeds or cooking a sample of seeds from individual plants (Dhulappanavar 1976; Ghose and Butany 1952); heating leaf tissue in water or eluting the aroma from leaf tissue with diluted potassium hydroxide (KOH) (Sood and Siddiq 1978, Hien *et al*., 2006); and heating several half-cut seeds in fresh water (Wanchana *et al*., 2005). However, these sensory evaluation methods are not consistent or reliable because the aroma is subject to human preference. A more reliable method is based on a gas chromatography for quantification of volatile compounds from 100 g cooked rice (Petrov *et al*., 1996). To be more practical to breeders, a method for 1 g of samples was developed (Mahatheeranont *et al*., 2001). From this small volume, 1 ppb of 2-acetyl-1-pyrroline can be detected with repeatability. Recently, a histochemical method was also developed to localize 2AP in plant tissues (Nadaf *et al*., 2006). This method is based on to the structure of 2AP that includes a reactive methyl ketone group, which reacts with 2,4-dinitrophenyl hydrazine to produce an orange-red coloured compound, 2-acetyl-phenyl hydrazone.

6. **The Genetics of Rice Aroma**

The majority of studies which have focused on the genetics of fragrance in rice determined fragrance is due to a single recessive gene (Sood and Sidiq, 1978; Berner and Hoff, 1986; Ahn *et al*., 1992; Bollich *et al*., 1992; Lorieux *et al*., 1996; Garland *et al*., 2000; Cordeiro *et al*., 2002; Jin *et al*., 2003). Genetic mapping of grain aroma were reported as a quantitative trait based on sensory tests. The grain 2AP density was identified in three map locations (Lorieux *et al.,* 1996). The major QTL mapped on chromosome 8 coincided with the consensus genetic map based on sensory test on chromosome 8 (Chen *et al*., 2006; Lorieux *et al.,* 1996). In addition, two minor QTLs were localized on chromosome 4 and 12 (Lorieux *et al.,* 1996). The 4.5 cM map interval between RG1 and RG 28 on the chromosome 8 was considered a critical region for map-based cloning.

The first mapping of grain aroma took place in 1992 (Ahn *et al*., 1992) and the gene responsible for grain aroma was identified 12 years later, with The first and only successful map-based cloning of the gene controlling 2AP (Vanavichit *et al*., 2004, 2005). By taking advantage of within-family segregation for 2AP from the F6 to the F13 generations of the cross between Jasmine and a non-aromatic rice. The original 1.13-Mb region flanked by RG 1 and RG 28 was narrowed down to 82.8 kb, where three KDML 105 Bacterial Artificial Chromosome (BAC) clones were shotgun sequenced, and three candidate genes were identified (Vanavichit *et al*., 2005). ORF3, later named as Os2AP, was determined to be responsible for grain aroma in aromatic rice, because double recombinations within ORF3 resulted in the disappearance of 2AP. Comparative sequence analysis of ORF3 between KDML 105 and Nipponbare revealed that the 4.5-kb genomic sequence contained 15 exons of the 1512bp coding sequence that translated into 503 amino acid in non aromatic Nipponbare. In aromatic KDML 105 and within the exon 7 of *Os2AP*, two important mutation events were found at positions 730(A to T) and 732 (T to A), followed by the 8- bp deletion “GATTAGGC starting at position 734. A second map-based cloning approach cloning was also reported; in a cross between aromatic Kyeema and a cultivar of non- aromatic rice, grain aroma was mapped on chromosome 8 between

SSR markers RM 515 and SSR J07 (Bradbury *et al*., 2005). The *in silico* physical map consisted of four Nipponbare BAC clones spanning the 386-kb flanked SSR markers RM 215 and SSR J07. Re-sequencing one of the BAC clones revealed 17 genes. However, significant sequence variation was identified in only one clone, which was later identified as BAD2, a betadine aldehyde dehydrogenase (BADH) homologe. Based on their similarity at both the nucleotide and amino acid levels, Os2AP and BAD2 were considered the same gene. A third-map based cloning effort was reported using *in silico* physical mapping within critical region by comparing only genomic-end sequence of Nipponbare (Wanchara *et al*., 2005) and by comparing genomic sequences between Nipponbare and 93-11(Chen *et al*., 2008). The conclusion that BAD2 was the best aroma candidate locus identified in the *Azucena japonica* cultivar was also reached using fine-scale mapping using Azucena x IR64 (Bourgis *et al.*, 2008).

7. Mutations in aromatic gene

It was primarily reported that the 8-bp deletion and 3 single nucleotide polymorphisms (SNPs) in exon 7 of *Os2AP* have led to the introduction of premature stop codon to produce a truncated protein which result in abrogation of the function of the enzyme Os2AP consequently accumulate substrate 2AP in fragrant varieties, while the functional *Os2AP* gene codes for a 503 amino acid mature protein which consumes the substrate in non-fragrant varieties.

Identification of the gene for fragrance and availability of large aromatic rice gene pool has created world-wide interest to look for allelic variants at this locus. In addition to 8-bp deletion in exon 7, several variations including a 7-bp insertion in exon 8 (Amarawathi *et al*., 2008); a 7-bp deletion in exon 2 (Shi *et al*., 2008); absence of MITE (miniature interspersed transposable element) in promoter (Bourgis *et al*., 2008); two new SNPs in the central section of intron 8 (Sun *et al*., 2008); a TT deletion in intron 2 and a repeated (AT)n insert in intron 4 (Chen *et al*., 2008) of *badh2* were reported in various fragrant varieties. With an exception to the 8-bp deletion reported as the genetic cause for aroma, it was also found five indigenous

aromatic rice genotypes did not carry this deletion (Sakthivel *et al*., 2006). Similar exceptions were also observed by others in some fragrant varieties (Kuo *et al*., 2005), Navarro *et al*., 2007 and Fitzgerald *et al*., 2008). These studies indicate the existence of allelic/genetic diversity for fragrance in aromatic rice gene pool.

8. Gene Expression and role of *Os2AP* in determining aromatic trait

Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. These products are often proteins, but in non-protein coding genes such as ribosomal RNA (rRNA), transfer RNA (tRNA) or small nuclear RNA (snRNA) genes, the product is a functional RNA. Several steps in the gene expression process may be modulated, including the transcription, RNA splicing, translation, and post-translational modification of a protein. In genetics, gene expression is the most fundamental level at which the genotype gives rise to the phenotype. Measuring gene expression is an important part of many life sciences - the ability to quantify the level at which a particular gene is expressed within a cell, tissue or organism can give a huge amount of information.

The first approach investigating how the aromatic gene functions was achieved by comparing the isogenic lines A 117 and NA 10, which differ only in the 27kb genomic region containing aromatic gene Os2AP (Vanavichit *et al*., 2005). Transcription analysis of Os2AP and flanking candidate genes revealed the differential expression of Os2AP in all parts of rice plant. The compound 2AP is naturally expressed starting from young seedling to the grain-filling period and accumulates in mature grains. The pattern of 2AP expression was consistent with the constitutive expression of functional Os2AP in all plant organs. However, one exception was in the roots (Chen *et al*., 2008). Other researchers detected 2AP and Os2AP transcripts at low levels from rice roots and culture media (Vanavichit *et al*., 2005). In the investigation of transcription levels at different growth stages and in different tissues, total RNA was extracted from young plants (10 days), adult leaves (30 day) and roots (14 days), and flowering panicles of Nipponbare and transgenic plants and used for RT-PCR analysis. The reduction of Os2AP transcripts was highly

significant from 10-20 days after pollination (DAP) (Vanavichit *et al*., 2005). In connection with 8bp-deletion in exon 7 of aromatic allele, the suppressive expression of Os2AP results from premature stop codon at position 753, which shortened full-length peptide to 252 amino acids in aromatic rice (Bradbury *et al*., 2005, Vanavichit *et al*., 2005). Moreover, the suppression of Os2AP gene expression in by RNA interference enhances the accumulation of 2AP in non-aromatic Nipponbare, that could accumulate 2AP in a range of 0.05-0.20 ppm (Vanavichit *et al*., 2005). By contrast, reduction of 2AP level is detected in a transgenic aromatic rice line when it was transformed with functional *Badh2* gene (Chen *et al*., 2008). These finding fully supports genetic mapping evidence to indicate role of Os2AP or BADH2 in 2-AP accumulation in rice. Levels of partial *Badh2* transcripts of aromatic Thai rice cv. KDML105 and its mutants of different aroma levels also corresponds to sensory-tested aroma levels of the rice when analyzed by RT-PCR, which the lower aroma level, the higher transcript level (Srivong *et al*., 2006 and 2007). The *Badh2* expression at translation level using Western blot analysis has been reported that the BADH2 protein was not detectable in Chinese aromatic rice, cv. Wuxiangjing, but nonaromatic rice, cv. Nanjing11 (Chen *et al*., 2008).

To indirectly determine the loss of Os2AP or BADH2 function in aromatic rice, it was of interest to assay BADH activity in rice extracts of aromatic and non aromatic rice. The *Badh2* gene product, BADH2, it is suggested to be responsible to gamma- aminobutyraldehyde (AB-ald) conversion to gamma-aminobutyric acid (GABA). According to Chen *et al*., 2008, gamma- aminobutyraldehyde (AB-ald) is known to be maintained in an equimolar ratio with an immediate 2AP precursor, Δ^1-pyrroline, and the AB-ald levels appears to be an important factor regulating the rate of 2AP biosynthesis. They suggested that the functional BADH2 enzyme (coded by the aroma gene *Fgr*) inhibits 2AP biosynthesis in non-fragrant rice by converting AB-ald, a presumed 2AP precursor, to GABA while the non-functional *badh2* (coded by *fgr*) result in AB-ald accumulation leading to the formation of 2AP in fragrant rice. Bradbury *et al*., 2008 also suggested that gamma- aminobutyraldehyde (AB-ald)is an effective substrate for BADH2 and that accumulation and spontaneous cyclisation of AB-ald to form Δ^1-pyrroline due to a non-functional BADH2 enzyme as the likely

cause of 2AP accumulation in rice. However, in another study, increased expression of Δ^1-pyrroline-5-carboxylate synthetase in fragrant varieties compared with non-fragrant varieties, as well as concomitant elevated concentrations of its product, led to the conclusion that Δ^1-pyrroline-5-carboxylate, usually the immediate precursor of proline synthesized from glutamate, reacts directly with methylglyoxal to form 2AP (Huang *et al*., 2008), with no direct role proposed for BADH2.

9. Molecular markers for classifying aromatic and non-aromatic rice

The advantages of applying markers in selection include managing target traits through genotyping, ensuring that the selection is independent of environmental factors and developmental stages, maintaining recessive alleles in backcrossing, and pyramiding multiple monogenic traits or several quantitative trait loci (QTL) for a single target trait (Xu and Crouch 2008). Since the rice fragrant trait is controlled by a recessive fgr gene and fragrance itself is very difficult to be evaluated, researchers have been striving to develop molecular markers to assist breeders in their selection for fragrant rice genotypes. In an effort to develop PCR based markers for *fgr*, Garland *et al*. (2000) identified a one base pair polymorphism within the RFLP clone RG28. This marker required the use of expensive capillary electrophoresis equipment to discriminate between PCR products that differed by one base pair and because it was physically removed from *fgr* was not 100% accurate in discriminating between non-fragrant and fragrant plants. Cordeiro *et al.* (2002) identified the highly polymorphic single sequence repeat (SSR) markers linked to the fgr gene. By converting the SNP detected between fragrant and non-fragrant rice varieties, Jin *et al.* (2003) developed a SNP marker for the potential use in selecting for the fragrant genotype.

The ideal markers would be the so-called functional markers developed from functional domain(s) within the target gene as these would provide an absolute differentiation between phenotypes (Andersen and Luebberstedt 2003). After the discovery of the aroma gene, the following functional molecular markers have been developed: Aromarker (Vanavichit *et al.*, 2008), allele specific amplification (ASA)

marker (Bradbury 2005b), FMbadh2-E2A and FMbadh2-E2B (Shi *et al*., 2008). However, these markers only cover two types of aromatic alleles, including an 8-bp deletion in exon 7 (for the first two markers) and a 7-bp deletion in exon 2 (for the latter marker) and these markers might not cover the other types of mutations, if they exist.

10. Microsatellite markers for assessment of genetic diversity

Assessment of genetic diversity is an integral part of plant breeding as exploitation of the genetic diversity helps the plant breeder to develop new varieties. Selection of diverse parents with desirable qualities can be done based on analysis using morphological, biochemical and/or DNA markers. DNA markers are the most widely used and are predominant due to their abundance and repeatability; they remain unaffected across different stages, seasons, locations and agronomic practices. Of the several classes of DNA markers available, microsatellite or simple sequence repeat (SSR) markers are considered as most amenable for several applications including genetic diversity studies due to their multiallelic nature, high reproducibility, co-dominant inheritance, abundance and extensive genome coverage (McCouch *et al.*, 2002) Polymorphisms are easily detected as length variations of fragments, usually on polyacrylamide gels. A carefully chosen set of SSR markers providing genome-wide coverage will facilitate an unbiased assay of genetic diversity, thus giving a robust, unambiguous molecular description of rice cultivars. SSR markers have been effectively used to study genetic diversity among closely related rice cultivars (Zhu *et al*., 2004, Bonow *et al*., 2009) and distantly related genotypes(Giarrocco *et al*., 2007, Jayamani *et al*., 2007). However, only a few efforts have been made to characterize aromatic rice varieties including basmati varieties (Agarwal *et al*., 2002, Siwach *et al*., 2004).

11. Sequencing technology for allelic diversity

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). Analysis of single nucleotide polymorphism (SNP) and small insertion/deletion (InDel), which are the basis of most differences between alleles, has been simplified by recent developments in sequencing technology. 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. In comparison to individual SNPs, haplotype analysis of a group of linked SNPs is more informative in determining association with the phenotypes. Haplotype and genetic diversity analysis across the BADH2 region demonstrated that the haplotype groups can be classified into two clusters; *Japonica* Gene Haplotype Cluster (Jap_GH) and *Indica* Gene Haplotype Cluster (Ind_GH) (Kovach *et al*., 2009). The study on SNP variation in the BADH1 gene on chromosome 4 (homologue of BADH2 on chromosome 8) demonstrated two groups of protein haplotypes representing indica rice varieties and the japonica group of varieties. It was also identified that the SNP haplotype and protein haplotype group having the loss of function allele in BADH1 gene also related to the accumulation of aromatic compound 2AP. (Singh *et al*., 2010).

MATERIALS AND METHODS

Part I: Identification of aroma allele in Myanmar aromatic rice

Plant Materials

Two sets of rice samples were used in this study. The first set was a group of 52 Myanmar landrace rice varieties that belong to Isozyme group V (Khush *et al*., 2003); this group was kindly provided by the Plant Biotechnology Center at the Myanma Agriculture Service in Myanmar. The first set used the tracking system of the International Rice Germplasm Collection (IRGC) accession numbers. The second set consisted of 25 diverse aromatic rice varieties from unknown Isozyme groups that were collected from different geographical regions of Myanmar; this set of rice varieties was provided by the Seed Bank at the Department of Agricultural Research in Myanmar. Based on the variety names and passport data, all varieties in the second set were considered to be aromatic rice.

PCR genotyping for *Os2AP* gene mutations

The first set of 52 Myanmar varieties together with the two control varieties, KDML105 (aromatic) and ISL-10 (non-aromatic), were used for characterizing the *Os2AP* gene mutations. All plants were grown in an experimental field at Kasetsart University (Nakhon Pathom, Thailand) in July 2008.

Leaf samples for all plants were collected from 3-week-old seedlings. Genomic DNA was extracted from 1 g of leaf tissues according to the DNA Trap method developed by the DNA Technology Laboratory at Kasetsart University (Nakhon Pathom, Thailand). The cut leaf sample was put in 1.5 ml tube. The leaf tissue frozen in liquid nitrogen and powdered with a sterile plastic rod was mixed with 500ul extraction buffer. The extract was then homogenized for 10 s with a vortex and incubated at 65°C for one hour. After 1/10 volume of Neutralizing buffer was added, the tube was placed on ice and incubated for 10 min. Centrifuged the

tube at 10,000 rpm for 10 min and then transferred the supernatant into a new 1.5-ml tube. The 200 µl of Bind-mix solution was added and mixed thoroughly. After briefly centrifuged and discarded the supernatant, pellet was washed by 1 ml of propanol-wash solution. Then the pellet was collected by briefly centrifugation, and the supernatant was removed. Washed pellets again with 1 ml of 95% ethanol, and air-dried. DNA was eluted from the pellet by adding 100 ul of dH_2O and incubating at 65 ^{o}C for 30 min.

Two PCR markers, Aromarker (Vanavichit *et al*., 2008) and FMbadh2-E2B (Shi *et al*., 2008), were developed for detecting the 8-bp deletion in exon 7 and the 7-bp deletion in exon 2 of *Os2AP*, respectively. These markers were used to amplify genomic DNA from the 52 rice varieties. PCR was performed in a total volume of 10 µl containing 2 µl of genomic DNA (50 ng/µl), 1 µl of 10x buffer, 1 µl of 25 mM $MgCl_2$, 2 µl of 1 mM dNTPs, 0.5 µl of each primer (5 µM) and 0.1 µl of Taq DNA polymerase (Fermatas; Life Science, USA). The PCR analysis was initiated by denaturation at 94°C for 3 min followed by 35 cycles of 94 $^{\circ}$C for 30 s, 55 $^{\circ}$C for 30 s and 72 $^{\circ}$C for 1 min. A final incubation at 72 $^{\circ}$C for 7 min was allowed for the completion of primer extension. A portion (2 µl) of the PCR products was resolved on a 4.5% polyacrylamide gel and visually examined after silver staining. The amplified PCR products were resolved by 4.5% polyacrylamide gel electrophoresis (PAGE) and visualized by silver staining.

Aroma evaluation by sensory test and 2AP quantification

The different sensory test methods were performed in different tissues for aroma evaluation. For the first method, the aroma was evaluated from one-month-old leaves by incubating cut leaves in a diluted KOH solution and noting the aroma according to the method of Sood and Siddiq (1978). For the second method, the aroma was evaluated by chewing premature and mature seeds. For the third method, aroma was evaluated by direct smelling of two-month-old callus samples cultured on Murashige and Skoog media (1972). All sensory methods were performed by five trained panels. To get the more accurate result, the 2AP content was analyzed from

grain samples (5 g) using an improved GC-MS method developed for detecting 2AP in a small scale of sample (as little as 0.5 g; Wongpornchai *et al*., 2004).

Reverse Transcription PCR (RT-PCR)

To investigate transcription levels at different growth stages and in different tissues, total RNA was extracted from callus (two-month-old cultured in Murashige and Skoog (1962) media), leaves (one-month-old) and roots from aromatic and non-aromatic varieties using a High Pure RNA Isolation Kit (Roche Diagnostics GmbH, Roche Applied Science, Germany). Reverse transcription polymerase chain reaction (RT-PCR) was performed according to the manufacturer's protocol for the Superscript™ III First Strand Synthesis System for RT-PCR (Invitrogen ™ Life Technologies, UK). AromarkerBig (5 'TGCTCCTTTGTCATCACACC 3') and (5' CCATGCAACCATCCTTTCTT 3') was used as the *Os2AP* gene-specific primer. Actin primers (5'AGGGCTGTTTTCCCTAGTATCGTG3' and 5' GATGGCATGAGGAGG GGCAT 3') were used as a positive internal control. RT-PCR products were analyzed on a 1% agarose gel in 1x TBE buffer.

Sequencing of the *Os2AP* gene in selected aromatic rice varieties

The full-length *Os2AP* gene was sequenced in aromatic varieties from Pathein Nyunt, Yangon Saba and Kyet Paung. Primer pairs designed to amplify fragments covering the entire *Os2AP* gene were provided by the Rice Gene Discovery Unit at Kasetsart University in Thailand. Amplified PCR products were purified by a Gel/PCR DNA Fragment Extraction Kit (Geneaid, Taiwan) and cloned into a pGEM®-T Easy cloning vector (Promega, Madison, WI). Plasmids from positive colonies were purified using a High Speed Plasmid Minikit QIAprep Spin Miniprep Kit (QIAGEN, USA) and then sequenced using ABI PRISM® BigDye™ Terminators (Applied Biosystem/Perkin-Elmer, San Jose, CA, USA). The sequences of each variety were assembled based on identical overlapping regions using the CAP3 Sequence Assembly Program (Huang X and Madan A. 1999). The genomic sequence of the non-aroma allele of the *Os2AP* gene on chromosome 8 (GenBank Accession

No. AP005537) of *japonica* rice (*Oryza sativa japonica* cultivar Nipponbare) was retrieved from the NCBI website (http://www.ncbi.nlm.nih.gov). Multiple sequence alignment of the *Os2AP* alleles among Myanmar aromatic rice varieties and the non-aromatic rice variety, Nipponbare, was performed using the CLUSTALW program (Thompson *et al*., 1994).

