

Development of allele-specific SNP markers for *PPR10* gene at *Rf4* locus of Fertility Restorer Gene for Identification of Maintainer and Restorer lines

Saengtong Ponjaroenkit*, Kanokwan Janpen, Chotipa Sakulsingharoj and Varaporn Sangtong

¹Program in Genetics, Faculty of Science, Maejo University, Chiang Mai 50290, Thailand.

*Corresponding author: saengtong@mju.ac.th

ABSTRACT

Cytoplasmic male sterility (CMS) -associated genes are located in the mitochondrial genome whereas restorer of fertility (*Rf*) genes are located in the nuclear genome. Two major restorer gene loci of Wild abortive-CMS (WA-CMS), *Rf3* and *Rf4*, are required for the production of viable pollen. The *Rf4* locus of *indica* rice cultivars were reported, which contains four RNA-binding pentatricopeptide repeat protein (PPR) encoding genes. Preliminary sequence analysis of reported *PPR* gene at *Rf4* locus from maintainer and restorer lines revealed the difference in size of protein only PPR10 protein that resulted from single nucleotide polymorphism (SNP). In this study, *PPR10* gene was isolated from four Thai rice cultivars by Polymerase Chain Reaction (PCR) of genomic DNA. The nucleotide sequences of cloned *PPR10* gene from Khao Dawk Mali 105, Nahng Mon S-4, Pathum Thani 1 and Suphan Buri 1 were analyzed by ClustalX program. One of fifteen SNPs at position 1,392 caused nonsense mutation and classified PPR10 protein into two groups by the length. Tetra-primer amplification refractory mutation system-polymerase chain reaction (ARMS-PCR) is fast, reliable and low cost. Then, four primers of ARMS-PCR were developed to use as an allele-specific DNA marker for *Rf4* locus. The tetra-primer ARMS-PCR products of five maintainer cultivars and four restorer cultivars were G allele (416 bp) and T allele (242 bp), respectively. Therefore, tetra-primer ARMS-PCR could be allele specific *PPR10* gene marker and used to identify maintainer and restorer Thai rice cultivars of WA-CMS. These allele-specific *PPR10* gene markers will be used for germplasm genotyping and hybrid rice breeding program.

Keywords: allele-specific SNP marker; tetra-primer amplification refractory mutation system-polymerase chain reaction (ARMS-PCR); restorer of fertility (*Rf*) gene; pentatricopeptide repeat (PPR) protein; wild abortive- cytoplasmic male sterility (WA-CMS)

INTRODUCTION

Cytoplasmic male sterility (CMS) is maternal inheritance. CMS-associated genes are located in the mitochondrial genome whereas restorer of fertility (*Rf*) genes are located in the nuclear genome. CMS plant is unable to produce functional pollen. Several CMS systems with different cytoplasmic/nuclear combination have been used in hybrid rice production and wild abortive-CMS (WA-CMS) system has been widely used. WA-CMS system, which consists of CMS, maintainer and fertility restorer lines, has been applied for the commercial production of hybrid seeds. Two major restorer gene loci of WA-CMS, *Rf3* and *Rf4*, which are required for the production of viable pollen were mapped to chromosome 1 and 10 respectively (El-Namaky *et al.*, 2016). The *Rf4* locus plays slightly more important role than *Rf3* locus in restorer ability (Cai *et al.*, 2013). The majority of restorer fertility (RF) or restorer fertility like (RFL) proteins belongs to the RNA-binding pentatricopeptide repeat protein (PPR) family. Nuclear encoded RF proteins are imported into mitochondria and bind to the CMS transcripts, preventing translation or inducing RNA cleavage. The PPR proteins can be divided into P and PLS subfamilies. PLS-class proteins are involved in RNA editing whereas P-class proteins are involved in stability of organelle transcripts and intron splicing. The RF or RFL proteins belong to the P-class PPR subfamily and are characterized by the presence of tandem array of 15-20 PPR motif (Melonek *et al.*, 2016).