Development of an aroma-specific functional marker for Myanmar aromatic rice

The genomic sequence of the aromatic variety, Pathein Nyunt, was used as a reference sequence for the aroma allele with the 3-bp insertion. The PCR marker, 3In2AP, was developed to detect this sequence variation. The forward (5' GTCCTGTTCAATCTTGCAGC 3') and reverse (5' CTTGATGCAACCATGTCATA3') primers were designed using Primer 3 (version 0.4; http://frodo.wi.mit.edu/primer3) by applying the constraints of 20 nucleotides in length and a GC content of 40–60% to avoid primer self-complementarity. The PCR marker was verified by amplifying the genomic DNA extracted from the leaf, seed and callus samples of different aromatic and non-aromatic rice varieties. PCR was performed in a total volume of 10 µl containing 2 µl of genomic DNA (50 ng/µl), 1 µl of 10x buffer, 1 µl of 25 mM $MgCl_2$, 2 µl of 1 mM dNTPs, 0.5 µl of each primer (5 µM) and 0.1 µl of Taq DNA polymerase (Fermatas; Life Science, USA). The PCR analysis was initiated by denaturation at 94 $^{\circ}$C for 3 min followed by 35 cycles of 94 $^{\circ}$C for 30 s, 55 $^{\circ}$C for 30 s and 72 $^{\circ}$C for 1 min. A final incubation at 72 $^{\circ}$C for 7 min was allowed for the completion of primer extension. A portion (2 µl) of the PCR products was resolved on a 4.5% polyacrylamide gel and visually examined after silver staining.

Application of functional markers for genotyping F_2 populations for aroma

To evaluate the accuracy and reliability of the newly developed marker, 3In2AP, 160 F_2 plants derived from a cross of Pathein Nyunt (aromatic variety; 3-bp insertion) and ISL-10 (non-aromatic variety; isogenic line derived from KDML105 x CT9993 crosses) were genotyped. The PCR amplicons were resolved on a 4.5% polyacrylamide gel, and the segregation ratio according to the banding patterns was

observed. Phenotypic evaluation was also performed by a sensory test with the leaves of 160 F_2 plants according to the method of Sood and Siddiq (1978) and by chewing the F_3 seeds of individual F_2 lines.

Screening aromatic alleles in aromatic rice germplasms from Myanmar

To validate the utility of the 3In2AP marker and to screen the aromatic rice germplasms from Myanmar, 25 diverse aromatic rice varieties collected from different geographical regions of Myanmar were analyzed. Based on the variety names and passport data, all varieties in this set were considered as aromatic rice. The sensory tests, based on leaf samples, were repeated for confirmation following the protocol of Sood and Siddiq (1978).

Genomic DNA was extracted from 1 g of leaf tissues according to the DNA Trap method. Three markers for the different aroma alleles (Aromarker for the 8-bp deletion and 3In2AP for the 3-bp insertion) were used to amplify the target regions on the *Os2AP* gene. KDML105 (aromatic variety with the 8-bp deletion), Pathein Nyunt (aromatic variety with the 3-bp insertion) and ISL-10 (non-aromatic variety) were used as positive and negative control varieties. PCR was performed in a total volume of 10 µl containing 2 µl of genomic DNA (50 ng/µl), 1 µl of 10x buffer, 0.4 µl of 25 mM $MgCl_2$, 2 µl of 1 mM dNTPs, 0.4 µl of each primer (5 µM) and 0.1 µl of Taq DNA polymerase. The PCR analysis was initialized by denaturation at 94 °C for 3 min followed by 35 cycles of 94 $^{\circ}$C for 30 s, 55 $^{\circ}$C for 30 s and 72 $^{\circ}$C for 2 min followed by a 7 min incubation at 72 $^{\circ}$C. A portion (2 µl) of the PCR products was subjected to electrophoresis on a 4.5% polyacrylamide gel, and the banding pattern was examined after silver staining.

Part II: Genetic diversity of aromatic rice varieties

Plant Materials

The total of 147 varieties from Myanmar listed in Appendix Table 1 was used in this study. A first subset of 96 accessions including 62 aromatic and 34 non-aromatic varieties originating from different geographical regions in Myanmar were provided by the Myanmar Seed Bank ("MSB" ID prefix). The isozyme group to which these accessions belonged was unknown. The aromatic/non aromatic quality of the varieties of this set had already been determined by the Myanmar Seed Bank based on sensory tests. The sensory tests, based on leaf samples, were repeated for confirmation following the protocol of Sood and Siddiq (1978). The second subset of 51 accessions including 3 aromatic, 48 non aromatic varieties from Myanmar were provided by Myanma Agriculture Service ("IRGC" ID prefix). According to the results of Khush *et al*. (2003), all the members of the second subset belong to isozymic group 5. The aromatic/non aromatic quality of the varieties of this two set was determined in experiment part 1.

This set of Myanmar varieties was completed with 80 varieties listed in Appendix Table 2 extracted from a core collection representing the varietal group diversity of *Oryza sativa* for which the enzymatic group had already been determined (Glaszmann *et al*. 1995). This set, hereafter referred to as "reference set", was composed of 20 accessions from group 1 (indica), 9 from group 2 (aus/boro), 31 from group 5 (sadri/basmati), and 20 from group 6 (japonica), so group 5 accessions were over-represented. The information on the aromatic/non-aromatic nature of these accessions came from Kovach *et al*. (2009) and, for those that were not tested by these authors, from breeders' knowledge (Appendix Table 2). Those for which the information was not available were labeled "unknown" (u).

Genomic DNA Extraction

DNA was extracted from one plant per accession using the MATAB method (Risterucci *et al*., 2000). Five hundred mg of tissue frozen in liquid nitrogen and powdered with a mortar was mixed with 5 ml of extraction buffer (1.4 *M* NaCl, 100 m*M* Tris HCl pH 8.0, 20 m*M* EDTA, 10 m*M* Na_2SO_3, 1% PEG 6000, 2% MATAB) preheated to 74°C. The extract was then homogenized for 10 s with a vortex and incubated 30 min at 74°C; after being cooled to 20°C, an equal volume of chloroform-isoamyl alcohol (24:1 v/v) was added, followed by emulsification. The tube was then centrifuged at 7000 *g* for 30 min and the supernatant was precipitated at –20°C overnight after the addition of an equal volume of isopropanol. The DNA was removed with a glass hook and re-suspended in 1 ml of 0.7 *M* NaCl, 50 m*M* TRIS-HCl, 10 m*M* EDTA, pH 7.0 buffer.

PCR amplification and genotyping by SSR markers

A set of 19 SSR markers distributed on the 12 chromosomes was genotyped on the 147 Myanmar accessions and the 31 reference set accessions from group 5. The details of SSR markers are presented in Table 3. The other 49 reference set accessions had already been genotyped with the same markers (Courtois *et al*., unpublished results). The genotyping was performed according to the protocol of Roy *et al*. (1996) implemented with the automated infrared fluorescence technology of LI-COR 3200 sequencers (Li-COR, Lincoln, Nebraska, USA) at CIRAD genotyping and robotics platform, Montpellier, France. Primer sequences were retrieved from the Gramene database (www.gramene.org). For a given SSR locus, the forward primer was designed with a 5'-end M13 tail (5' CACGACGTTGTAAAACGAC 3').

PCR amplifications were performed in a Eppendorf Mastercycler (Eppendorf, Hamburg, Germany) on 25 ng of DNA in a final volume of 10 µl of buffer (10 mM Tris-HCl (pH 8), 100 mM KCl, and 0.5 mM MgCl2) containing 0.08 µM of the M13-tailed primer, 0.1 µM of the other primer, 200 µM of dNTP, 1 U of Taq DNA polymerase (Invitrogen, Carlsbad, California, USA), and 0.1 µM of M13 primer-

fluorescent dye IR700 or IR800 (Eurofins-MWG, Ebersberg, Germany). The PCR programme included an initial denaturation cycle at 95 °C for 4 min, 35 cycles at 94 °C for 1 min, Tm for 1 min, and 72 °C for 1 min, and a final elongation step at 72 °C for 8 min. IR700 or IR800-labeled PCR products were diluted 7-fold and 5-fold respectively, subjected to electrophoresis in a 6.5% polyacrylamide gel and then sized by the IR fluorescence scanning system of the sequencer. Allele calling was performed twice by two different persons based on five DNA pools of known allele size included in each gel and used as standards.

PCR amplification and genotyping by functional markers of aroma

To characterize the *Os2AP* gene mutations, the 178 accessions including Myanmar accessions and reference varieties were also genotyped with two other primer pairs, the first one to detect an 8 bp deletion and the second one to detect a 3 bp insertion in the coding sequence of the *Os2AP* gene. The sequences of the two primer pairs are listed in Appendix Table 3. The protocol for PCR amplification and genotyping were performed as mentioned in SSR genotyping.

Statistical analyses

(a) Genetic distance between accessions and population structure

PowerMarker version 3.25 (Liu and Muse 2005) was used to calculate the allele number, allele frequencies and PIC (polymorphism information content) of each marker.

Both a distance-based and a model-based approach were used to assess the genetic structure of the whole population (Myanmar accessions and the reference set). A dissimilarity matrix was computed using a shared allele index with DARwin software (Perrier and Jacquemoud-Collet 2006). An unweighted neighbor-joining (NJ) tree was built based on this dissimilarity matrix. The number of sub-populations, K, in the population was assessed in parallel using the model-based program

Structure v2.3 (Pritchard *et al*., 2000). The program was run with the following parameters: haploid data, possibility of admixture and correlated allelic frequencies. A run was composed of 100,000 burn-ins followed by 100,000 iterations. Ten replicates were performed for each value of K, with K varying from 1 to 10. For each run, the percentages of admixture in the K sub-populations of all accessions were computed. The best replicate giving the maximum likelihood were chosen as the final result for each K. An accession was discretely classified in a sub-population when the admixture percentage of the accession for the sub-population concerned was above 80%. If the maximum value was less than 80%, it was classified as admixed (m).

To determine the most likely K value, the criteria proposed by Evanno *et al*. (2005) was also used related to the first and second order rates of change of the likelihood function with respect to K.

(b) Genetic variation within and between populations

The hierarchical distribution of the molecular variance within and between the sub-populations defined by Structure was assessed by analysis of molecular variance (AMOVA) with Arlequin (Excoffier *et al*., 2006). To evaluate the genetic differentiation between these populations, pair-wise F_{ST} (Wright 1978) were computed with the same software, using 1000 permutations to determine their significance.

(c) Distinction between two-subpopulations

XLSTAT tree classification tools (Breimann *et al*., 1984) under Excel (http://www.xlstat.com/) was used to determine which markers or marker combinations best distinguished sub-populations in Isozyme group 5 rice varieties.

Sequencing

The *Os2AP* or *BADH2* gene (gene Os08g32870) was retrieved from Gramene; position in MSU v6 chromosome 8: 20,377,081 to 20,383,348 bp). The 91 accessions: 50 aromatic, 7 non-aromatic and 3 unknowns from the Myanmar collection (Appendix Table 1), and 17 aromatic, 5 non-aromatic including Nipponbare, and 9 unknowns from the reference set (Appendix Table 2) were sequenced. The 12 primer pairs covering the whole *Os2AP* gene listed in Appendix Table 3 came from Vanavichit *et al*. (2008), Kovach *et al*. (2009) or Myint *et al*. (2012). Amplification products were purified and sequenced at GATC, Germany. The sequences, except that of Nipponbare, were deposited in the EMBL-Genbank database under the accession numbers [GenBank: JQ308346 to JQ308435].

Haplotype analysis

Sequence quality control, alignment and nucleotide polymorphism detection were performed using Codon Code Aligner v1.6.3 (http://www.codoncode.com/index.htm) with minimum Phred scores set at 20. Polymorphism sites were numbered starting with position 1 for the A of the ATG of the first exon (position in MSU v6: 20,377,257) based on the gene annotation in Gremene. Haplotypes were manually determined. Haplotype networks, representing unique alleles separated by mutational steps, were constructed with NETWORK v4.5 (http://www.fluxus-technology.com/sharenet.htm) using the median-joining method (Bandelt *et al*. 1999), with equal weight for all sites and an eta-parameter of 0. Both SNP and Indels were used, with the Indels coded in biallelic form as SNPs.

RESULTS AND DISCUSSION

Part I: Identification of aroma allele in Myanmar aromatic rice

Results

Aroma evaluation and *Os2AP* genotyping of Myanmar rice varieties

A collection of 52 Myanmar landrace rice varieties belonging to Isozyme group 5 (Khush *et al*., 2003), was used for the characterization of *Os2AP* gene alleles. The aroma characteristics of the varieties were evaluated by sensory test methods as follows: incubating cut leaves in diluted KOH and chewing several seeds and detecting aroma by sniffing callus tissue. Only three Myanmar rice varieties (Pathein Nyunt, Yangon Saba and Kyet Paung) out of the 52 varieties were detected and confirmed by sensory methods to be aromatic rice. As analyzed by GC-MS, the 2AP content was present at an intermediate level in the grains of the two aromatic varieties (0.92 ppm for Pathein Nyunt and 0.86 ppm for Yangon Saba) when compared to Thai Jasmine rice (1.3 ppm for KDML105) and another Myanmar rice (1.46 ppm for Kyet Paung).

All of the 52 varieties were subsequently genotyped by the two aromatic gene-specific markers to detect the 8-bp deletion in exon 7 and the 7-bp deletion in exon 2. The deletions were not present in all of the non-aromatic varieties. Interestingly, out of the three aromatic rice varieties, only the Kyet Paung variety contained the 8-bp deletion in exon 7. The other two aromatic rice varieties, Pathein Nyunt and Yangon Saba, did not contain the 8-bp deletion or 7-bp deletion.

Expression of *Os2AP* in the two Myanmar aromatic rice varieties

The expression of *Os2AP* was analyzed by RT-PCR in the two Myanmar aromatic rice varieties, Pathein Nyunt and Yangon Saba (these two varieties did not

contain gene mutations in exon 2 or exon 7), to investigate whether this gene had a role in determining the aromatic trait in these two varieties. RT-PCR analysis demonstrated that the expression level of the *Os2AP* gene in the two varieties did not differ from that of the non-aromatic rice, ISL-10, but the gene expression level in KDML105 (the aromatic variety containing the 8-bp deletion) was highly reduced (Figure 1).

Sequence variation in the *Os2AP* gene in the two aromatic Myanmar rice varieties

The *Os2AP* gene in the two aromatic rice varieties was sequenced to further investigate whether the other type of gene mutation was present in this gene, even though no effect on gene expression was observed. Gene sequencing revealed a 3-bp insertion (TAT) in exon 13 at the 1257-1259 position of the 1512-bp cDNA sequence in both the Pathein Nyunt and Yangon Saba varieties. The *Os2AP* gene was also sequenced in the aromatic variety (Kyet Paung). The sequencing result confirmed the presence of an 8-bp deletion in this aromatic variety (Figure 2).

The genomic DNA sequence of two Myanmar aromatic rice varieties were translated to protein by using Expasy translation tool (http://web.expasy.org/translate/). The multiple alignments of protein sequences of *Os2AP* compared among aromatic and non-aromatic rice varieties. The 3-bp insertion found in the two Myanmar aromatic rice varieties was an in-frame translation that caused an additional amino acid (tyrosine, Y) at the position 420 of amino acid sequence to be translated into the cDNA sequence of the *Os2AP* gene of the two aromatic varieties. The total amino acid sequence length encoded by the *Os2AP* gene in the two varieties was 504 amino acids (aa), and the normal length found in the non-aromatic variety was 503 aa (Figure 3).

Development of a functional marker specific to the 3-bp insertion in exon 13 of *Os2AP*

A new PCR marker, 3In2AP, was developed for detecting the 3-bp insertion (TAT) in exon 13 of the *Os2AP* gene (Figure 4 a). The primer was validated by amplifying the genomic DNA of the two aromatic varieties, Pathein Nyunt and Yangon Saba, as well as KDML105 and ISL-10, which were used as controls. As a result, different banding patterns were observed between the two Myanmar aromatic varieties as compared to KDML105 and ISL-10. The sizes of the PCR products amplified by the 3In2AP marker were 197 bp for aromatic varieties with the 3-bp insertion and 194 bp for KDML105 and ISL-10 (Figure 4b).

Application of the 3In2AP marker for genotyping the F_2 population

The aromatic variety, Pathein Nyunt, and non-aromatic variety, ISL-10, were used as parental lines to produce an F_2 population segregating for aroma. The 3In2AP marker was used to genotype the parental lines and F_1 plants. The polymorphic PCR bands were 197 bp in Pathein Nyunt and 194 bp in ISL-10, and both bands (197 and 194 bp) were observed in the F_1 hybrid. The marker was then used to genotype 160 lines of the F_2 population derived from a single F_1 plant of the cross between Pathein Nyunt and ISL-10. As a result, 33 lines were identified as homozygous to the aromatic variety, Pathein Nyunt, and 80 lines were heterozygous. Moreover, 47 lines carried the allele of ISL-10 (Fig. 5). The segregation ratio of the three genotypes in the F_2 population agreed with the expected ratio of 1:2:1 ($\chi^2 = 2.450$; $P = 0.2938$), which corresponded to the Mendelian segregation of a single gene (Table 1). The phenotypic evaluation performed by a sensory test with the leaf samples of F_2 plants showed that 33 lines were aromatic and 127 lines were non-aromatic. The numbers of the aromatic and non-aromatic F_2 lines showed a ratio of 1:3 ($\chi^2 = 1.633$; $P = 0.2013$). The segregation of phenotypes also agreed with the Mendelian segregation ratio for a single recessive gene that controls a trait (Table 1). According to these results, the phenotypes and genotypes of all 160 F_2 progenies were correlated.

All 33 aromatic lines contained the same 3-bp insertion allele, which suggests that this single recessive allele is associated with the aroma phenotype (Figure 5).

Application of the 3In2AP marker to screen aromatic rice germplasm from Myanmar

For screening the aromatic rice germplasm from Myanmar, Aromarker and 3In2AP were used to detect the two different aroma alleles in 25 aromatic rice varieties. The KDML105 and Pathein Nyunt aromatic varieties were used as positive control varieties, and ISL-10 was used as a negative control variety. The functional markers clearly showed that this set of Myanmar aromatic varieties carries either the 8-bp deletion or the 3-bp insertion aroma alleles; three varieties, which were not detected by the two functional markers, were exceptions. Sixteen aromatic varieties carried the 3-bp insertion aromatic allele, and six aromatic varieties carried the 8-bp deletion aromatic allele (Table 2). The distribution of the aromatic rice varieties containing the two different aromatic alleles in different geographical regions of Myanmar is shown in Figure 6.

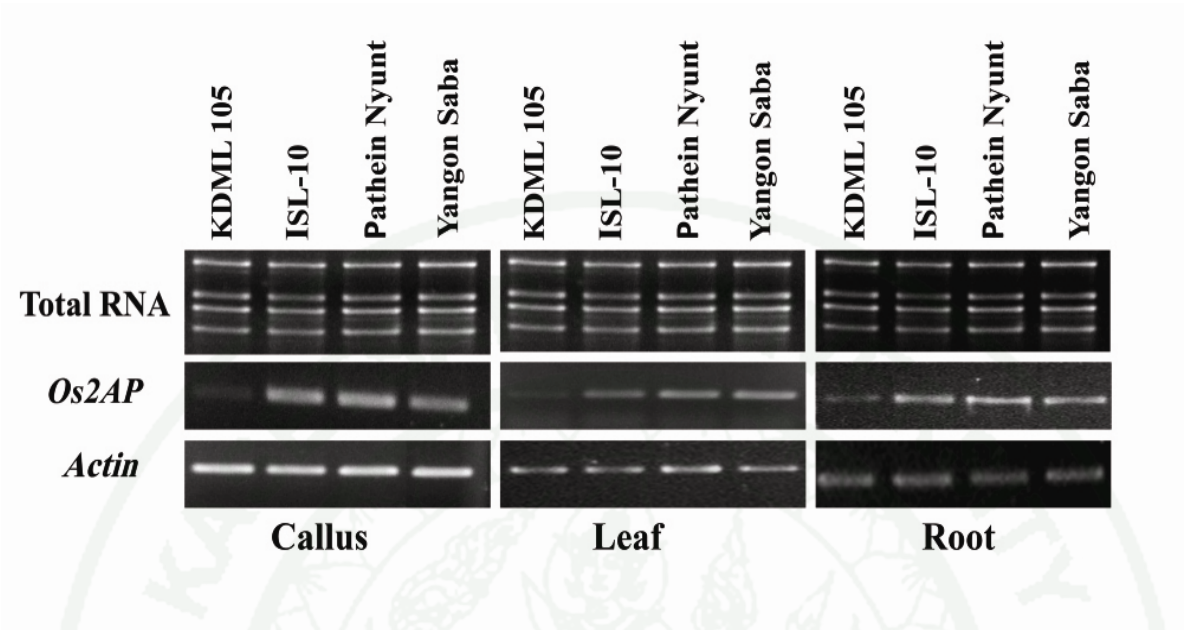

Figure 1 RT-PCR analysis of *Os2AP* gene expression in callus, leaf and root between aromatic and non-aromatic lines.
KDML 105 is the aromatic variety with the 8-bp deletion, and ISL-10 is a non-aromatic variety. Pathein Nyunt and Yangon Saba are aromatic varieties with the 3-bp insertion. The actin gene was used as a control because it was expressed at the same level in all tissues.

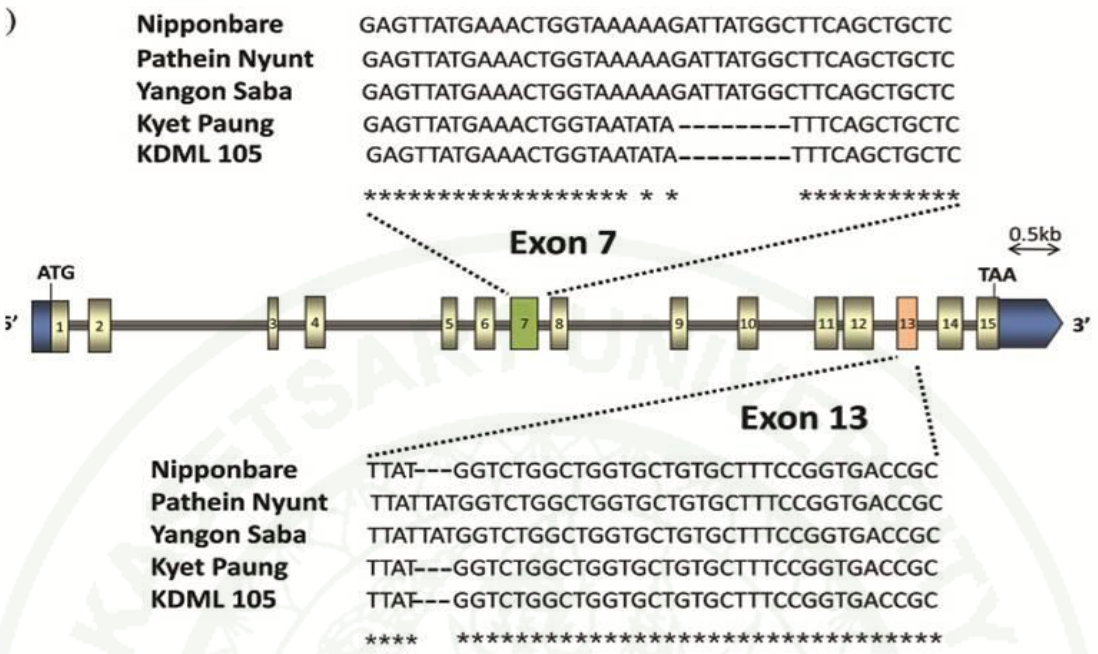

Figure 2 Nucleotide sequence alignments showing two aromatic alleles; 8-bp deletion in exon 7 and 3-bp insertion in exon 13 of *Os2AP* gene. Nipponbare is a non- aromatic rice variety. KDML105 is an aromatic rice variety containing the 8-bp deletion. Kyet Paung is a Myanmar aromatic rice variety containing the 8-bp deletion. Pathein Nyunt and Yangon Saba are Myanmar aromatic rice varieties containing the 3-bp insertion.

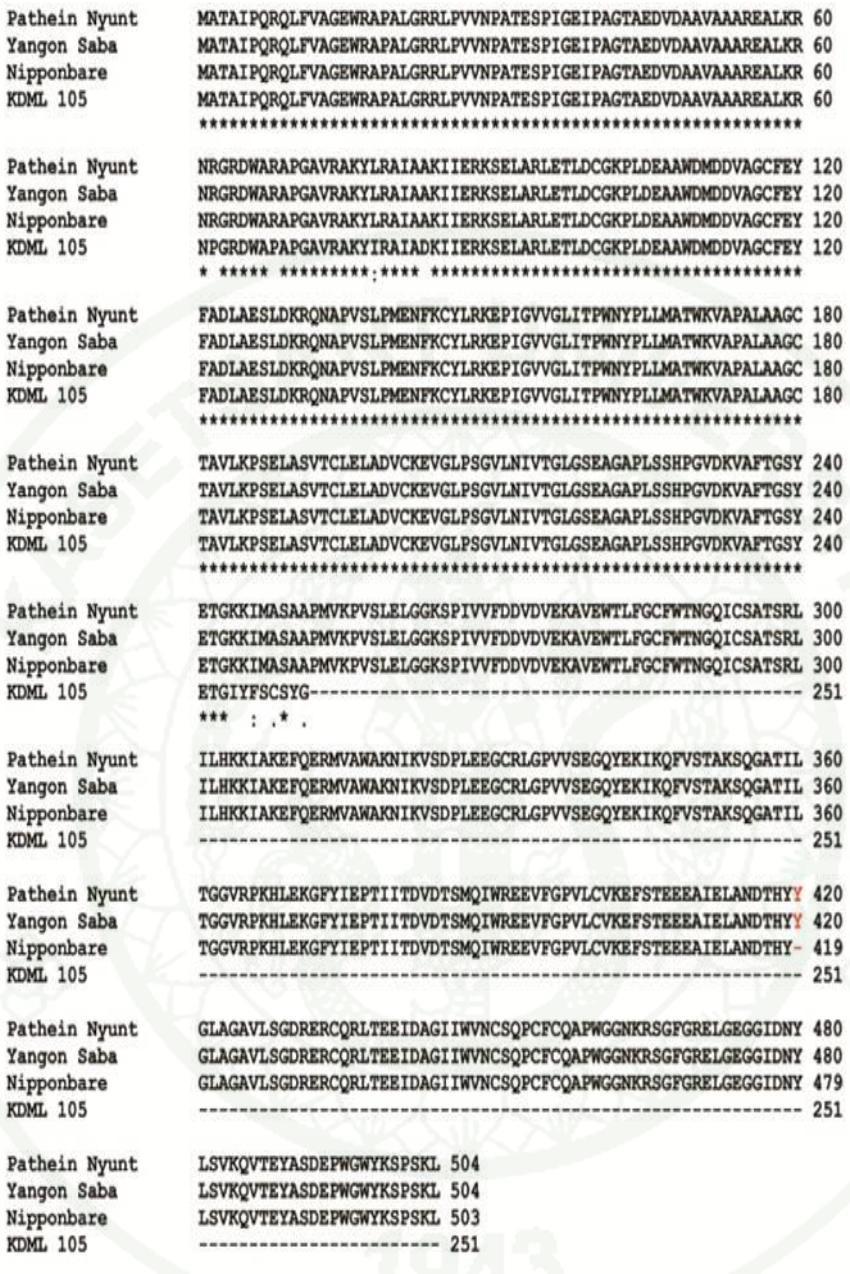

Figure 3 Multiple alignments of deduced protein sequences of *Os2AP* compared among aromatic and non-aromatic rice varieties. An additional amino acid, tyrosine (Y), in the Pathein Nyunt and Yangon Saba varieties is highlighted in red.

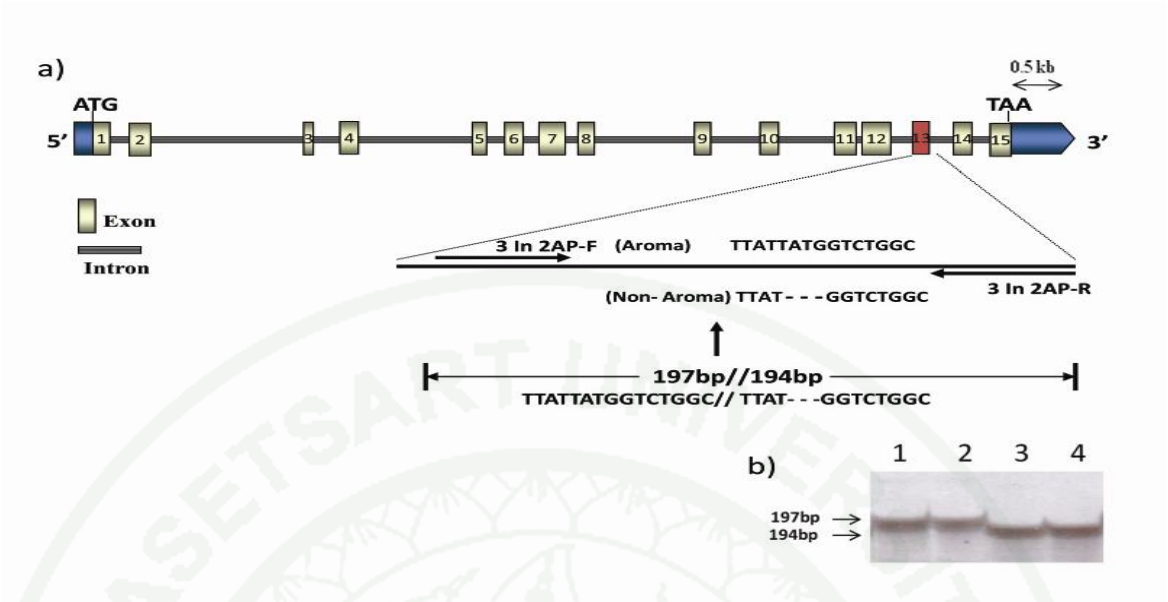

Figure 4 (a) The schematic of the 3In2AP marker location on exon 13 of *Os2AP*. The sequence variation, 3-bp insertion (TAT), and expected sizes of the PCR products are shown.

(b) The banding patterns of the 3In2AP marker are provided. The follows lanes are shown: Lane 1, Pathein Nyunt (aromatic rice with 3-bp insertion); Lane 2, Yangon Saba (aromatic rice with 3-bp insertion); Lane 3, KDML105 (aromatic rice with 8-bp deletion); and Lane 4, ISL-10 (non-aromatic rice). The PCR product sizes are indicated by arrows.

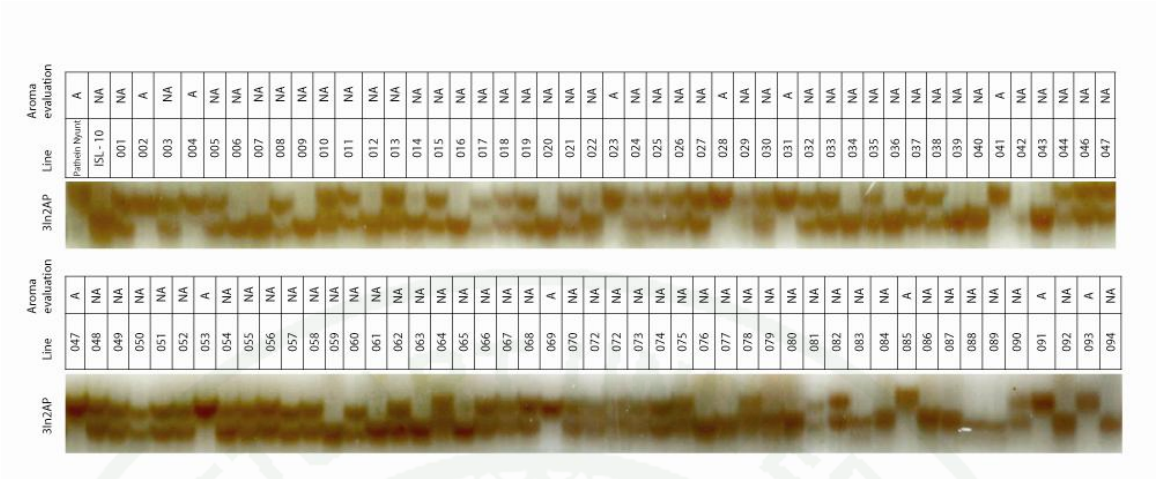

Figure 5 The segregation pattern, as revealed by the 3In2AP marker, of the representative 94 F_2 plants from the cross between Pathein Nyunt and ISL-10. The aroma phenotypes are shown as A (aromatic) and NA (non-aromatic).

Table 1 Chi-squared test for phenotypic and genotypic segregation in the F_2 plants derived from the Pathein Nyunt and ISL-10 cross.

	Number			
	Observed	Expected	$\chi 2$	P-value
Phenotype (expected ratio is 1:3)				
Aromatic	33	40	1.225	
Non-aromatic	127	120	0.408	
Total	160	160	1.633	0.2013
Genotype (expected ratio is 1:2:1)				
Pathein Nyunt	33	40	1.225	
Pathein Nyunt /ISL-10	80	80	0.000	
ISL-10	47	40	1.225	
Total	160	160	2.450	0.2938

Table 2 List of Myanmar aromatic rice varieties with different aroma alleles on the *Os2AP* gene.

No.	Accession No.	Variety	Mutation	Region
1	MSB 682	Sabani Hmwe	8-bp deletion	Southern coastal region
2	MSB 804	Paw San Bay Kyar	3-bp insertion	Delta area
3	MSB 1128	Paw San Yin	3-bp insertion	Delta area
4	MSB 1158	Bahan Hmwe	8-bp deletion	Delta area
5	MSB 1207	Paw San Bay Kyar	8-bp deletion	Delta area
6	MSB 1382	Nga Bya Yin	3-bp insertion	Semiarid region
		Kauk Hnyin Net		
7	MSB 1789	Nga Kywe Yin	3-bp insertion	Eastern Plateau region
8	MSB 1791	Tyaung pyan Yin	3-bp insertion	Delta area
9	MSB 1863	Paw San Bay Kyar	8-bp deletion	Delta area
10	MSB 1871	Nga Pyar gyi	3-bp insertion	Semiarid region
11	MSB 2297	Kamar Kyi Saw	3-bp insertion	Eastern Plateau region
12	MSB 2579	Paw San Bay Kyar	3-bp insertion	Delta area
13	MSB 2620	Paw San Hmwe	3-bp insertion	Delta area
14	MSB 3163	Paw San Bay Kyar	3-bp insertion	Delta area
15	MSB 3708	Bay Kyar Paw San	3-bp insertion	Southern Plain region
16	MSB 6701	KaukHnyin Hmwe	8-bp deletion	Eastern Plateau region
17	MSB 6097	Kauk Hnyin Phyu	3-bp insertion	Eastern Plateau region
18	MSB 6801	Kauk Hnyin Net	3-bp insertion	Eastern Plateau region
19	MSB 6973	Paw San Bay Kyar	3-bp insertion	Western coastal region
20	IRGC 33552	Pathein Nyunt	3-bp insertion	Delta Area
21	IRGC 33858	Yangon Saba	3-bp insertion	Southern Plain region
22	IRGC 58069	Kyet Paung	8-bp deletion	Semiarid region
23	MSB 2502	Paw San Hmwe	Neither	Delta Area
24	MSB 2877	Paw San Bay Kyar	Neither	Delta Area
25	MSB 1628	Taung Pyan Hmwe	Neither	Eastern Plateau region

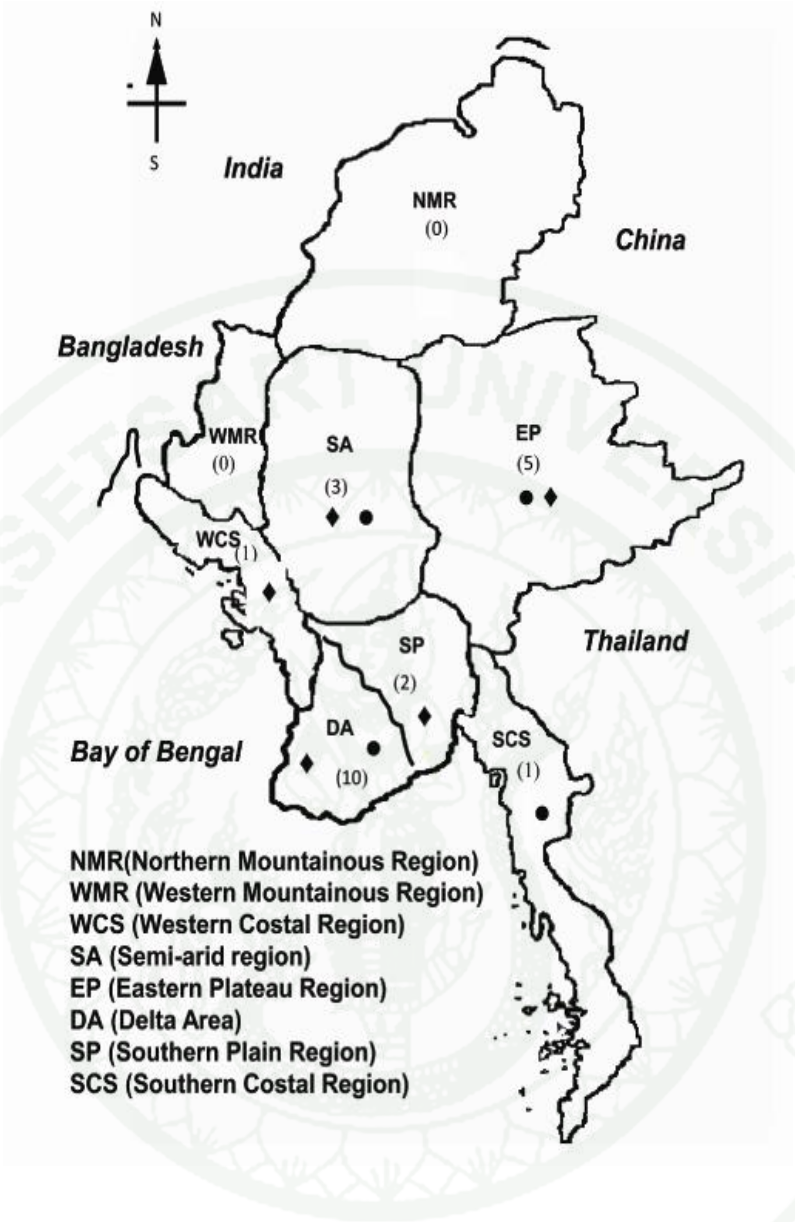

Figure 6 The geographical locations of the aromatic varieties collected in Myanmar. The number of varieties collected from each region is shown.
(♦) Collection site of aromatic rice varieties with the 3-bp insertion.
(●) Collection site of aromatic rice varieties with the 8-bp deletion.