Presently, *Rf4* locus of *indica* rice cultivars, IR24 Minghui63 and 93-11, were cloned and sequenced. This *Rf4* locus contains four PPR encoding genes which are *PPR7/PPR458*, *PPR8/PPR782b*, *PPR9/PPR782a* and *PPR10/PPR454* (Kazama and Toriyama 2014; Tang *et al.*, 2014). Then, genomic fragments of each *PPR* gene of IR24 and Minghui63 (restorer) were introduced into Taichung 65 and Jin23A (CMS), respectively by Agrobacterium-mediated transformation. Only transgenic lines of

PPR9/PPR782a showed partially restored male fertility (Kazama and Toriyama 2014; Tang *et al.*, 2014). This result indicated that *PPR9/PPR782a* gene may be a major restorer gene at *Rf4* locus and other *PPR* genes may play a minor role. They may exhibit additive effect in restorer fertility ability. However, *Rf4* can often restore fertility of other CMS systems (BT-CMS, HL-CMS) and encodes PPR proteins which are highly identical to these of *Rf1a* gene of BT-CMS (Huang *et al.*, 2014). Therefore, other *PPR* genes such as *PPR10* gene could be restorer lines in other CMS lines or other CMS systems.

DNA markers that enable to determine *Rf3* and *Rf4* loci of maintainer and restorer lines are mostly SSR markers (Fan *et al.*, 2015; Kiani 2015; Waza and Jaiswal 2016). The RM171 and RM 258 were used to detect *Rf4* locus for WA-CMS in 300 rice cultivars. Currently, one set of PCR-based codominant markers of *Rf4* locus that could distinguish maintainers from restorers was reported (Pranathi *et al.*, 2016). A specific gene marker has not been reported for any *Rf* gene.

Single nucleotide polymorphisms (SNPs) are the most common form of DNA variation. SNPs can be used in plants as gene markers for many breeding applications, population studies, and germplasm screening. Tetra-primer amplification refractory mutation system-polymerase chain reaction (ARMS-PCR) is a simple and economical method to detect SNPs. This technique is based on the use of four primers in a single reaction and followed by agarose gel electrophoresis (Medrano and de Oliveira 2014). This method was used for SNP genotyping in barley (Chiapparino *et al.*, 2004). Our preliminary sequence analysis of reported *PPR7*, *PPR9* and *PPR10* genes from maintainer and restorer lines revealed the difference of protein size of these lines only in *PPR10* protein that resulted from SNP whereas the size of *PPR7* and *PPR9* proteins was the same in both rice lines (data not shown). Therefore, *PPR10* gene could be a first candidate gene for gene specific DNA marker development of *Rf4* locus from Thai rice cultivars. Compared with other methods for SNPs genotyping, tetra-primer ARMS-PCR is fast, reliable and low cost. In this study, the tetra-primer ARMS-PCR was used to develop allele-specific DNA markers for *PPR10* gene, which is one of fertility restorer genes at *Rf4* locus. The *PPR10* genes were isolated from Thai rice cultivars which may different from the previous reported *PPR10* gene. Then, sequences of isolated *PPR10* genes were used to develop the *PPR10* specific gene marker that can classify maintainer and restorer lines of Thai rice cultivars.

MATERIALS AND METHODS

Plant Materials

The maintainer and restorer lines for WA-CMS were classified by crossing between CMS lines (IR58025A) as a female parent and Thai rice cultivars as a male parent. The F₁ pollen was stained with 2% IKI₂ solution. If F₁ had sterile pollen, male parents were identified as a maintainer line. In contrast, if F₁ had fertile pollen, male parents were identified as a restorer line. In this study, Khao Dawk Mali 105 (KDML105), Jao Hom Nin (JHN), Nahng Mon S-4 (NMS4) and Pin Gaew 56 (PG56) were classified as maintainers. Pathum Thani 1 (PTT1), Suphan Buri 1 (SPR1), Look Daeng Pattani (LDP) and RD47 were classified as restorers (Pantuli 2013). Moreover, the hybrid of IR58025A x PTT1 was used to validate the allele specific gene marker.

DNA extraction

Genomic DNA was extracted from the leaves of seedlings by the cetyl trimethylammonium bromide method with minor modifications (Hwang and Kim 2000).