Discussion

A single recessive gene located on chromosome 8 controls the aroma trait in rice. After the gene was identified, many scientists have studied a large collection of aromatic rice varieties. The 8-bp deletion in exon 7, which leads to 2AP accumulation in aromatic rice, has been found as the major allele in most aromatic accessions, and other types of mutations with less frequency have also been reported (Bradbury *et al*., 2005a; Bourgis *et al*., 2008; Shi *et al.,* 2008; Kovach *et al.,* 2009). Interestingly, the present study found that there were twice as many Myanmar aromatic rice varieties containing the 3-bp insertion allele as those containing the 8-bp deletion allele. It is likely that the aromatic rice varieties with the 3-bp insertion in the *Os2AP* gene (resulting in a tyrosine insertion into the protein at the position 420 of 503 amino acid) contained less 2AP in the grains as compared to the varieties containing the 8-bp deletion. The *Os2AP* expression level in the aromatic varieties with the 3-bp insertion was not reduced as compared to the non-aromatic variety, which suggested that the Os2AP protein could be produced in these aromatic lines. However, a recent study has demonstrated that the enzyme encoded by *Os2AP* from aromatic rice containing the 3-bp insertion exhibits less catalytic efficiency toward 4-aminobutanal. The tyrosine (Y) insertion at the position 420 could alter the conformation of a loop located near the NAD^+ binding site resulting to the slight reduction of the binding affinity toward NAD^+ and substrates (Wongpanya *et al*., 2011). The mutant may account for the accumulation of 1-pyrroline, the immediate substrate of 2AP, generating aroma in the Myanmar aromatic rice. It is possible that the *Os2AP* gene with the 3-bp insertion can partially function and that the *Os2AP* gene with the 8-bp deletion cannot function at all, which may explain the lower 2AP level in the varieties containing the 3-bp insertion as compared with the variety containing the 8-bp deletion.

The identification of multiple mutations for 2AP enables rice breeding programs to actively select for multiple genetic sources of 2AP in different aromatic rice varieties. The availability of a functional marker to detect the different mutations will facilitate marker-assisted selection aimed at developing aromatic rice varieties. A

functional marker to detect the 3-bp insertion was developed in this study. The validity of the newly developed marker, 3In2AP, was tested on F_2 segregation lines of the Pathein Nyunt and ISL-10 cross. The 3In2AP marker predicted the phenotype of all the F_2 plants segregating for aroma, and it clearly distinguished the homozygotes from heterozygous F_2 plants. Due to its co-dominant nature, this marker can be potentially useful in testing large segregating materials of aromatic rice improvement programs through a marker-assisted breeding program. Moreover, the assay for genotyping rice aroma using the 3In2AP marker is a simple and robust method for screening rice populations for aroma. The PCR products can be easily and inexpensively analyzed by polyacrylamide gel electrophoresis (PAGE) with silver staining or by sophisticated high-throughput experiments.

The screening of aromatic varieties from Myanmar demonstrated that the majority of aromatic rice from seven different Myanmar regions had the 3-bp insertion in exon 13 and that several varieties from three other regions had the 8-bp deletion in exon 7. Most of the prominent aromatic varieties, such as Paw San Hmwe (MSB 2620), Paw San Yin (MSB 1128) and Paw San Bay Kyar (MSB 804), and landrace varieties, such as Pathein Nyunt (IRGC 33552) and Yangon Saba (IRGC 33858), contained the 3-bp insertion. The collection site of these varieties with the 3-bp insertion aromatic allele was the delta region of Myanmar. Moreover, the Paw San Bay Kyar (MSB 1863) and Taung Pyan Yin (MSB 1791) varieties collected from the same delta region carried the aroma allele with the 8-bp deletion. The locally adapted aromatic rice varieties (sticky aromatic rice) in the Eastern Plateau region, Kauk Hnyin Phyu (MSB 6097) and Kauk Hnyin Net (MSB 6801), contained the 3-bp insertion, and the Kauk Hnyin Hmwe (MSB 6701) variety contained the 8-bp deletion. In a recent study on the origin and evolution of aroma in rice, Kovach *et al*. (2009) reported that five Myanmar aromatic varieties, Paw San Hmwe (IRGC 97793), Paw San Hmwe (N/A), Yangon Saba (IRGC 33858), Emahta Longyu (IRGC 33064) and Balugyun IRRI (IRGC 32960), contain the 3-bp insertion and that three varieties, Bokehmwe ((IRGC 32985), Ma Waine Ohn (IRGC 33357) and Paw San Hmwe (IRGC 33571), contain the 8-bp deletion in exon 7. The results of this study and previous studies demonstrate that the 3-bp insertion is the major aroma allele in

Myanmar rice varieties and that most of the prominent aromatic rice varieties of Myanmar originated in the delta regions of Myanmar. In summary, the molecular information collected from Myanmar rice varieties can be used to study a large panel of aromatic rice varieties and their diversity at a whole genome level. Furthermore, the functional marker developed in this study can be used for aromatic rice breeding for improvement with respect to certain characters and also for the genetic resource conservation program.

Part II: Genetic diversity of aromatic rice varieties

Results

Genotyping

(a) Genetic distance between accessions and population structure

A total of 147 varieties from Myanmar and 80 varieties from the reference set were genotyped with 19 markers. The rate of missing data was 2.4%. The percentage of heterozygosity was 1.4%. The PIC of each marker in the different sets is listed in Table 6.

The number of alleles per marker and the PIC were generally slightly lower in the Myanmar collection than in the reference set. Focusing on group 5 accessions, differences between the two sets were also observed from marker to marker but the Myanmar collection was globally as diverse as the reference set.

We built an NJ tree and analyzed it using the accessions with known enzymatic grouping as references (Figure 7). A clear structure was revealed with one cluster of isozyme group 1 accessions, one cluster of isozyme group 2 accessions, one cluster of isozyme group 6 accessions, two clusters of isozyme group 5 accessions, and a few intermediates. The first cluster of isozyme group 5 accessions that we called 5A was mainly composed of varieties from South Asia (India, Nepal, Bhutan, and Pakistan), Iran and Madagascar, and one variety from Myanmar (Kyet Paung), and the second cluster that we called 5B was strictly composed of varieties from Myanmar. A few accessions among those said to belong to the Myanmar enzymatic group 5 did not cluster with the expected group: eight accessions in group 5 (Kaukkyi Meedon, Ngakyein Thee Shay, Ngakywe (IRGC 33468), Ngayun Taungpyan, Shwewagyi, Shewewayin, Wunkyaw, Meedon Yoyo) clustered with group 1 (as did two accessions in reference set group 5, T26 and Chote Dhan), two accessions clustered with group 2 (Thonsatoe, Yat Sauk Saba), one with group 6 (San Kar Yan

Thae), and three additional accessions (Ngakywe D25-4, Nga Kywe Gyi, Ngakywe (IRGC 33467)) were intermediates between the groups. The Myanmar accessions for which the isozyme group was unknown, clustered massively with either group 1 (61 accessions) or group 5B (32 accessions); only two of them clustered with group 5A (Namathalay and Thee Htat Pin Khine), and one was located with the intermediates (Paw San Hwme (MSB 2923).

Based on Structure results and Evanno's criteria (Figure 8), the most likely number of sub-populations was five. We analyzed the results again based on the isozyme-based organization of the reference set. The five sub-populations detected corresponded to group 1, group 2, group 5A, group 5B and group 6 (Figure 9). The projection of Structure assignation for K=5 on the NJ tree was in excellent agreement with the clusters identified, including for those genotypes from enzymatic group 5 which had appeared in an unexpected NJ tree cluster. Seventeen varieties (7.5%) were admixed (Figure 10).

(b) Differentiation between sub-populations

Hierarchical analysis of molecular variance revealed a highly significant genetic differentiation among the five sub-populations with 45% of the variance due to differences among sub-populations and 55% due to difference within sub-populations. The pair-wise F_{ST} varied between 0.31 and 0.63, indicating high differentiation between sub-populations. Sub-population 5B showed a high and similar level of differentiation from group 1 (FST of 0.52), group 6 (0.50) and from the 5A sub-population (0.51) and an even higher level from group 2 (0.63). Sub-population 5A had moderate but similar levels of differentiation as group 1 (FST of 0.36) group 2 (0.34) and group 6 (0.31). These high F_{ST} were consistent with the strong structure revealed by the NJ tree.

(c) Distinction between two group 5 sub-populations

The distinction between two sub-populations among accessions belonging to isozyme group 5 was unexpected. We used classification and regression tree methods (Breimann *et al*., 1984) to determine which markers best distinguished sub-population 5A from 5B. RM11, located on chromosome 7, enabled to *a posteriori* correctly predict 100% of the *a priori* grouping, with allele 140 present only in sub-population 5B and alleles 142, 144 and 148 present only in sub-population 5A (Figure 11). RM510 located on chromosome 6, had the same capacity with alleles 141 and 143, which were present only in sub-population 5B, and with alleles 131 and 139, which were present only in sub-population 5A (Figure 12). The next step was not as decisive, and three markers (RM237, RM 215 and RM338) located on chromosomes 1, 9 and 3 respectively were needed to predict 98% of the grouping (Figure 13).

(d)Aromatic /non-aromatic nature of the accessions and mutations in *Os2AP*

The aromatic/non-aromatic nature of most varieties from Myanmar (Appendix Table 1) was determined. A total of 64 of the 147 accessions (44%) were aromatic: 26 accessions from sub-population 1 (39%), two accessions from population 5A (50%), 33 accessions from population 5B (56%), and three admixed accessions (23%).

Two mutations on *Os2AP* gene (the 8 bp deletion in exon 7 and 3 bp insertion in exon 13) known to be associated with the presence of aroma were genotyped. We tested the 147 Myanmar accessions and the 31 accessions from group 5 of the reference set (Appendix Tables 1 and 2). The result showed that all the accessions that had either the 8 bp deletion or the 3 bp insertion were aromatic. However, thirty four aromatic varieties had neither of the mutations.

It should be noted that different accessions bearing the same name or root name, which generally indicates popular varieties, sometimes corresponded to distinct varietal types, differing in terms of aroma, functional mutation, or even assignment to a sub-population (e.g. Paw San Hmwe or Ngakywe)

Sequencing

To explain the presence of accessions that were aromatic even though they did not have any of the known mutations, we sequenced the *Os2AP* gene. Despite using high-fidelity long read DNA polymerase and specific PCR conditions, we were not able to consistently amplify the first two exons, probably because of the very high CG content of this zone. We amplified 4842 bp from intron 2 to 3' UTR, corresponding to 65% of the gene sequence and 78% of the coding sequence.

Among the 91 sequenced accessions listed in Appendix Table 1 and 2, we identified 63 polymorphisms corresponding to 54 SNPs, eight indels, and one SSR. Among the 63 polymorphisms, 26 were singletons almost exclusively found in two varieties: Kywet Thwar (9 specific polymorphisms), and Firooz (13 specific polymorphisms) with one in Thee Htat Pin Khine and two in Bu Saba. Nipponbare itself carried the minor allele at two SNPs, one of which was a singleton. Among the polymorphisms detected, very few were in the coding sequence: three SNPs adjacent to one 8 bp indel in exon 7, one SNP in exon 9 (singleton in a non-aromatic accession), one SNP in exon 10 (singleton in an aromatic accession), and one 3 bp indel in exon 13. One SNP, one 4 bp indel (singleton in a non-aromatic accession) and one 43 bp indel were located in exon 15 in the 3'UTR region based on the MSU annotation.

Haplotype network

After eliminating the singletons and the SSR, we obtained 36 polymorphisms representing 15 haplotypes (Table 4). Six haplotypes were variety specific. Two corresponded to Kywet Thwar (H10) and Firooz (H11). The other four were Bay Kyar (H12), which appeared to have recombined at the end of the sequence, and Nga Cheik (H13), Bu Saba (H14) and Nipponbare (H15). The remaining nine haplotypes were used to draw the haplotype network (Figures 14). Two groups of haplotypes were detected (H1 to H5 and H6 to H9). On the basis of the SNP data that we had in common with Kovach *et al*. (2009), theses two groups relate to the Japonica and the

Indica types, respectively. In our result, H1 strictly included only aromatic accessions that carried the well known 3 SNPS -8 bp deletion polymorphism. This group composed with many accessions in sub-population 5A, but including some accessions from group 1. H5 was also strictly composed of aromatic varieties with the 3 bp insertion which appeared to be specific to Myanmar accessions. The other haplotypes in the first group, which did not carry mutations in the coding sequence, were composed of a mixture of aromatic and non-aromatic accessions. In the second group, mutations in the coding sequence were found only at the very end of the last exon and did not correlate with the presence of fragrance (composite haplotypes). Few haplotypes were specific to a structure group; H8 was specific to 5A accessions (but included only two accessions), and H5 was strictly specific to 5B accessions.

A total of 24 accessions from Myanmar were aromatic but did not carry any mutations in the coding sequence of the segments sequenced in this study. They belonged to different haplotypes and different varietal groups (12 from group 1, 9 from group 5B, and 3 admixed). In addition, the three Kiriminy accessions from Madagascar belonging to group 5A (Kiriminy 1133, Kiriminy de 4 mois, and Kiriminy type Bengaly) did not carry any mutation.

Table 3 Characteristics of the SSR loci including their repeat motif, the number of alleles per locus, and PIC value in the two rice collections (Myanmar and reference set)

Marker	Chr.	SSR motif	Reference set accessions[1]		Myanmar collection[2]		Reference set group 5[3]		Myanmar group 5[4]	
			No of alleles	PIC	No of alleles	PIC	No of alleles	PIC	No of alleles	PIC
RM1	1	(GA)26	13	0.82	16	0.73	5	0.50	9	0.48
RM5	1	(GA)14	11	0.85	9	0.82	8	0.72	9	0.83
RM11	7	(GA)17	10	0.78	9	0.67	4	0.36	7	0.37
RM25	8	(GA)18	11	0.69	10	0.75	6	0.45	6	0.64
RM44	8	(GA)16	11	0.72	5	0.67	3	0.42	5	0.53
RM124	4	(TC)10	4	0.56	3	0.42	4	0.60	3	0.30
RM154	2	(GA)21	15	0.89	12	0.85	9	0.74	10	0.82
RM215	9	(CT)16	7	0.66	7	0.52	4	0.50	6	0.45
RM237	1	(CT)18	8	0.79	11	0.60	5	0.60	5	0.32
RM271	10	(GA)15	7	0.69	8	0.53	3	0.37	7	0.33
RM287	11	(GA)21	11	0.61	12	0.70	4	0.23	9	0.60
RM316	9	Complex	9	0.73	7	0.30	7	0.58	6	0.22
RM338	3	(CTT)6	3	0.40	2	0.24	2	0.17	2	0.36
RM431	1	(AG)16	10	0.75	9	0.67	5	0.47	7	0.62
RM447	8	(CTT)8	8	0.65	4	0.48	4	0.49	4	0.39
RM474	10	(AT)13	14	0.87	12	0.71	6	0.67	7	0.55
RM510	6	(GA)15	7	0.58	6	0.47	3	0.30	4	0.39
RM538	5	(GA)14	11	0.69	11	0.68	6	0.48	6	0.41
RM1227	12	(AG)15	8	0.77	6	0.72	7	0.74	6	0.67
Total			178		159		95		118	
Average			9.3	0.71	8.4	0.61	5.0	0.49	6.2	0.49

[1] 80 accessions; [2] 147 accessions; [3] 31 accessions; [4] 51 accessions

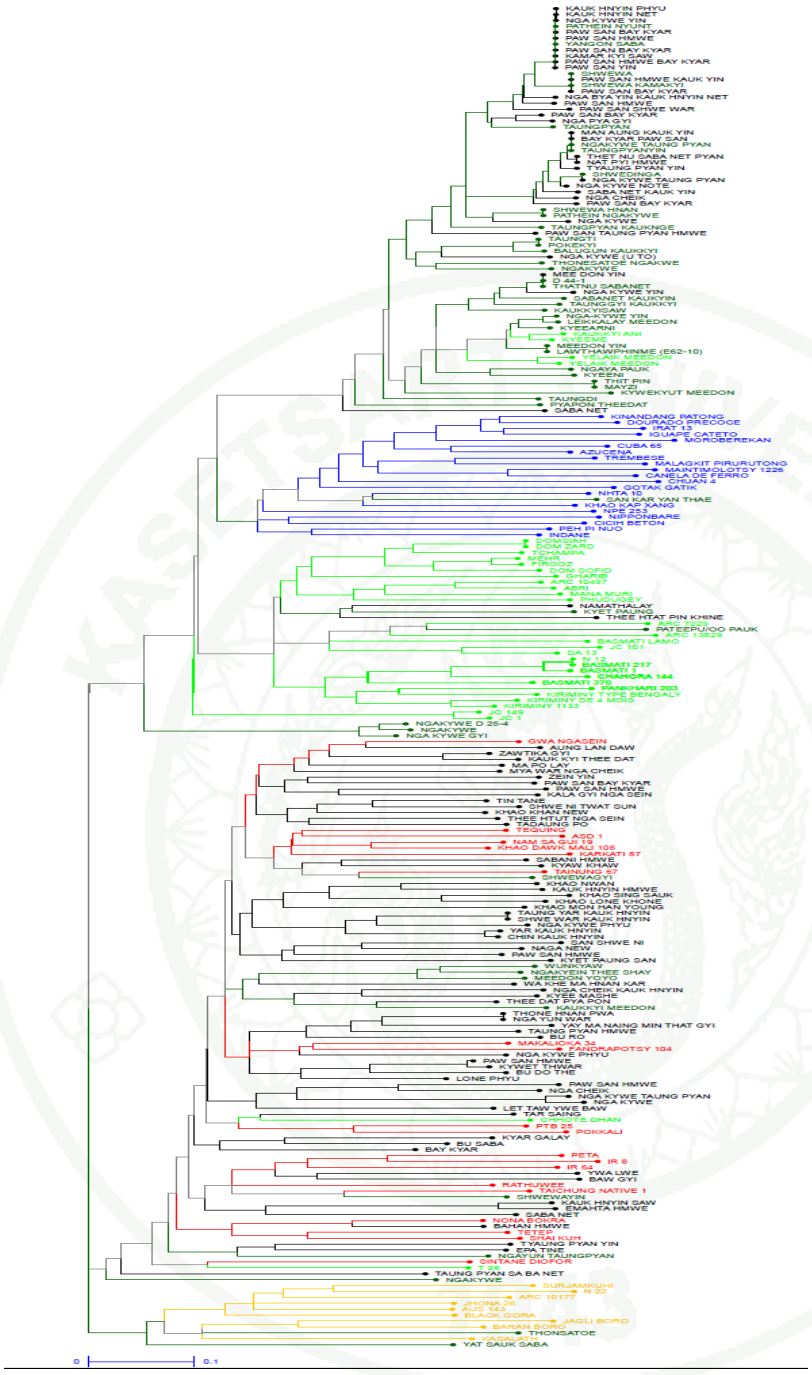

Figure 7 NJ tree representing the relative position of Myanmar accessions and the reference set.

Colors according to enzymatic groups: 1 = indica accessions in red, 2=aus/boro in yellow, 5=sadri/basmati in light green (reference set) or dark green (Myanmar accessions based on data from Khush *et al*. (2003), 6=japonica in blue; unknown group in black.

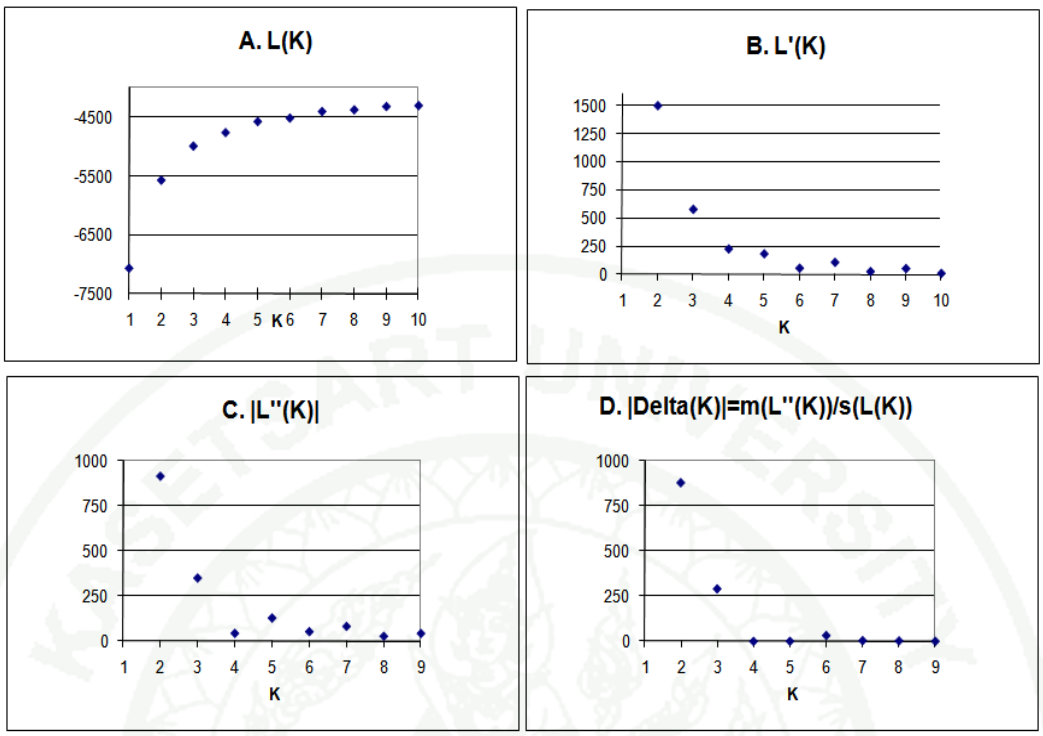

Figure 8 Evolution of the criteria that enabled detection of the likely number of subpopulations (K) in the 227 accessions using 19 SSR markers for K values varying from 1 to 10. A: L(k) = mean of the likelihood distribution LnP(D) over 10 runs for each K value.B: L'(K) = rate of change of the likelihood function with respect to K.
C:|L''(K)| = second order rate of change of L(K) with respect to K; Delta(k) = mean |L''(K)|/standard deviation (L(K).