Cloning and Sequence analysis of *PPR10* gene

A comparison of rice *PPR10* genes from Minghui63 (KJ680251), 93-11 (KJ680253) and IR24 (AB900791) reported in GenBank was performed to identify conserved regions. These two regions were used to design primers for amplification of *PPR10* gene from four Thai rice cultivars which were KDML105, NMS4, PTT1 and SPR1.

The PCR reaction was performed in total volume of 20 µl containing 50 ng of genomic DNA, 1X of Thermo Scientific Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific), 0.5 µM of each primer. The PCR condition was as follows: 98 °C for 30 sec, followed by 35 cycles of 98 °C for 10 sec, 60 °C for 10 sec and 72 °C for 90 sec, and final 72 °C for 5 min. PCR products were analyzed by 1% (w/v) agarose gel electrophoresis.

The expected bands about 2.2 kb of *PPR10* gene were purified from gel and cloned into pBluescript II SK vector. Two recombinant clones of each rice cultivar *PPR10* gene were sequenced by 1st BASE (Malaysia) using universal primers and *PPR10* specific primer (Os10g0495100R). The 2.2 contiguous sequence of each rice cultivar *PPR10* gene was assembled by BioEdit program.

The nucleotide and amino acid sequences of cloned *PPR10* gene from KDML105, NMS4, PTT1, and SPR1 were analyzed and compared with *PPR10*

genes reported in GenBank (Minghui63, 93-11 and IR24). Multiple sequence alignments were performed using ClustalX2.1 program (Larkin *et al.*, 2007). The InterProScan from <https://www.ebi.ac.uk/interpro/interproscan> was used to find PPR motif within PPR10 protein.

Primer design and validation of SNP genotyping

The nucleotide sequence comparison revealed one SNP (G/T) causing nonsense mutation where GAA was changed to TAA (stop codon). Then, tetra-primer ARMS-PCR was employed to identify this SNP. In this method four primers were used to amplify a target fragment from DNA containing the SNP, generating amplicons representing each of the two allelic forms. Primers were used to amplify fragments of different sizes for each allele that can be resolved by agarose gel electrophoresis. Tetra -primers were designed using BatchPrimer3 v1.0 program (<http://batchprimer3.bioinformatics.ucdavis>).

The gene fragment was amplified from outer forward primer and outer reverse primer. The G specific allele was amplified from inner forward primer and outer reverse primer whereas the T specific allele was amplified from outer forward primer and inner reverse primer (Figure 1).

The tetra-primer ARMS-PCR reaction was performed in total volume of 20 μ l containing 20 ng of genomic DNA, 1X MYTAQ RED MIX (BIOLINE, UK), 0.125 μ M of each inner primer, 0.5 μ M of each outer primer. The PCR condition was as follows: 94 $^{\circ}$ C for 5 min, followed by 45 cycles of 94 $^{\circ}$ C for 30 sec, 60 $^{\circ}$ C for 15 sec and 72 $^{\circ}$ C for 30 sec, and final 72 $^{\circ}$ C for 5 min. PCR products were analyzed by 2% (w/v) agarose gel electrophoresis. The PCR products were purified and sequenced to verify the amplicons. Then, primer set was tested on five maintainer cultivars, four restorer cultivars and one hybrid of IR58025A x PTT1 to validate the accuracy of the tetra-primer ARMS-PCR.