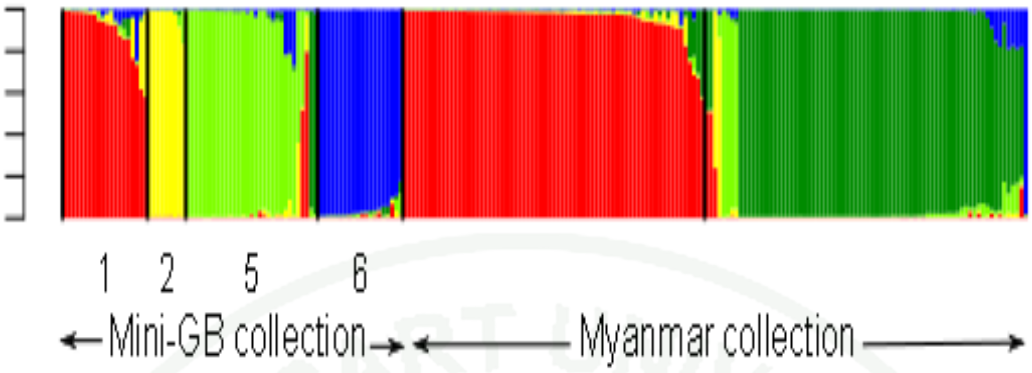

Figure 9 Graph of estimated membership fraction for K=5 (run with the highest likelihood). The first part of the graph represents the mini-GB collection with a vertical bar separating enzymatic groups. The second part of the graph includes the Myanmar accessions. The five different colors correspond to the five different groups of Structure. Enzymatic groups: 1=indica, 2=aus/boro, 5=sadri/basmati; 6=japonica (Glaszmann, 1987).

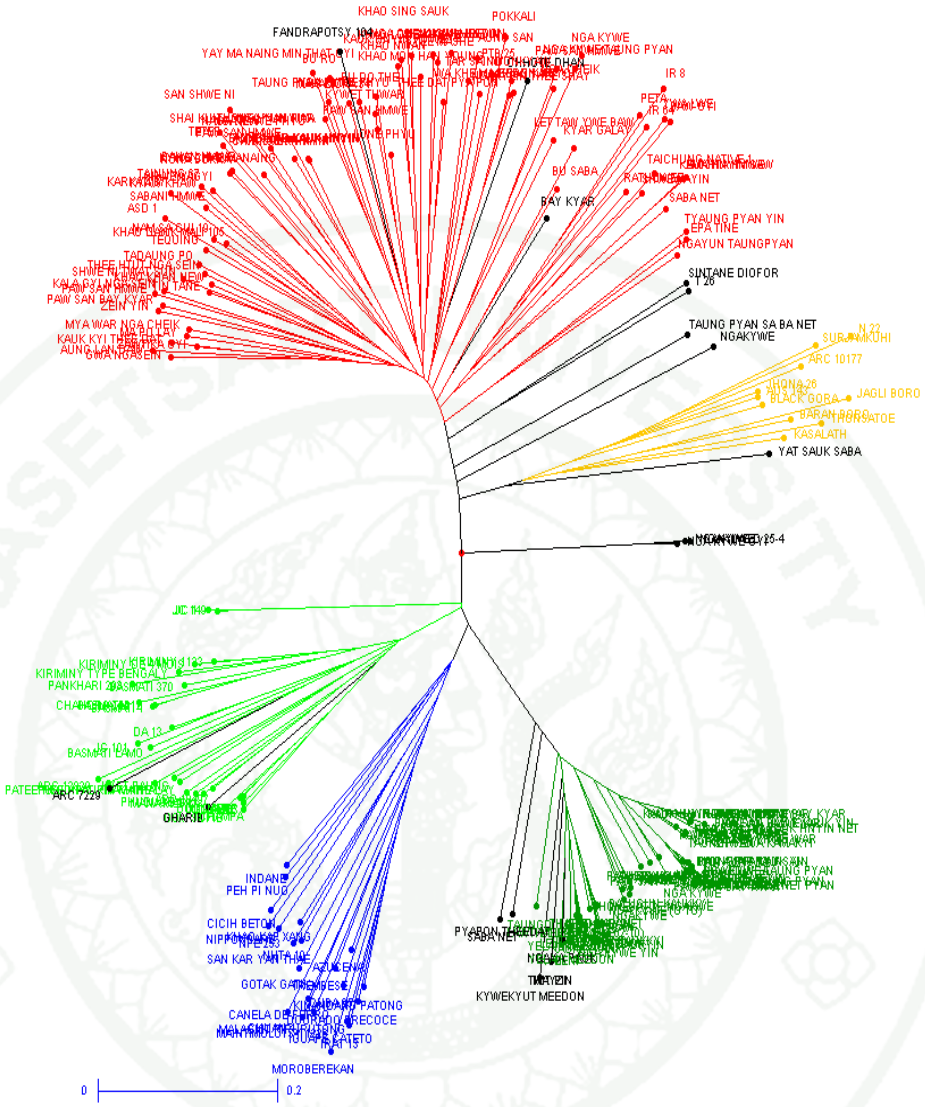

Figure 10 Organisation according to Structure groups.

colors on the basis of isozyme groups; red= Indica (group 1),

yellow = group 2, palegreen = group 5A (Basmati),

dark green = group 5B (Myanmar) , blue = Japonica

black = admixed

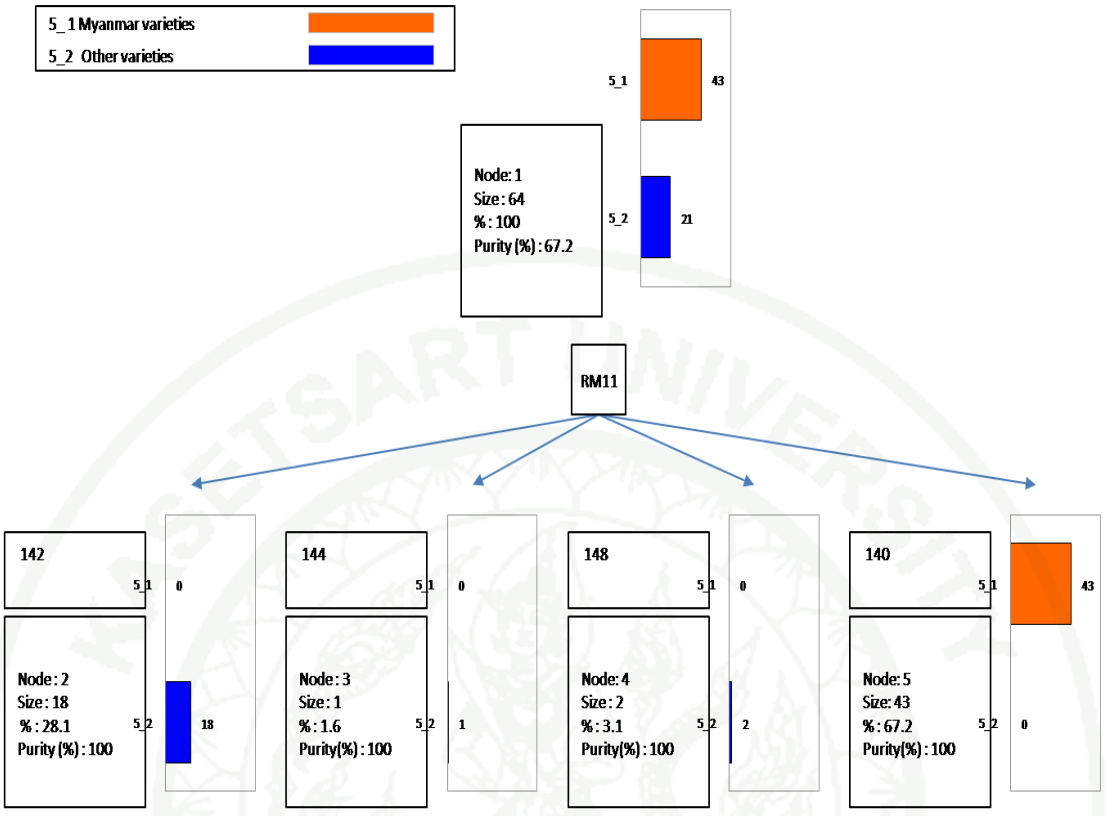

Figure 11 Classification and regression tree showing the split of group5 into two sub-populations distinguished by RM 11.

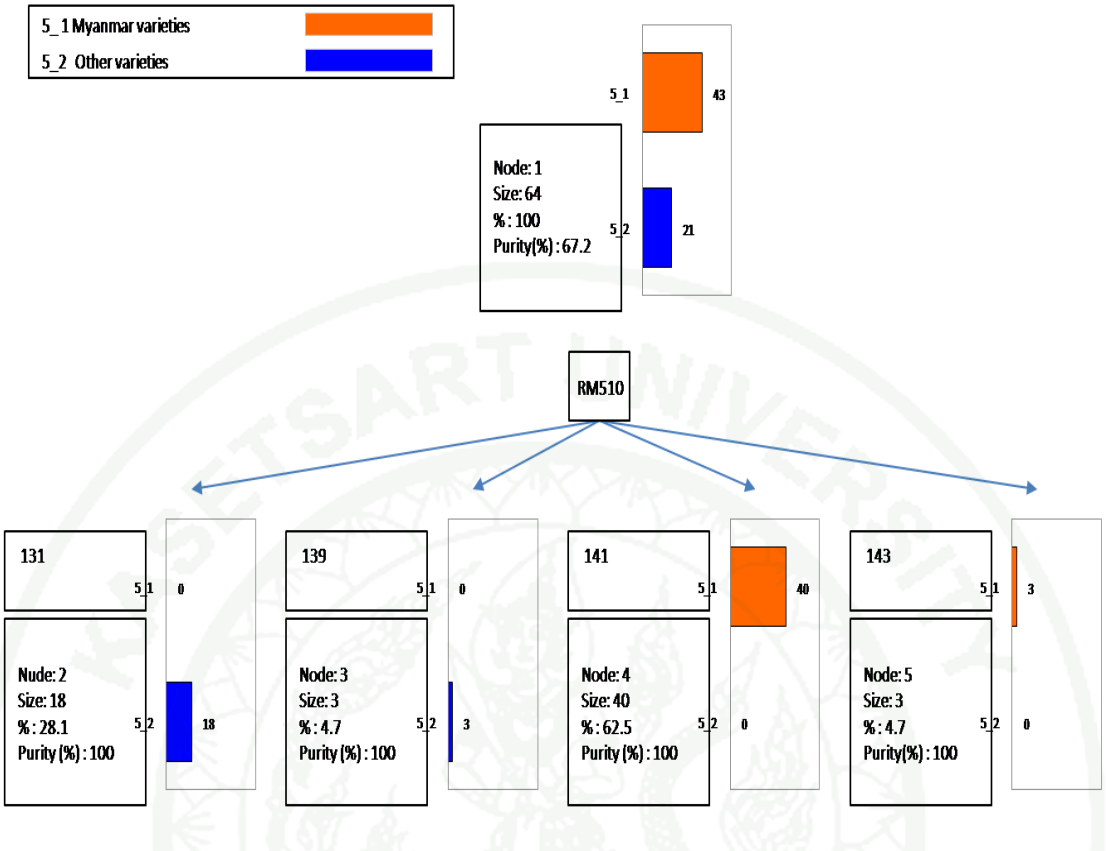

Figure 12 Classification and regression tree showing the split of group5 into two sub-populations distinguished by RM 510.

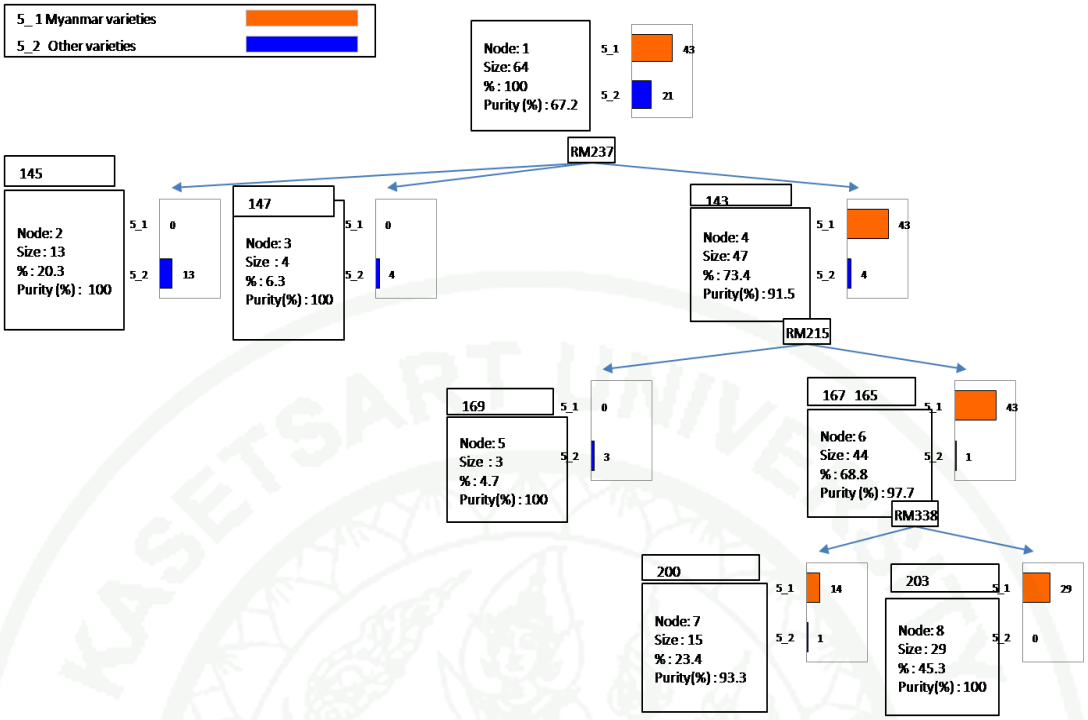

Figure 13 Classification and regression tree showing the split of group5 into two sub-populations distinguished by RM 237, RM 215, RM 338.

Table 4 Position of the polymorphisms and haplotypes in the 91 accessions

Pos.	Zone	P	Nature	Freq.	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15
695	I2	snp	nc	2.2	C	C	C	C	C	C	C	C	C	C	C	G	G	C	C
825	I2	snp	nc	7.7	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C
829	I2	snp	nc	3.3	T	T	T	T	T	T	T	T	T	T	T	C	C	C	T
839	I2	indel	nc	29.7	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2
876	I2	snp	nc	2.2	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A
915	I2	snp	nc	25.3	C	C	C	C	C	T	T	T	T	T	T	C	C	C	C
918	I2	snp	nc	2.2	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T
1082	I2	snp	nc	21.9	T	T	T	T	T	C	C	C	T	T	T	T	T	T	T
1112	I2	snp	nc	25.3	A	A	A	A	A	T	T	T	T	T	T	A	A	A	A
1113	I2	indel	nc	25.3	3	3	3	3	3	0	0	0	0	0	0	3	3	3	3
1316	I2	snp	nc	7.7	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G
1340	I2	snp	nc	23.1	C	C	C	C	C	G	G	G	G	C	C	C	C	C	C
1819	I4	snp	nc	25.3	T	T	T	T	T	A	A	A	A	A	A	T	T	T	T
2142	I4	snp	nc	26.4	T	T	T	T	T	A	A	A	A	A	A	T	T	T	A
2144	I4	snp	nc	8.8	T	T	T	T	T	A	T	T	T	T	A	T	T	T	T
2205	I4	snp	nc	26.4	A	A	A	A	A	G	G	G	G	G	G	G	A	A	A
2219	I4	snp	nc	25.3	G	G	G	G	G	A	A	A	A	A	A	G	G	G	G
2889	E7	snp	c	25.3	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A
2891	E7	snp	c	25.3	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A

Table 4 (Continued)

Pos.	Zone	P	Nature	Freq.	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15
2893	E7	snp	c	24.2	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A
2894	E7	indel	c	25.3	0	8	8	8	8	8	8	8	8	8	8	8	8	8	8
3003	I7	indel	nc	7.7	T	T	T	T	T	0	T	T	T	T	T	T	T	T	T
3031	I7	snp	nc	3.3	G	G	G	G	G	G	G	G	G	G	G	T	T	T	G
3219	I8	snp	nc	12.1	C	C	C	T	C	C	C	C	C	C	C	T	T	C	C
3220	I8	snp	nc	25.3	G	G	G	G	G	A	A	A	A	A	A	G	G	G	G
3372	I8	snp	nc	25.3	C	C	C	C	C	G	G	G	G	G	G	C	C	C	C
3467	I8	snp	nc	25.3	T	T	T	T	T	C	C	C	C	C	C	T	T	T	T
3988	I9	snp	nc	2.2	T	T	T	T	T	T	T	T	T	G	G	T	T	T	T
4423	I10	snp	nc	2.2	C	C	C	C	C	C	C	C	C	G	G	C	C	C	C
4452	I10	snp	nc	2.2	A	A	A	A	A	A	A	A	A	A	A	G	G	A	A
5233	E13	indel	c	14.3	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
5491	I13	snp	nc	2.2	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A
5531	I13	snp	nc	29.7	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A
5688	I14	snp	nc	4.4	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
6049	3'UTR	snp	nc	3.3	T	T	T	T	T	T	T	T	T	A	T	T	A	A	T
6086	3'UTR	indel	nc	24.2	43	43	43	43	43	0	0	0	0	43	43	0	43	43	43

Pos. = position of the polymorphism (position 1 for the A of the ATG of the first exon corresponding to position 20,377,257 in MSU v6).
Indels of more than one bp: site 839: 2 = TT; site 1113: 3 = TTA; site 2894: 8= ATTATGGC; site 5233: 3=TTA;
site 6086: 43= GTTCTCTCCGTATCGGCTTGTGGTGTTTCAACCTTAAGACC
Zone: I = intron; E = exon Nature: nc= non-coding; c= coding region

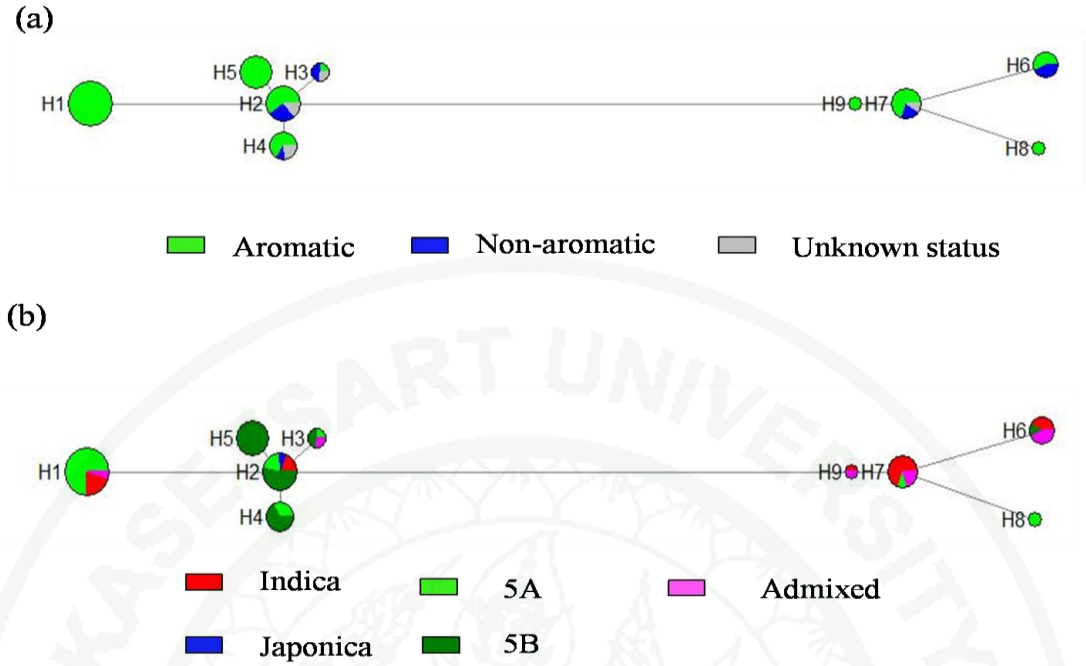

Figure 14 Haplotype network for the gene Os2AP Each node represents a haplotype, its size being proportional to its frequency.

(a) Distribution according to the aromatic nature of the accessions
Green: aromatic; blue: non-aromatic; grey: unknown

(b) Distribution according to Structure groups
Red: 1 (indica); blue: 6 (japonica); light green=5A; dark green = 5B; violet: admixed

Discussion

The genetic diversity of 147 accessions from Myanmar were analyzed at neutral markers as well as the *Os2AP* gene, which is responsible for aroma. Based on isozymes, Khush *et al*. (2003) classified the 1354 accessions from Myanmar stored in the IRRI genebank and found 1115 (82.4%) belonging to group 1 (indica), 17 (1.3%) to group 2, 52 (3.8%) to group 5, 30 (2.2%) to group 6 (japonica) and 140 (10.3%) to group 0 (intermediates). In comparison with this representation in the reference world collection, the set in this study clearly over-represented group 5 accessions and under-represented japonica and intermediates, in accordance with our focus on aromatic rice.

The analyses in this study led to unexpected and original results, both in terms of varietal groups and distribution of aromatic varieties. The results of Structure and of the distance-based NJ tree clearly separated group 5 accessions into two sub-groups. These two sub-groups carry diagnostic alleles on at least two loci located on different chromosomes, eliminating the possibility that this could result from miscoding. The accessions in group 5A originate from a range of countries along the Himalayan foothills, while the accessions in group 5B appear to be limited to Myanmar. The fact that this group had not been identified previously is probably due to the small number of Myanmar accessions in earlier studies. This result shows the interest of broadening the samples analyzed with molecular markers to obtain a finer view of the genetic structure of *O. sativa*. The broadest systematic study to date remains the initial isozyme survey reported in 1987-1988. While proposing group 5 as a new group comprising varieties previously considered to be Indicas, the author of that study commented on the loose coherence of this group with a geographic cline placing Myanmar varieties at the eastern edge, with a near-specific allele at locus *Amp3*; yet this was the only locus with this pattern, providing insufficient grounds for distinguishing sub-groups on the basis of multi-locus linkage disequilibrium.

The position of *Amp3* has not yet been clearly defined on the rice genome; based on recombination mapping, it should be located between 5.5 and 7.5 Mb on

chromosome 6, but a clear association with an annotated gene in the region remains to be established. RM510, whose polymorphism clearly separates 5A from 5B, is located on chromosome 6, in position 2.8 Mb, i.e. approximately 3 to 5 Mb from *Amp3*; linkage disequilibrium between two loci (*Amp3* and *Est2*) on chromosome 6 was highlighted long ago, as was its relation to varietal classification (Glaszmann 1988).

Among the Myanmar accessions, a very high proportion (44%) is aromatic. This high proportion was expected because our sampling focused on aromatic accessions. The results showed that the very high proportion of aromatic accession were in group 1. The proportion (39%) is almost as high as in groups 5A and 5B (50 and 56%). Low yield is the main concern in aromatic varieties belonging to group 5 (Singh *et al*., 2000), while rice top yielders in tropical irrigated conditions are generally indica varieties. The development of new aromatic varieties with an indica background is difficult using varieties from group 5 as donors because of inter-group compatibility problems. The identification of many new aromatic varieties belonging to group 1 may considerably broaden the range of varieties that breeders can use.

Sequencing on part of the *Os2AP* gene in most of the aromatic accessions was also performed in this study. The diversity of the gene has already been studied in depth by Kovach *et al*. (2009), who identified 10 aroma-associated alleles. Among the polymorphisms affecting the coding sequence in our sample, we found two mutations, the 3 SNPs/8 bp deletion (badh2.1) and 3 bp insertion (badh2.8) already identified by these authors. But the 3 bp insertion (badh2.8) was present at a much higher frequency in our sample. It was recently demonstrated that this insertion did not modify the gene expression but the addition of a tyrosine to the peptide interfered with cofactor NAD^+ binding, lowered enzyme activity and led to accumulation of 2-acetyl-1-pyrroline, although at a lower level than in accessions carrying badh2.8 (Vanavichit and Yoshihashi 2010, Wongpanya *et al*., 2011, Myint *et al*., 2012). We found two additional deletions at the very end of the gene that were specific to Myanmar accessions, a deletion of 3 bp in three accessions and another deletion of 43 bp in 22 accessions. However, it is unclear whether these mutations are in the 3'UTR or outside the gene. Neither of these two mutations controls the production of aroma since both groups include both aromatic and non-aromatic accessions.

The badh2.8 allele is typical of Myanmar aromatic accessions belonging to group 5B. It is present in the most popular and widely grown aromatic varieties such as Paw San Hmwe and its declinations (Paw San Yin, Paw San Bay Kyar). The same situation was encountered among Chinese accessions with a specific allele that has not yet been found elsewhere (Shi *et al.*, 2008). This suggests that the trait was selected on several occasions. While the 8 bp deletion allele was quite successful, since it is encountered in a large diversity of countries from Madagascar to China and in various genetic backgrounds (isozyme group 5A mostly but also groups 1 and 6), the badh2.8 allele appears to have remained more localized, i.e. limited to Myanmar and group 5B. This may be due to limited past varietal exchanges between Myanmar and neighboring countries but this hypothesis is unlikely since the badh2.1 allele was found in some of the indica aromatic accessions from Myanmar. It may result from the presence of this allele in less interesting backgrounds than those of the successful badh2.1 allele, or from a lower expression of the aroma due to the different nature of the mutations (amino acid addition in badh2.8 versus stop codon in badh2.1).

Several aromatic accessions did not carry any mutation in the coding segments we sequenced. The reality of their aromatic nature can thus be questioned. Simple and fast methods such as the sniffing method or the one proposed by Sood and Siddiq (1978) are commonly used and have been shown to be in very good agreement with gas chromatography (Lorieux *et al.*, 1996) but false positives and false negatives (because of the transient nature of aroma expression) are known to occur. In addition, the link between aroma and 2-acetyl-1-pyrroline is not always direct. Some varieties can be aromatic despite a low level of 2-acetyl-1-pyrroline, probably because of higher levels of other compounds (Sakthivel *et al*. 2009). The Kiriminy varieties from Madagascar (group 5A) that do not carry the badh2.1 mutation are nevertheless said to be aromatic (N. Ahmadi, Cirad, personal communication) but whether the aroma comes from 2-acetyl-1-pyrroline or from another compound has not yet been determined for this variety.

Assuming that the great majority of those accessions are indeed aromatic, the mutations are located either in the unsequenced coding part or in the promoter. However, only two rare mutations (singletons), both specific to tropical japonicas,

were found by Kovach *et al*. (2009) in the first and second exons that we did not sequence. Shi *et al*. (2008) found a 7 bp deletion in exon 2 but the deletion appeared to be specific to Chinese accessions. The possibility that the functional mutation differs from those already described cannot be ignored. A few accessions that did not have functional *Os2AP* mutation that could explain their aromatic nature were also detected in previous studies (Kovach *et al*. 2009; Sakthivel *et al*. 2009). Fitzgerald *et al*. (2008) postulated that production of 2-acetyl-1- pyroline could be driven by alleles at two different genes and not only by different alleles of *Os2AP*. Rice has indeed a second *BADH1* gene located on chromosome 4 that codes for the *BADH* enzyme and acts in a similar way to *BADH2* or *Os2AP* but is regulated differently. A minor QTL for aroma l was detected by Lorieux *et al*. (2000) on the same region of chromosome 4. Singh *et al*. (2010) found a haplotype of *BADH1* associated with aroma. However, they also showed that the mutation in *BADH2* whose expression is constitutive) was a primary requirement and that *BADH1* mostly appeared to modify aroma intensity. But the acetyl-pyrroline biosynthesis pathway is complex and mutations in other genes than *BADH* paralogues may be responsible for the aroma, as suggested by Sakthivel *et al*. (2009). The development of mapping populations involving accessions that do not carry known mutations in BADH2 or Os2AP as a parent could help solve this issue.

CONCLUSION

The present study shows remarkable features for both allelic diversity in aroma trait and genetic diversity of Myanmar aromatic rice varieties. Myanmar aromatic accessions are composed in group 1 (indica), group 5A and group 5B. The varieties belonging in group 1 and group 5A carry the well known 3 SNPS -8 bp deletion polymorphism. The 3 bp insertion allele is found as major allele in Myanmar aromatic varieties which is strictly composed in the new varietal group 5B. The origin of this aroma allele is assumed as in Delta region of Myanmar. We also found two additional deletions at the very end of the *Os2AP* gene that were specific to Myanmar accessions, a deletion of 3 bp in three accessions and another deletion of 43 bp in 22 accessions. However, it is unclear whether these mutations are in the 3'UTR or outside the gene and whether these two mutations control the production of aroma in Myanmar aromatic varieties.

In contrast to 8bp-deleted aromatic lines, the *Os2AP* gene in the aromatic rice varieties with 3bp insertion is expressed normally. It is noted that the aromatic lines with 3bp insertion have lower amount of 2AP compared to those with 8bp deleted lines. The 3bp insertion is in frame with translation and adds an additional amino acid (Tyrosine) into the peptide. Based on the report on Wongpanya *et al*., 2011, the predicted three dimensional structure revealed that the additional tyrosine is perfectly situated in the NAD binding pocket of the NAD binding domain. The Os2AP enzyme activity in this type of mutation shows lower enzyme activity than in the wild type, which causes the accumulation of 2AP in these varieties.

Knowledge of the most likely genetic cause of aroma has allowed the development a perfect assay for aroma in rice. A functional marker for detecting the new allele in Myanmar aromatic varieties could be developed in this study and tested in an aroma-segregating F_2 population and screening in diverse types of rice varieties. This marker allows determination of the genotypic status of an individual rice plant, either homozygous aroma, homozygous non-aroma or heterozygous non-aroma, has practical utility for rice breeders. The assay is a

simple robust method for screening rice to determine its aroma status across a wide range of rice varieties and within segregating populations using DNA isolated from rice following simple, inexpensive and rapid extraction protocols. The PCR products can be analysed easily and inexpensively on polyacrylamide gel or alternatively using more sophisticated high throughput equipment, making the assay a very versatile tool.

In summary, with one new group and the new alleles found in Myanmar, our study illustrates that the Himalayan foothills contain series of non-Indica and non-Japonica varietal types that bear novel variations for useful traits. These new alleles can be of interest for both local and more general plant breeding programs.

LITERATURE CITED

Agarwal, R. K., V. V. Shenoy, J. Ramadevi, R. Rajkumar and L. Singh. 2002. Molecular characterization of some Indian Basmati and other elite rice genotypes using fluorescent – AFLP. **Theor. Appl. Genet.** 105: 680–690.

Ahn, S. N., C.N. Bollich and S.D. Tanksley. 1992. RFLP Tagging of a Gene for Aroma in Rice. **Theor. Appl. Genet.** 84**:** 825-828.

Amarawathi, Y., R. Singh, A.K. Singh, V.P. Singh, T. Mohapatra and T.R. Sharma. 2008. Mapping of quantitative trait loci for basmati quality traits in rice (Oryza sativa L.). **Mol. Breed.** 21: 49–65.

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

Baishya, S., Sachdev Archana, R.P. Johari and S. L. Mehta. 2000. RAPD analysis of aromatic and non aromatic rice (Oryza sativa L). **J. Plant Biochem. Biotech.** 9: 23–6.

Bandelt, H.J, P. Forster and A. Rohl. 1999. Median-joining networks for inferring intraspecic phylogenies. **Mol. Biol. Evol.** 16: 37–48.

Berner, D. K. and B.J Hoff. 1999. Inheritance of scent in American long grain rice. **Crop. Sci**. 986 26: 876–878.

Bhattacharjee, P., R.S. Singhal and P.R Kulkarni. 2002. Basmati rice: a review. **J. Food Sci.Tech.** 37: 1-12

Bollich, C.N., J.N. Rutger, B.D. and Webb. 1992. Developments in rice research in the United States. **International Rice Commission Newsletter.** 41: 32-34

Bonow, S., V. R. Edila, P. Von, M. G. C. Vieira, and B. Vosman. 2009. Microsatellite markers in and around rice genes: applications in variety identification and DUS testing. **Crop. Sci**. **49**: 880–886.

Bourgis, F., R. Guyot, H. Gherbi, Tailliez, I .Amabile, J. Salse, M. Lorieux, Delseny and A. Ghesquière. 2008. Characterization of the major fragrance gene from an aromatic japonica rice and analysis of its diversity in Asian cultivated rice. **Theor. Appl. Genet**. 117: 353-368.

Bradbury, L.M.T, T.L. Fitzgerald, R.J. Henry, Q. Jin and D.L.E Waters. 2005. The gene for fragrance in rice. **Plant. Biotech J** 3: 363–370.

Bradbury, L.M.T, S.A. Gillies, D.J. Brushett, D.L.E. Waters and R.J. Henry. 2008. Inactivation of an aminoaldehyde dehydrogenase is responsible for fragrance in rice. **Plant. Mol. Biol.** 68: 439–449.

Breiman, L, J.H. Friedman, R. Olshen and C.J. Stone. 1984. **Classification and regression trees**. Wadsworth and Brooks, Pacific Grove, California.

Buttery, R.G., L.C. Ling and B.O. Juliano. 1982. 2-Acetyl-1-pyrroline: an important aroma component of cooked rice. **Chem Ind (London)**: 958-959.

Buttery, R.G., B.O. Juliano and L.C. Ling. 1983b. Identification of Rice Aroma Compound 2- Acetyl-1-Pyrroline in Pandan Leaves. **Chemistry & Industry:** 478-478.

Buttery, R.G. and L.C. Ling. 1995. Volatile Flavor Components of Corn Tortillas and Related Products. **J. Agri. Food Chem**. 43: 1878-1882.

Chang, T.T. 1976. The origin, evolution, cultivation, dissemination, and diversification of Asian and African ricies. **Euphytica** 25: 425-441.

Chen, S., Y. Yang, W. Shi, Q .Ji, F. He, Z. Zhang, Z.Cheng, X. Liu and M. Xu. 2008. *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. **Plant Cell.** 20: 1850–1861.

Cordeiro, G.M, M.J. Christopher, R.J. Henry and R.F. Reinke. 2002. Identification of microsatellite markers for fragrance in rice by analysis of the rice genome sequence. **Mol. Breed.** 9: 245-250

Dhulappanavar, C.V. 1976. Inheritance of scent in rice. **Euphytica** 25: 659–662.

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.

Excoffier, L., G .Laval and S. Schneider. 2006. Arlequin v3.1. **An integrated software for population genetic data analysis**. Available source: http://cmpg.unibe.ch/software/arlequin3

Fitzgerald, M.A., N.R. Sackville Hamilton, M.N. Calingacion, Verhoeven, H.A. and V.M. Butardo. 2008. Is there a second fragrance gene in rice? **Plant Biotechn. J** 6: 416-423.

Garland, S., L. Lewin, A. Blakeney, R. Reinke and R. Henry. 2000. PCR-based molecular markers for the fragrance gene in rice (Oryza sativa. L.). **Theor. Appl. Genet**. 101: 364-371.

Garris, A.J. T.H. Tai, J .Coburn, S. Kresovich and S. McCouch. 2005. Genetic structure and diversity of *O. sativa*. **Genetics** 169: 1631-1638.

Gay, F., I. Maraval , S. Roques, Z. Gunata, R. Boulanger, A. Audebert and C. Mestres. 2010. Effect of salinity on yield and 2-acetyl-1-pyrroline content in the grains of three fragrant rice cultivars in Camargue. **Field Crop Research** 117: 154-160.

Ghose, R.L.M. and W.T. Butany. 1952. Studies on the inheritance of some Characters in rice (*Oryza sativa* L.). **Indian J. Genet. Plant Breed.** 12: 26-30.

Giarrocco, L. E., M. A. Marassi and G. L. Salerno. 2007. Assessment of the genetic diversity in argentine rice cultivars with SSR markers. **Crop Sci.** 47: 853–858.

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

Glaszmann, J.C. 1988. Geographic pattern of variation among Asian native rice cultivars based on fifteen isozyme loci. **Genome** 30: 782-792.

Glaszmann, J.C and M. Arraudeau. 1986. Rice plant type variation: Japonica-Javanica relationships. **Rice Genetic Newsl** 3: 41-43.

Glaszmann, J.C, T. Mew, H. Hibino, C.K. Kim, T.I. Mew, C.H. Vera Cruz, J.L. Notteghem and J.M. Bonman. 1995. **Molecular variation as a diverse source of disease resistance in cultivated rice.** In: Rice Genetics III, IRRI, Los Baños, Philippines: 460-466.

Ghose, R.L.M and W.T. Butany. 1952. Studies on the inheritance of some characters in rice (*Oryza sativa* L.) **Indian J. Genet. Plant Breed.** 12: 26–30.

Hien, N.L., T. Yoshihashi, W.A. Sarhadi, and Y.Hirata . 2006a. Sensory test for aroma and quantitative analysis of 2-acetyl-1-pyrroline in Asian aromatic rice varieties. **Plant Prod. Sci.** 9: 294–297.

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

Huang, X. and A . Madan. 1999. CAP3: A DNA sequence assembly program. **Genome Res**. 9: 868–877.

Huang, T.C., C.S Teng., J.L. Chang, H.S. Chuang, Ho CT and Wu ML. 2008. Biosynthetic mechanism of 2-acetyl-1-pyrroline and its relationship with Δ1-pyrroline-5-carboxylic acid and methylglyoxal in aromatic rice (Oryza sativa L.) callus. **J Agric. Food Chem.** 56: 7399–7404.

Isshiki, M., Y. Yamamoto, H.Satoh and K. Shimamoto. 2001. Nonsense-mediated decay of mutant waxy mRNA in rice. **Plant Physiol.** 125: 1388–1395.

Jayamani, P., S. Negrão, M. Martins, B. Maçãs, and M. Oliveira. 2007. Genetic relatedness of Portuguese rice accessions from diverse origins as assessed by micro satellite markers. **Crop Sci.** 47: 879–884.

Jin, Q.S., D. Waters, G.M. Cordeiro., R.J. Henry and R.F. Reinke. 2003. A single nucleotide polymorphism (SNP) marker linked to the fragrance gene in rice (Oryza sativa L.). **Plant Science** 165: 359-364

Juliano, B.O. 2007. **Rice chemistry and quality**. Philrice, Philippines: 402.

Kato, S., H. Kosaha and S. Hara. 1928. **On the affinity of rice varieties as shown by the fertility of rice plants**. Bulletin of Sciences of Central Agricultural Kyushu Imperial University 2: 241-276.

Khush, G.S., D. Brar, P.S. Virk, S.X. Tang, S.S. Malik, G.A. Busto, Y.T. Lee, R .McNally, L.N. Trinh, Y. Jiang and M.A.M. Shat. 2003. **Classifying rice germplasm by isozyme polymorphism and origin of cultivated rice**. IRRI Discussion Paper 46: 279.

Kovach, M.J., M.N. Calingacion, M.A. Fitzgerald and S.R. McCouch. 2009. The origin and evolution of fragrance in rice. **PNAS 106**(34): 14444-14449.

Kuo, S.M., S.Y. Chou, A.Z. Wang, T.H. Tseng, F.S. Chueh and H.E. Yen. 2005. **The betaine aldehyde dehydrogenase (BAD2) gene is not responsible for the aroma trait of SA0420 rice mutant derived by sodium azide mutagenesis**. 5th International Rice Genetics Symposium. Philippines: IRRI: 166–7. Nov 19–23.