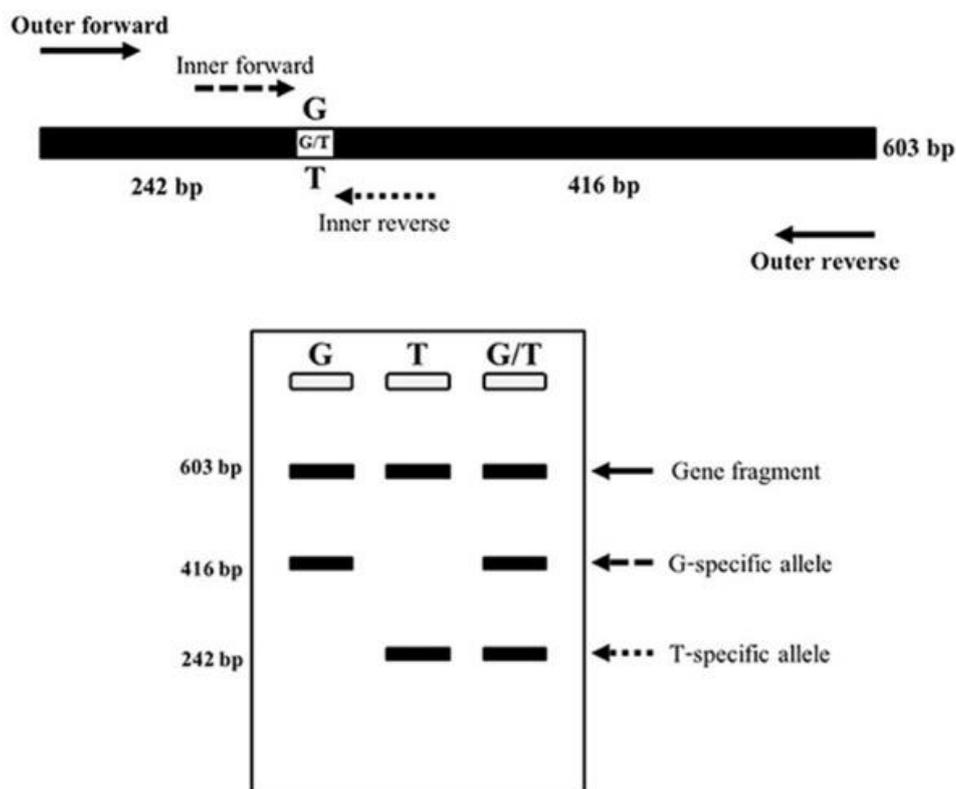


Figure 1. Diagram representation of tetra- primer ARMS-PCR for allele-specific SNP genotyping. Four primers are used: the two outer primers amplify a gene fragment that contains an SNP (white box). The inner primers are designed to amplify the G or T allele. The outer forward primer and inner reverse primer were used to amplify an amplicon representing the T allele whereas the inner forward primer and outer reverse primer were used to amplify an amplicon representing the G allele. The two allele specific amplicons differ in length that can be discriminated by gel electrophoresis. (Adapted from (Chiapparino *et al.*, 2004)).

RESULTS AND DISCUSSIONS

Cloning and sequence analysis of *PPR10* gene

Our preliminary study of *PPR7*, *PPR9* and *PPR10* genes from maintainer (93-11 and Nipponbare) and restorer (IR24 and Minghui63) lines revealed that the difference in size of protein was only found in *PPR10* protein that resulted from SNP. On the other hand, *PPR7* and *PPR9* protein sizes were the same in both rice lines. Therefore, *PPR10* gene could be a first candidate gene for DNA marker development of *Rf4* locus from four Thai rice cultivars. However, *PPR10* gene from Thai rice cultivars may be not the same as reported in GenBank. In this study, the isolation of *PPR10* gene was performed from Thai rice cultivars. Since *PPR10* gene is intronless gene. Then, gene amplification

was performed using PCR of genomic DNA.

The *PPR10* nucleotide sequence from GenBank comparison revealed conserved region which was used to design primers of *PPR10* gene, *PPR10*-Forward and *PPR10*-Reverse primers as shown in Table 1. The *PPR10*-Forward and *PPR*-Reverse primers were located 29 nucleotides upstream of start codon and 149 nucleotides downstream of stop codon of *PPR10* gene, respectively and covered full-length coding region of *PPR10* gene. These primers were used to amplify *PPR10* gene from Thai rice cultivars. Then, the 2.2 kb expected band was cloned and sequenced. The nucleotide sequences of cloned *PPR10* gene from KDML105 and NMS4 were 2,188 bp whereas those from PTT1 and SPR1 were 2,194 bp.

Table 1 List of primers used in this study.