Liu, K. and S.V. Muse. 2005. PowerMarker: Integrated analysis environment for genetic marker data. **Bioinformatics** 31(9): 2128-2129.

Lorieux, M., M. Petrov, N. Huang, E. Guiderdoni and A . Ghesquière. 1996. Aroma in rice: genetic analysis of quantitative trait. **Theor. Appl. Genet.** 93: 1145-1151.

Mahatheeranont, S., S. Keawsa-ard and K. Dumri. 2001. Quantification of the rice aroma compound, 2-acetyl-1-pyrroline, in un-cooked Khao Dawk Mali 105 brown rice. **Agric. Food Chem**. 49: 773-779.

Masuda, R. 1989. Cold storage of vegetables, Vegetable soybean, **Refrigeration**. 64: 359-376 .

Matsuo, T. 1952. Genecological studies on cultivated rice. **Bull Nat Inst Agric Sci Jpn** 3: 1-111.

Masumoto, K., A. Fujita, K. Kawakami, T. Mikami, M. Nomura. 2004. Differences in aromatic components of ten brands of rice according to annual production. **Food Science and Technology Research**. 10**:** 474-478

McCouch, S. R. *et al*. 2002. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). **DNA Res**. 9: 199–207.

Murashige, T. and F. Skoog. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. **Physiol Plant** 15: 473–497.

Myint, K. M., S.Arikit, S. Wanchana, T. Yoshihashi, K. Choowongkomon and A. Vanavichit. 2012. A PCR-based marker for a locus conferring the aroma in Myanmar rice (*Oryza sativa. L.*). **Theor. Appl. Genet. (**In press).

Nadaf, A.B., S. Krishnan and K.V. Wakte. 2006. Histochemical and biochemical analysis of major aroma compound (2-acetyl-1-pyrroline) in basmati and other scented rice (*Oryza sativa* L.). **Current Science**. 91(11): 1533-1536.

Nakagahra, M. 1978. The differentiation, classification and center of genetic diversity of cultivated rice by isozyme analyses. **Trop Agr Res** Ser 11: 77–82.

Navarro, M., V. Butardo, C. Bounphanousay, R. Reano, R.S. Hamilton and H. Verhoeven. 2007. The good, the BAD and the fragrant – understanding fragrance in rice. **Int. Network on quality rices** – clearing old hurdles with new science: improving rice grain quality. Philippines: IRRI: 16. Abstract, April 17–19.

Niu, X., W. Tang, W. Huang, G. Ren, Q. Wang, D. Luo, Y. Xiao, S .Yang, F.Wang, B.R. Lu, F. Gao, T. Lu and Y. Liu. 2008. RNAi-directed down-regulation of OsBADH2 results in aroma (2-acetyl-1-pyrroline) production in rice (*Oryza sativa* L.). **BMC Plant Biol.** 8: 100.

Oka, H.I. 1958. Varietal variation and classification of cultivated rice. **Ind J genet and Plant Breeding** 18: 78-79.

Oka, H.I. 1988. **Indica-Japonica differentiation of rice cultivars**. In: Origin of cultivated rice. Oka HI (ed). Japan Scientific Society Press Tokyo: 141-179.

Perrier, X. and J.P. Jacquemoud-Collet. 2006. **DARwin software**. http://darwin.cirad.fr/

Petrov, M., M. Danzart, P. Giampaoli, J. Fayre, and H. Richard. 1996. Rice aroma analysis. Discrimination between a scented and non-scented rice. **Sci Aliments** 16: 347-360.

Philpot, K., A. Aryan and J. Oliver. 2005. Grain quality of fragrant rice: distribution of 2- acetyl-1-pyrroline indifferent tissues of paddy rice. **In The 12th Royal Australian Chemical Institute (RACI) Convention**, Sydney, AUSTRALIA.

Pritchard, J.K., M. Stephens and P. Donnelly. 2000a. Inference of population structure using multilocus genotype data. **Genetics** 155: 945–959.

Reinke, R.F., L.A. Welsch, J.E. Reece, L.G. Lewin and A.B. Blakeney. 1991. Procedure for quality selection of aromatic rice varieties. **Int Rice Res Newsl** 16: 10–1.

Risterucci, A.M., L. Grivet, J.A.K. N'Goran, I. Pieretti, M.H. Flament and C. Lanaud. 2000. A high density linkage map of *Theobroma cacao* L. **Theor. Appl. Genet**. 101: 948-955.

Romanczyk, L.J., C.A. Mcclell, L.S. Post and W.M. Aitken. 1995. Formation of 2-Acetyl- 1-Pyrroline by Several Bacillus-Cereus Strains Isolated from Cocoa Fermentation Boxes. **J. Agric. Food Chem.** 43**:** 469-475.

Roy, R., D.L. Steffens, B. Gratside, G.Y. Jang and J.A. Brumbaugh. 1996. Producing STR locus patterns from bloodstains and other forensic samples using an infrared fluorescent automated DNA sequencer. **J. Forensic Sci.** 41: 418–424.

Sakthivel, K., C.N. Neeraja, N. Shobha Rani, K. Kaladhar and R.M. Sundaram and S.M. Balachandran. 2006. **Allele mining for aroma gene BAD-2 in Indian rice cultivars**. 2nd International Rice Congress, New Delhi, Oct 9–13: 316. Abstract.

Sakthivel, K., R.M. Sundaram, N. Shobha Rani, S.M. Balachandran and C.N. Neeraja. 2009. Genetic and molecular basis of fragrance in rice. **Biotechnol Adv** 27(4): 468–473.

Shi, W., Y. Yang, S. Chen and M. Xu. 2008. Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties**. Mol. Breed.** 22: 185–192.

Shobha Rani, N., M.K. Pandey, G.S.V. Prasad and I. Sudharshan. 2006. Historical significance, grain quality features and precision breeding for improvement of export quality basmati varieties in India. **Indian J. Crop Sci** 1: 29–41.

Singh, A., P.K. Singh, R. Singh, A. Pandit, A.K. Mahato, D.K. Gupta, K. Tyagi, A.K. Singh, N.K. Singh and T.R. Sharma. 2010. SNP haplotypes of the *BADH1* gene and their association with aroma in rice (*O. sativa* L.). **Mol. Breeding** 26: 325-338.

Singh, R.K., U.S. Singh, G.S. Khush. 2000. **Aromatic rices**. Oxford and IBH Publishing Co, New Delhi, India: 292.

Siwach, P., J. Sunita, , N. Saini, V. K. Chaudhary and R. K. Jain. 2004. Allelic diversity among basmati and non-basmati long grain *Indica* rice varieties using microsattellite markers. **J. Plant Biochem. Biotech.** 13: 25–32.

Sood, B.C. and E.A. Siddiq. 1978. A rapid technique for scent determination in rice. **Indian J. Genet Plant Breed.** 38: 268–271.

Srivong, P., P. Waihalong, P. Pongdontri and P. Wangsomnuk. 2006. **Comparison of partial BADH2 transcripts of Khao Dok Mali 105 and its mutants of different aroma levels.** Poster abstract P202 in Joint third AOHUPO and forth structural biology and functional genomics conference, December 4-7, 2006, University cultural centre, National University of Singapore, Singapore: 294.

Srivong, P., P. Waisalong, P. Wangsomnuk and P. Pongdontri . 2007. **Comparison of aroma level of Khao Dok Mali 105 and its mutants by expression level of partial BADH2 gene**. Poster Presentation in 1st BMB conference: Biochemistry and molecular biology for the integration of life, April 26-27, 2007, Bangkok, Thailand: 48.

Sun, S.H., F.Y. Gao, X.J. Lu, X.J. Wu, X.D. Wang, G.J. Ren and H. Luo. 2008. Genetic analysis and gene fine mapping of aroma in rice (Oryza sativa L. Cyperales, Poaceae). **Genet. Mol. Biol.** 31: 532–8.

Thompson, J.D., D.G. Higgins and T.J Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Res.** 22: 4673–4680.

Tun, Y.T. 2006. **Characterization of Myanmar rice landraces based on analyses of genetic diversity of various traits**. PhD thesis, Tokyo University of Agriculture: 248.

Unnevher, L.J, B. Duff and B.O Juliano. 1992. **Consumer demand for rice grain quality**. IRRI, Manila Philippines and IDRC, Ottawa, Canada: 248.

Vanavichit, A., W. Kamolsukyunyong, S. Wanchana, S. Tragoonrung , T. Toojinda, S. Wongpornchai, S. Ruengphayak, T. Toojinda and S. Tragoonrung. 2004. **Discovering genes for rice aroma**. In Proceedings of the 1st International Conference on Rice for the future, 31 August-3 September 2001: 71 80,

Kasetsart university, Bangkok, Thailand.

Vanavichit, A., S. Tragoonrung, T. Toojinda, S. Wanchana, W. Kamolsukyunyong. 2005. **Transgenic rice plants with reduced expression of *Os2AP* and elevated levels of 2-acetyl-1-pyrroline**. US patent No. 7,319,181.

Vanavichit, A. and T. Yoshihashi. 2010. Molecular aspects of fragrance and aroma in rice. **Advanc. Bota. Res.** 56: 49-73.

Wanchana, S., W. Kamolsukyunyong , S. Ruengphayak, T. Toojinda, S. Tragoonrung and A. Vanavichit. 2005. A rapid construction of a physical contig across a 4.5 cM region for rice grain aroma facilitates marker enrichment for positional cloning. **Sci. Asia.** 31: 299–306.

Weber, D.J., R. Rohilla and U.S. Singh, 2000. **Chemistry and Biochemistry of Aroma in Scented Rice**. In: Aromatic Rices, Singh, R.K., U.S. Singh and G.S. Khush (Eds.). Science Publishers Inc., Enfield, NH, USA: 29-46.

Widjaja, R., Craske, J.D and M. Wootton. 1996. Comparative studies on volatile components of non-fragrant and fragrant rices. **J. Sci. Food.Agric.** 70: 151-161.

Wilkie, K. and M. Wootton. 2004. Flavour qualities of new Australian fragrant rice cultivars -http://www.rirdc.gov.au/reports/RIC/04-160sum.html. In. Australian Government - Rural Industries Research and Development Corporation (RIRDC)

Wongpanya, R., N. Boonyalai, N. Thammachuchourat, N. Horata, S. Arikit, K.M. Myint, A. Vanavichit and K. Choowongkomon. 2011. Biochemical and enzymatic study of rice BADH wild-type and mutants: an insight into fragrance in rice. **Protein J** 30: 529-538.

Wongpornchai, S., T. Sriseadk and S. Choonvisase. 2003. Identification and Quantitation of the Rice Aroma Compound, 2-Acetyl-1-pyrroline, in Bread Flowers (*Vallaris glabra Ktze*). **J. Agric. Food Chem.** 51: 457-462

Wright, S. 1978. **Evolution and the genetics of populations. Variability within and among natural population**. Chicago: University of Chicago Press: 590

Xu, Y., J.H. Crouch. 2008. Marker-assisted selection in plant breeding: from publications to practice. **Crop Sci** 48: 391–407.

Yamanaka, S., S.A. Jatoi, S.S. Yi, S.L. Kothari, T. Htut and K.N. Watanabe. 2011. Genetic diversity of Myanmar rice and their implementation on management methods. **African J. Biotechn.** 10(8): 1290-1298.

Yoshihashi, T. 2002. Quantitative analysis on 2-acetyl-1-pyrroline of an aromatic rice by stable isotope dilution method and model studies on its formation during cooking. **J**. **Food Sci.** 67: 619–622.

Yoshihashi, T., N.T.T. Huong, V. Surojanametakul, P. Tungtrakul and W. Varanyanond. 2005. Effect of storage conditions on 2-acetyl-1-pyrroline content in aromatic rice variety, Khao Dawk Mali 105. **J. Food Sci.** 70: S34-S37.

Zhu, M., Y. Wang, Y. Zhu and B. R. Lu. 2004. Estimating genetic diversity of rice land races from Yunnan by SSR assay and its implication for conservation. **Acta Bot. Sin.** 46: 1458–1467.

APPENDIX

Appendix Table 1 List of Myanmar accessions assayed for characterization of genetic diversity

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
MSB_311	AUNG LAN DAW	Bago Division	RL	u	1	Na	A	A	ns
MSB_1158	BAHAN HMWE	Ayeyarwady Division	RL	u	1	Ar	P	A	H1
IRGC_32959	BALUGUN KAUKKYI	Ayeyarwady delta zone	U	5	5B	Na	A	A	ns
MSB_537	BAW GYI	Ayeyarwady Division	RL	u	1	Na	A	A	ns
MSB_1774	BAY KYAR	Sagaing Division	RL	u	m	Ar	A	A	H12
MSB_3708	BAY KYAR PAW SAN	Yangon Division	RL	u	5B	Ar	A	P	H5
MSB_211	BU DO THE	Shan State	RU	u	1	Na	A	A	ns
MSB_769	BU RO	Kayin State	RL	u	1	Na	A	A	ns
MSB_2587	BU SABA	Tanitharyi Division	RL	u	1	Na	A	A	H14
MSB_325	CHIN KAUK HNYIN	Sagaing Division	U	u	1	Ar	A	A	H7
IRGC_6804	D 44-1	Southern plain zone	U	5	5B	Na	A	A	H3
MSB_1898	EMAHTA HMWE	Kachin State	RL	u	1	Ar	A	A	H7
MSB_635	EPA TINE	Kayin State	RL	u	1	Na	A	A	ns
MSB_2512	KALA GYI NGA SEIN	Mon State	RL	u	1	Na	A	A	ns

Appendix Table 1 (Continued)

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
MSB_2297	KAMAR KYI SAW	Mon State	RL	u	5B	Ar	A	P	H5
MSB_6071	KAUK HNYIN HMWE	Shan State	U	u	1	Ar	P	A	H1
MSB_6108	KAUK HNYIN NET	Shan State	U	u	5B	Ar	A	P	H5
MSB_6097	KAUK HNYIN PHYU	Shan State	U	u	5B	Ar	A	P	H5
MSB_371	KAUK HNYIN SAW	Kachin State	U	u	1	Ar	u	A	Ns
MSB_498	KAUK KYI THEE DAT	Bago Division	RL	u	1	Na	A	A	Ns
IRGC_33191	KAUKKYI MEEDON	Ayeyarwady delta zone	U	5	1	Na	A	A	Ns
IRGC_33192	KAUKKYISAW	Southern coastal zone	U	5	5B	u	A	A	H6
MSB_240	KHAO KHAN NEW	Shan State	RU	u	1	Na	A	A	Ns
MSB_263	KHAO LONE KHONE	Shan State	RU	u	1	Na	A	A	Ns
MSB_158	KHAO MON HAN YOUNG	Kayah State	RU	u	1	Na	A	A	Ns
MSB_519	KHAO NWAN	Shan State	RU	u	1	Na	A	A	Ns
MSB_218	KHAO SING SAUK	Shan State	RU	u	1	Na	A	A	Ns
MSB_353	KYAR GALAY	Mon State	RL	u	1	Na	A	A	Ns

Appendix Table 1 (Continued)

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
MSB_728	KYAW KHAW	Kachin State	RU	u	1	Na	A	A	Ns
MSB_230	KYEE MASHE	Mandalay Division	RL	u	1	Na	A	A	Ns
IRGC_33280	KYEEARNI	Southern plain zone	U	5	5B	Na	A	A	Ns
IRGC_33282	KYEEME	U	U	5	5B	Na	u	u	Ns
IRGC_33283	KYEENI	Southern plain zone	U	5	5B	Na	A	A	Ns
IRGC_58069	KYET PAUNG	Semi-arid zone	U	5	5A	Ar	P	A	H1
MSB_235	KYET PAUNG SAN	Sagaing Division	RL	u	1	Ar	A	A	H7
IRGC_33287	KYWEKYUT MEEDON	Ayeyarwady delta zone	u	5	m	Na	A	A	Ns
MSB_1932	KYWET THWAR	Yangon Division	u	u	1	Ar	A	A	H10
IRGC_33301	LAWTHAWPHINME (E62-10)	Southern coastal zone	u	5	5B	Na	A	A	Ns
IRGC_33304	LEIKKALAY MEEDON	Southern plain zone	u	5	5B	Na	A	A	H2
MSB_2560	LET TAW YWE BAW	Mon State	RL	u	1	Na	A	A	Ns
MSB_577	LONE PHYU	Mon State	u	u	1	Na	A	A	Ns
MSB_285	MA PO LAY	Ayeyarwady Division	RL	u	1	Na	A	A	Ns

Appendix Table 1 (Continued)

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
MSB_7293	MAN AUNG KAUK YIN	Rakhine State	u	u	5B	Ar	A	A	H4
IRGC_33373	MAYZI	Western coastal zone	u	5	M	Na	A	A	Ns
MSB_398	MEE DON YIN	Mandalay Division	u	u	5B	Ar	A	A	H3
IRGC_33379	MEEDON YIN	Semi-arid zone	u	5	5B	Na	A	A	Ns
IRGC_33380	MEEDON YOYO	Ayeyarwady delta zone	u	5	1	Na	A	A	Ns
MSB_829	MYA WAR NGA CHEIK	Sagaing Division	u	u	1	Ar	u	A	H6
MSB_109	NAGA NEW	Rakhine State	RL	u	1	Na	A	A	Ns
MSB_2283	NAMATHALAY	Sagaing Division	u	u	5A	Ar	P	A	H1
MSB_1809	NAT PYI HMWE	Bago Division	u	u	5B	Ar	u	u	H4
MSB_1382	NGA BYA YIN KAUK HNYIN NET	Sagaing Division	u	u	5B	Ar	A	P	H5
MSB_1777	NGA CHEIK	Kachin State	u	u	5B	Ar	A	A	H4
MSB_5884	NGA CHEIK	Chin State	u	u	1	Ar	A	A	H13
MSB_6101	NGA CHEIK KAUK HNYIN	Shan State	u	u	1	Ar	A	A	Ns
MSB_1790	NGA KYWE	Bago Division	RL	u	1	Ar	u	A	Ns

Appendix Table 1 (Continued)

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
MSB_56	NGA KYWE	Mon State	RL	u	5B	Ar	A	A	H2
MSB_1105	NGA KYWE (U TO)	Ayeyarwady Division	RL	u	5B	Ar	A	A	Ns
IRGC_58152	NGA KYWE GYI	Western coastal zone	u	5	m	Na	A	A	H6
MSB_1638	NGA KYWE NOTE	Bago Division	RL	u	5B	Ar	A	A	H4
MSB_348	NGA KYWE PHYU	Ayeyarwady Division	u	u	1	Ar	A	A	Ns
MSB_352	NGA KYWE PHYU	Ayeyarwady Division	RL	u	1	Ar	A	A	H7
MSB_1114	NGA KYWE TAUNG PYAN	Ayeyarwady Division	RL	u	5B	Ar	A	A	Ns
MSB_1799	NGA KYWE TAUNG PYAN	Yangon Division	RL	u	1	Ar	A	A	H2
MSB_1789	NGA KYWE YIN	Kayin State	RL	u	5B	Ar	A	P	H5
MSB_410	NGA KYWE YIN	Ayeyarwady Division	RL	u	5B	Ar	A	A	Ns
MSB_1871	NGA PYA GYI	Mandalay Division	RL	u	5B	Ar	A	P	H5
MSB_404	NGA YUN WAR	Bago Division	RL	u	1	Na	A	A	Ns
IRGC_33466	NGAKYEIN THEE SHAY	Western coastal zone	u	5	1	Na	A	A	Ns
IRGC_33467	NGAKYWE	Ayeyarwady delta zone	u	5	m	Na	A	A	Ns

Appendix Table 1 (Continued)