Primer set	Name	Sequence (5'→3')	Allele	Amplicon size (bp)
<i>PPR10</i> gene	PPR10-Forward	TGCTGCTGCACCTGTCAGC	-	2,200
	PPR10-Reverse	GCCGATTAGGGTAGTATCGGGG	-	
	Os10g0495100R	CCTCAGCCTTCTCCCATTTG	-	-
Tetra primers	PPR10- Inner forward	AGAAGGCTGAGGAGITAATTTTTG	G	416
	PPR10- Inner reverse	CTGGACAGATGCCTTGATCCAACATTTA	T	242
	PPR10- Outer forward	CAAAATGAGGCAGCAAGGAT		603
	PPR10- Outer reverse	CCGATCGATGCACATATGAC		

The nucleotide sequences of *PPR10* genes from four rice cultivars were compared with three *PPR10* gene sequences from GenBank. The nucleotide sequence comparison of maintainers and restorers revealed 15 base substitutions or SNPs in coding region (position 29-1,745) which were nine transition and six transversion mutations. However, only five SNPs that caused missense or neutral mutations. One SNP at position 1,392 (G to T) caused nonsense mutation which resulted from transversion mutation as shown in Table 2. The SNP (T791-to-A) that caused nonsense mutation (TAT to TAA) was also observed in *Rf5/rf5* gene in the HL-CMS line (Huang *et al.*, 2014). The InDel mutation was found only at position 1,504 (downstream of stop codon) which did not affect *PPR10* protein from restorers (Figure 2A). The high frequency of transversion mutations may suggest that maintainers were diverged from restorers for a long time. The comparison of amino acids of *PPR10* protein revealed that all of maintainers and restorers had the same amino acid sequences (Figure 2B). The missense mutation happened more frequently than others (Table

2) which may influence RNA binding specificity. The study of *RF* gene at the same locus in different *Oryza* species showed diverse amino acid sequences that had different functions because they could bind different RNA targets (Melonek *et al.*, 2016). Moreover, based on SNP at position 1,392 (G¹³⁹² to T¹³⁹²), *PPR10* protein was classified into two groups which were 454 (T allele) and 569 (G allele) amino acids in length. Since *PPR10* was *PPR* protein, then *PPR* motif was searched by InterProScan program. The *PPR10* protein of restorers contained ten *PPR* motifs whereas maintainers contained thirteen *PPR* motifs (Figure 2B). The median number of *PPR* motifs per protein of P-class and RFLs were nine and thirteen motifs, respectively (Melonek *et al.*, 2016). From the number of *PPR* motifs, *PPR10* protein of maintainers and restorers may have different function. Therefore, *PPR10* gene of maintainers and restorers could have restorer ability in different CMS lines or CMS systems. There was reported that restorer lines containing *Rf4* can often restore the fertility of BT-CMS and HL-CMS (Huang *et al.*, 2014)

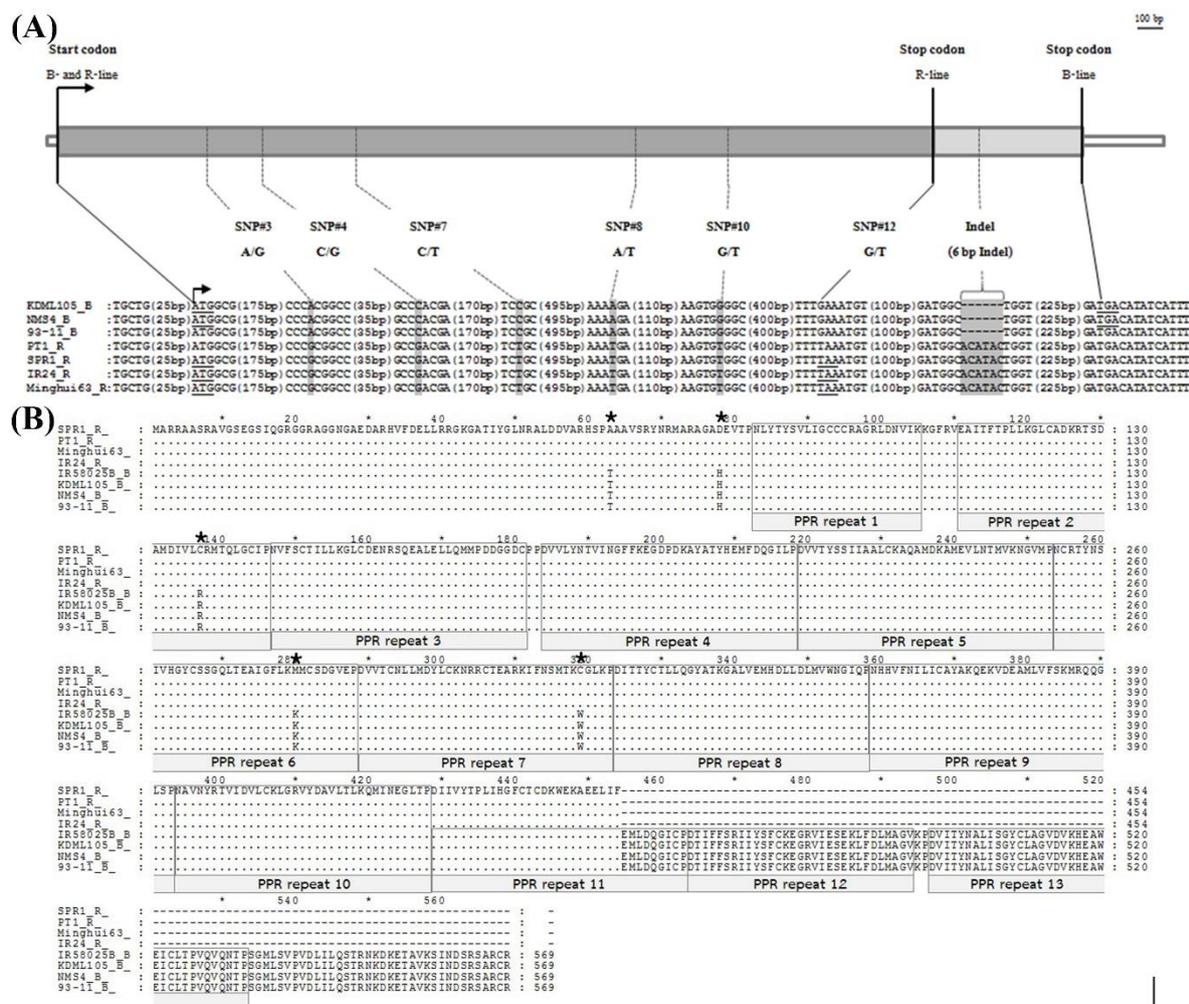


Figure 2. Nucleotide and amino acid sequence comparison of *PPR10* gene. The isolated *PPR10* genes were from KDML105, NMS4, SPR1, and PTT1 cultivars. The reported *PPR10* gene were from 93-11, IR24 and Minghui63. The KDML105, NMS4 and 93-11 were maintainers whereas SPR1, PTT1, IR24 and Minghui63 were restorers. (A) Schematic representation of 2.2 kb *PPR10* gene and grey boxes were coding regions. The nucleotide sequence comparison revealed five SNPs in coding region and six base pair deletion (InDel). (B) Amino acid comparison revealed five point mutations indicated by * and box indicated PPR motif.

Table 2 Five SNPs in coding region that affected amino acid of *PPR10* genes from 7 rice cultivars. The sequences of maintainers were used as reference lines.

Number (#)	Position	SNPs	Type of base substitution	Change in amino acid	Type of mutation
3	216	A to G	Transition	T to A	Missense
4	261	C to G	Transversion	H to D	Missense
7	438	C to T	Transition	R to C	Missense
8	868	A to T	Transversion	K to M	Missense
10	986	G to T	Transversion	W to C	Neutral
12	1,392	G to T	Transversion	E to Stop	Nonsense

The G allele was amplified from PPR10-Inner forward and PPR10- Outer reverse primer whereas T allele was amplified from PPR10-Outer forward and PPR10-Inner reverse primers. The amplicons of each allele were 416 and 242 bp, respectively. Then, the primer set was tested on five maintainer cultivars, four restorer cultivars and one hybrid of IR58025A x PTT1 to validate the accuracy of the tetra-primer ARMS-PCR. The PCR products of five maintainer cultivars and four restorer cultivars were 416 bp and 242 bp which were G and T alleles, respectively (Table 4). Moreover, the hybrid of IR58025A x PTT1 showed both G allele from IR58025A and T allele from PT1 (Figure 3B).

Table 4 The allele specific genotyping by tetra-primer ARMS-PCR.