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
IRGC_33468	NGAKYWE	Ayeyarwady delta zone	u	5	m	Na	A	A	Ns
IRGC_33469	NGAKYWE	Ayeyarwady delta zone	u	5	5B	Na	A	A	Ns
IRGC_5946	NGAKYWE D 25-4	Southern plain zone	u	5	m	u	A	A	H6
IRGC_11142	NGAKYWE TAUNG PYAN	Ayeyarwady delta zone	u	5	5B	Na	A	A	Ns
IRGC_33478	NGA-KYWE YIN	Eastern plateau	u	5	5B	Na	A	A	H2
IRGC_58154	NGAYA PAUK	Western coastal zone	u	5	m	Na	A	A	Ns
IRGC_33521	NGAYUN TAUNGPYAN	Southern plain zone	u	5	1	Na	A	A	Ns
IRGC_58172	PATEEPU/OO PAUK	Southern plain zone	u	5	5A	Na	A	A	Ns
IRGC_33551	PATHEIN NGAKYWE	Ayeyarwady delta zone	u	5	5B	Na	A	A	Ns
IRGC_33552	PATHEIN NYUNT	Ayeyarwady delta zone	u	5	5B	Ar	A	P	H5
MSB_2579	PAW SAN BAY KYAR	Ayeyarwady Division	RL	u	5B	Ar	A	P	H5
MSB_2877	PAW SAN BAY KYAR	Ayeyarwady Division	RL	u	1	Ar	u	A	H7
MSB_2925	PAW SAN BAY KYAR	Ayeyarwady Division	RL	u	5B	Ar	u	A	Ns
MSB_3163	PAW SAN BAY KYAR	u	RL	u	5B	Ar	u	P	Ns

Appendix Table 1 (Continued)

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
MSB_804	PAW SAN BAY KYAR	Ayeyarwady Division	RL	u	5B	Ar	u	P	H5
MSB_807	PAW SAN BAY KYAR	Ayeyarwady Division	RL	u	5B	Ar	A	A	H2
MSB_1207	PAW SAN HMWE	Ayeyarwady Division	u	u	5B	Ar	A	P	H5
MSB_1641	PAW SAN HMWE	Ayeyarwady Division	RL	u	1	Ar	A	A	Ns
MSB_2082	PAW SAN HMWE	u	RL	u	1	Ar	u	A	Ns
MSB_2502	PAW SAN HMWE	Ayeyarwady Division	RL	u	1	Ar	A	A	H6
MSB_2620	PAW SAN HMWE	Ayeyarwady Division	RL	u	5B	Ar	u	P	Ns
MSB_5802	PAW SAN HMWE	U	RL	u	1	Ar	u	A	Ns
MSB_6973	PAW SAN HMWE BAY KYAR	Rakhine State	RL	u	5B	Ar	u	P	H5
MSB_7005	PAW SAN HMWE KAUK YIN	Rakhine State	RL	u	5B	Ar	A	A	H2
MSB_2577	PAW SAN SHWE WAR	Ayeyarwady Division	u	u	5B	Ar	A	A	H2
MSB_2924	PAW SAN TAUNG PYAN HMWE	Yangon Division	u	u	5B	Ar	u	A	H2
MSB_1128	PAW SAN YIN	Ayeyarwady Division	RL	u	5B	Ar	A	P	H5
IRGC_33590	POKEKYI	Southern plain zone	u	5	5B	Na	A	A	Ns

Appendix Table 1 (Continued)

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
IRGC_33596	PYAPON THEEDAT	Ayeyarwady delta zone	u	5	m	Na	A	A	Ns
MSB_163	SABA NET	Mandalay Division	RL	u	1	Ar	A	A	Ns
MSB_2261	SABA NET	Kayah State	RL	u	m	Ar	u	A	H6
MSB_1915	SABA NET KAUK YIN	Mandalay Division	I	u	5B	Ar	u	A	Ns
IRGC_33605	SABANET KAUKYIN	Semi-arid zone	u	5	5B	Na	A	A	Ns
MSB_682	SABANI HMWE	Tanitharyi Division	RL	u	1	Ar	P	A	H1
IRGC_58191	SAN KAR YAN THAE	Southern plain zone	u	5	6	Na	A	A	H2
MSB_234	SAN SHWE NI	Bago Division	RL	u	1	Na	A	A	Ns
MSB_66	SHWE NI TWAT SUN	Yangon Division	RL	u	1	Na	A	A	Ns
MSB_75	SHWE WAR KAUK HNYIN	Yangon Division	u	u	1	Ar	A	A	H2
IRGC_33668	SHWEDINGA	Semi-arid zone	u	5	5B	Na	A	A	Ns
IRGC_33704	SHWEWA	Ayeyarwady delta zone	u	5	5B	Na	A	A	Ns
IRGC_33708	SHWEWA HNAN	Ayeyarwady delta zone	u	5	5B	Na	A	A	Ns
IRGC_33709	SHWEWA KAMAKYI	Ayeyarwady delta zone	u	5	5B	Na	A	A	Ns

Appendix Table 1 (Continued)

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
IRGC_33706	SHWEWAGYI	Ayeyarwady delta zone	u	5	1	Na	A	A	Ns
IRGC_33712	SHWEWAYIN	Southern plain zone	u	5	1	Na	A	A	Ns
MSB_217	TADAUNG PO	Ayeyarwady Division	D	u	1	Na	A	A	Ns
MSB_1403	TAR SAING	Kachin State	RU	u	1	Na	A	A	Ns
MSB_1628	TAUNG PYAN HMWE	Kayin State	RL	u	1	Ar	A	A	H9
MSB_1621	TAUNG PYAN SA BA NET	Bago Division	RL	u	m	Ar	A	A	H9
MSB_6096	TAUNG YAR KAUK HNYIN	Shan State	u	u	1	Ar	A	A	H2
IRGC_33745	TAUNGDI	Southern plain zone	u	5	5B	Na	A	A	Ns
IRGC_33747	TAUNGGYI KAUKKYI	Eastern plateau zone	u	5	5B	Na	A	A	Ns
IRGC_33752	TAUNGPYAN	Ayeyarwady delta zone	u	5	5B	Na	A	A	Ns
IRGC_33755	TAUNGPYAN KAUKNGE	Ayeyarwady delta zone	u	5	5B	Na	A	A	Ns
IRGC_33757	TAUNGPYANYIN	Southern plain zone	u	5	5B	Na	A	A	Ns
IRGC_33749	TAUNGTI	Southern plain zone	u	5	5B	Na	A	A	Ns
IRGC_33773	THATNU SABANET	Southern plain zone	u	5	5B	Na	A	A	Ns

Appendix Table 1 (Continued)

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
MSB_984	THEE DAT PYA PON	Ayeyarwady Division	RL	u	1	Na	A	A	Ns
MSB_647	THEE HTAT PIN KHINE	Sagaing Division	RL	u	5A	Na	A	A	H2
MSB_67	THEE HTAT NGA SEIN	Yangon Division	RL	u	1	Na	A	A	H6
MSB_1786	THET NU SABA NET PYAN	Yangon Division	RL	u	5B	Ar	u	A	H4
IRGC_33791	THIT PIN	Western coastal zone	u	5	m	Na	A	A	Ns
MSB_65	THONE HNAN PWA	Yangon Division	RL	u	1	Na	A	A	Ns
IRGC_33805	THONESATOE NGAKWE	Southern plain zone	u	5	5B	Na	A	A	Ns
IRGC_33802	THONSATOE	Semi-arid zone	u	5	2	Na	A	A	Ns
MSB_113	TIN TANE	Kachin State	RL	u	1	Na	A	A	Ns
MSB_1791	TYAUNG PYAN YIN	Ayeyarwady Division	RL	u	1	Ar	u	A	H1
MSB_1792	TYAUNG PYAN YIN	Bago Division	RL	u	5B	Ar	u	A	H4
MSB_925	WA KHE MA HNAN KAR	Ayeyarwady Division	D	u	1	Na	A	A	Ns
IRGC_33835	WUNKYAW	Southern plain zone	u	5	1	Na	A	A	Ns
IRGC_32293	YANGON SABA	Southern plain zone	u	5	5B	Ar	A	P	Ns

Appendix Table 1 (Continued)

ID	Variety name	Region of origin	Ecosystem	EG	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
MSB_788	YAR KAUK HNYIN	Shan State	RL	u	1	Ar	A	A	Ns
IRGC_58247	YAT SAUK SABA	Eastern plateau zone	u	5	m	Na	A	A	Ns
MSB_91	YAY MA NAING MIN THAR GYI	Yangon Division	D	u	1	Na	A	A	H7
IRGC_33888	YELAIK MEEDON	U	u	5	5B	u	A	A	H2
MSB_2528	YWA LWE	Kachin State	RU	u	1	Ar	A	A	H7
MSB_33	ZAWTIKA GYI	Yangon Division	RL	u	1	Na	A	A	Ns
MSB_233	ZEIN YIN	Bago Division	RL	u	1	Na	A	A	Ns

RL= rainfed lowland; RU = rainfed upland; D= deepwater; I=irrigated; u = unknown; Ar= aromatic; NA= non-aromatic;
Genotype: Del8 = 8 bp deletion in comparison with Nipponbare sequence; Ins3 = 3 bp insertion in comparison with Nipponbare sequence; A= Indel absent; P= Indel present; Haplotype: ns= not sequenced

Appendix Table 2 List of the varieties in the reference set assayed for characterization of genetic diversity

ID	Variety name	Country	Isozyme group	SSR genotyping	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
IRGC_32380	ABRI	Bhutan	5	YES	5A	u	A	A	H2
IRGC_12386	ARC 10177	India	2	NO	2	Na	u	u	ns
IRGC_12485	ARC 10497	India	5	YES	5A	u	A	A	H4
IRGC_42469	ARC 13829	India	5	YES	5A	u	A	A	H3
IRGC_12331	ARC 7229	India	5	NO	m	Na	u	u	H3
IRGC_6267	ASD 1	India	1	NO	1	Na	u	u	ns
IRGC_7017	AUS 143	India	2	YES	2	Na	u	u	ns
IRGC_328	AZUCENA	Philippines	6	NO	6	Ar	P	A	ns
IRGC_27509	BARAN BORO	Bangladesh	2	NO	2	Na	u	u	ns
IRGC_27798	BASMATI 1	Pakistan	5	YES	5A	Ar	P	A	H1
IRGC_53637	BASMATI 217	India	5	YES	5A	Ar	P	A	H1
IRGC_6426	BASMATI 370	Pakistan	5	NO	5A	Ar	P	A	H1
IRGC_58881	BASMATI LAMO	Nepal	5	YES	5A	Ar	P	A	H1
IRGC_40275	BLACK GORA	India	2	NO	2	Na	u	u	ns

Appendix Table 2 (Continued)

ID	Variety name	Country	Isozyme group	SSR genotyping	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
IRGC_50448	CANELA DE FERRO	Brazil	6	NO	6	Na	u	U	ns
IRGC_27869	CHAHORA 144	Pakistan	5	YES	5A	Ar	P	A	H1
IRGC_58930	CHHOTE DHAN	Nepal	5	YES	m	Na	A	A	H7
IRGC_17052	CHUAN 4	Taiwan	6	NO	6	Na	u	U	ns
IRGC_43372	CICIH BETON	Indonesia	6	NO	6	Na	u	U	ns
IRGC_10658	CUBA 65	Cuba	6	NO	6	Na	u	U	ns
IRGC_5857	DA 13	Bangladesh	5	YES	5A	Ar	P	A	H1
IRGC_12880	DOM SOFID	Iran	5	YES	5A	Ar	P	A	H1
IRGC_12881	DOM ZARD	Iran	5	YES	5A	Ar	P	A	H1
IRGC_32292	DOMSIAH	Iran	5	YES	5A	Ar	P	A	H1
IRGC_26011	DOURADO PRECOCE	Brazil	6	NO	6	Na	u	U	ns
IRGC_10984	FANDRAPOTSY 104	Madagascar	1	NO	m	Na	u	U	ns
IRGC_39261	FIROOZ	Iran	5	YES	5A	Na	A	A	H2
IRGC_32303	GHARIB	Iran	5	YES	m	Ar	P	A	H1

Appendix Table 2 (Continued)

ID	Variety name	Country	Isozyme group	SSR genotyping	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
IRGC_43397	GOTAK GATIK	Indonesia	6	NO	6	Na	u	u	ns
IRGC_33085	GWA NGASEIN	Myanmar	1	NO	1	Na	u	u	ns
IRGC_4122	IGUAPE CATETO	Brazil	6	NO	6	Na	u	u	ns
IRGC_33130	INDANE	Myanmar	6	NO	6	Na	u	u	ns
CIRAD_8351	IR 64	Philippines	1	NO	1	Na	A	A	ns
IRGC_10320	IR 8	Philippines	1	NO	1	Na	u	u	ns
IRGC_28508	IRAT 13	Côte d'Ivoire	6	NO	6	Na	u	u	ns
IRGC_27516	JAGLI BORO	Bangladesh	2	NO	2	Na	u	u	ns
IRGC_9091	JC 1	India	5	YES	5A	Ar	P	A	H1
IRGC_9060	JC 101	India	5	YES	5A	u	P	A	H1
IRGC_9070	JC 149	India	5	YES	5A	u	P	A	H1
IRGC_27967	JHONA 26	Pakistan	2	NO	2	Na	u	u	ns
IRGC_6618	KARKATI 87	Bangladesh	1	NO	1	Na	u	u	ns
IRGC_117617	KASALATH	India	2	NO	2	Na	u	u	ns

Appendix Table 2 (Continued)

ID	Variety name	Country	Isozyme group	SSR genotyping	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
IRGC_33188	KAUKKYI ANI	Myanmar	5	NO	5B	u	u	u	ns
IRGC_27748	KHAO DAWK MALI 105	Thailand	1	NO	1	Ar	P	A	H1
IRGC_23423	KHAO KAP XANG	Laos	6	NO	6	Na	u	u	ns
IRGC_23364	KINANDANG PATONG	Philippines	6	NO	6	Na	u	u	ns
CIRAD_5109	KIRIMINY 1133	Madagascar	5	YES	5A	Ar	A	A	H8
CIRAD_5110	KIRIMINY DE 4 MOIS	Madagascar	5	YES	5A	Ar	A	A	H8
CIRAD_5112	KIRIMINY TYPE BENGALY	Madagascar	5	YES	5A	Ar	A	A	H7
IRGC_11010	MAINTIMOLOTSY 1226	Madagascar	6	NO	6	Na	u	u	ns
IRGC_6087	MAKALIOKA 34	Madagascar	1	NO	1	Na	u	u	ns
IRGC_8182	MALAGKIT PIRURUTONG	Philippines	6	NO	6	Na	u	u	ns
IRGC_23973	MANA MURI	Nepal	5	YES	5A	u	A	A	H4
IRGC_12883	MEHR	Iran	5	YES	5A	u	P	A	H1
IRGC_12048	MOROBEREKAN	Guinea-Conakry	6	NO	6	Na	u	u	ns
IRGC_6298	N 12	India	5	YES	5A	Ar	P	A	H1

Appendix Table 2 (Continued)

ID	Variety name	Country	Isozyme group	SSR genotyping	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
IRGC_6264	N 22	India	2	NO	2	Na	u	u	ns
IRGC_11462	NAM SA GUI 19	Thailand	1	NO	1	Na	u	u	ns
IRGC_191	NHTA 10	India	6	NO	6	Na	u	u	ns
IRGC_12731	NIPPONBARE	Japan	6	NO	6	Na	u	u	H15
IRGC_22710	NONA BOKRA	India	1	NO	1	Na	u	u	ns
IRGC_38690	NPE 253	Pakistan	6	NO	6	Na	u	u	ns
IRGC_5999	PANKHARI 203	India	5	YES	5A	Ar	P	A	H1
IRGC_8266	PEH PI NUO	China	6	NO	6	Na	u	u	ns
IRGC_32571	PETA	Indonesia	1	NO	1	Na	u	u	ns
IRGC_32399	PHUDUGEY	Bhutan	5	YES	5A	Na	A	A	H4
IRGC_8948	POKKALI	Sri Lanka	1	NO	1	Na	u	u	ns
IRGC_6386	PTB 25	India	1	NO	1	Na	u	u	ns
IRGC_8952	RATHUWEE	Sri Lanka	1	NO	1	Na	u	u	ns
IRGC_8197	SHAI KUH	China	1	NO	1	Na	u	u	ns

Appendix Table 2 (Continued)

ID	Variety name	Country	Isozyme group	SSR genotyping	Structure group	Aroma	Genotype Del8	Genotype Ins3	Haplotype
IRGC_5418	SINTANE DIOFOR	Burkina Fasso	1	NO	m	Na	u	u	ns
IRGC_8256	SURJAMKUHI	India	2	NO	2	Na	u	u	ns
IRGC_46768	T 26	India	5	YES	m	u	A	A	H7
IRGC_105	TAICHUNG NATIVE 1	Taiwan	1	NO	1	Na	u	u	ns
IRGC_47743	TAINUNG 67	Taiwan	1	NO	1	Na	u	u	ns
IRGC_32362	TCHAMPA	India	5	YES	5A	Na	A	A	H2
IRGC_81093	TEQUING	China	1	NO	1	Na	u	u	ns
IRGC_32576	TETEP	Vietnam	1	NO	1	Na	u	u	ns
IRGC_43675	TREMBESE	Indonesia	6	NO	6	Na	u	u	ns
IRGC_33888	YELAIK MEEDON	Myanmar	5	NO	5B	u	A	A	H1

u = unknown; Ar= aromatic; NA= non-aromatic; Genotype: A= Indel absent; P= Indel present; Haplotype: ns= not sequenced

Appendix Table 3 List of the primers used to amplify segments of the *Os2AP* gene

Name	Forward primer	Reverse primer	Annealing temperature	Segment amplified	Source
INS3	GTCCTGTTCAATCTTGCAGC	CTTGATGCAACCATGTCATA	55°C	E13	Myint *et al*., 2012
DEL8	TGCTCCTTTGTCATCACACC	TTTCCACCAAGTTCCAGTGA	55°C	E7	Vanavichit *et al*., 2008
BADH 1	CGAAGTCCGTACCAACTGC	GGCCGTGAGCCATATACACT	55°C	E1-I1	Kovach *et al*., 2009
BADH 2	AGTTGGAAGCATGGCTGATT	CCAGCTCAGATTTCCTCTCG	55°C	E2-I2	Kovach *et al*., 2009
BADH 3	GATTGTGGGAAGCCTCTTGA	CGATAGGCTCTTTCCGAAGAT	55°C	E3-I3	Kovach *et al*., 2009
BADH 4	ATCTTCGGAAAGAGCCTATCG	AGGAGCTACCTTCCATGTTGC	55°C	E4-I4	Kovach *et al*., 2009
BADH 5	CTATCCTCTCCTGATGGCAAC	TGGCTACTAGAATGATGCTCAAAG	55°C	E5-E6	Kovach *et al*., 2009
BADH 6	TGCTCCTTTGTCATCACACC	GGTCCAAAAGCAACCAAAGA	55°C	E7-E8	Vanavichit *et al*., 2008
BADH 7	GCCATGCCAACTGAGTAAAG	CAATTTTATTCGCTCTGTGC	55°C	E8-E9	Vanavichit *et al*., 2008
BADH 8	TGCAACATCGCGTCTTATTC	GCAACTAGCAAGAGCATACACC	55°C	E9-E10	Vanavichit *et al*., 2008
BADH 9	GCCCGTTGTTAGTGAAGGAC	GTACCATCCCCACGGCTCAT	55°C	E10-I13	Vanavichit *et al*., 2008
BADH 10	ACCTGACATCATGCCTTTGG	CCGGTCATCAGCTAACTTCC	55°C	E14 -3'U	Vanavichit *et al*., 2008
BADH 11	GCCCGTTGTTAGTGAAGGAC	CAATGGCTTCTTCTTCAGTGC	55°C	I11-E12	Designed by us
BADH 12	CGAGCGATGCCAGAGATTA	CGTCACTTGCTTGACGCTTA	55°C	E13-E14	Designed by us

CIRRICULUM VITAE

NAME : Ms. Khin Myo Myint

BIRTH DATE : May 8, 1972

BIRTH PLACE : Myingyan, Mandalay Division, Myanmar

EDUCATION :

YEAR	INSTITUTE	DEGREE/DIPLOMA
1995	Yezin Agricultural Univ., Myanmar	B.Agr.Sc.
2000	Yezin Technological Univ., Myanmar	M.Sc

POSITION/TITLE : Senior Researcher

WORK PLACE : Plant Biotechnology Centre
Myanma Agriculture Service (MAS), Yangon,
Ministry of Agriculture and Irrigation, Myanmar

SCHOLARSHIP/AWARDS : Thai-French Trilateral Development Cooperation (TICA- TRF-FE) under the Royal Golden Jubilee Programme