Rice cultivar	Restore ability	SNP _{1,392}
IR58025B	Maintainer	G
KDML105		G
JHN		G
NMS4		G
PG56		G
PTT1	Restorer	T
SPR1		T
LDP		T
RD47		T
F ₁ (IR58025Ax PTT1)	-	G/T

Then, PCR products of 416 bp from KDML105 and these of 242 and 603 bp from LDP, were purified and sequenced. The nucleotide sequences of these fragments were as expected (data not shown). Therefore, tetra-primer ARMS-PCR could be the first allele specific *PPR10* gene marker and could be used to classify maintainer or restorer Thai rice cultivars of WA-CMS. The SSR markers were used to identify maintainer and restorer lines of WA-CMS which most of them were the linked part of *Rf4* locus. Then, they were not suitable across species (Miah *et al.*, 2013). This allele specific *PPR10* gene marker will be used for germplasm genotyping and the hybrid rice breeding program.

ACKNOWLEDGEMENTS

This study was supported by the grant from National Research Council of Thailand.

REFERENCES

Cai J, Liao QP, Dai ZJ, Zhu HT, Zeng RZ, Zhang ZM, Zhang G-Q (2013) Allelic differentiations and effects of the *Rf3* and *Rf4* genes on fertility

restoration in rice with wild abortive cytoplasmic male sterility. *Biol Plant* 57: 274–280.

Chen L-K, Yan X-C, Dai J-H, Chen S-P, Liu Y-Z, Wang H, Chen Z-Q, Guo T (2017) Significant association of the novel *Rf4*-targeted SNP marker with the restorer for WA-CMS in different rice backgrounds and its utilization in molecular screening. *J Integr Agric* 16: 2128–2135.

Chiapparino E, Lee D, Donini P (2004) Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS-PCR. *Genome* 47: 414–420.

El-Namaky R, Sedeek S, Moukoubi YD, Ortiz R, Manneh B (2016) Microsatellite-Aided Screening for Fertility Restoration Genes (*Rf*) Facilitates Hybrid Improvement. *Rice Sci* 23: 160–164.

Fan F, Li N, Wang J, Liu X, Liu J, Zhu Y, Li S (2015) Molecular marker-directed development of a novel cytoplasmic male sterile line in rice. *Mol Breed* 35: 212.

Huang J-Z, E Z-G, Zhang H-L, Shu Q-Y (2014) Workable male sterility systems for hybrid rice: Genetics, biochemistry, molecular biology, and utilization. *Rice* 7: 13.

Hwang S-K, Kim Y-M (2000) A simple and reliable method for preparation of cross-contamination-free plant genomic DNA for PCR-based detection of transgenes. *J Biochem Mol Biol* 33: 537–546.

Kazama T, Toriyama K (2014) A fertility restorer gene, *Rf4*, widely used for hybrid rice breeding encodes a pentatricopeptide repeat protein. *Rice* 7: 28.

Kiani G (2015) Validation of SSR Markers Linked to Restoring Fertility (*Rf*) Genes and Genotyping of Rice Lines at *Rf* Loci. *J Agric Sci Technol* 17: 1931–1938.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

Medrano RFV, de Oliveira CA (2014) Guidelines for the tetra-primer ARMS-PCR technique development. *Mol Biotechnol* 56: 599–608.

Melonek J, Stone JD, Small I (2016) Evolutionary plasticity of restorer-of-fertility-like proteins in rice. *Sci Rep* 6: 35152.

Miah G, Rafii MY, Ismail MR, Puteh AB, Rahim HA, Islam KN, Latif MA (2013) A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *Int J Mol Sci* 14: 22499–22528.

Pantuli S (2013) Selection Thailand rice varieties for maintainer lines and restore lines and improvement

- Thailand rice varieties for cytoplasmic male sterility lines. Special project Bachelor of Science Program in Agriculture (Agronomy) (Maejo University).
- Pranathi K, Viraktamath BC, Neeraja CN, Balachandran SM, Hari prasad AS, Koteswara RP, Revathi P, Senguttuvel P, Hajira S K, Balachiranjeevi C H, *et al.* (2016) Development and validation of candidate gene-specific markers for the major fertility restorer genes, Rf4 and Rf3 in rice. *Mol Breed* 36: 145.
- Tang H, Luo D, Zhou D, Zhang Q, Tian D, Zheng X, Chen L, Liu YG (2014) The Rice Restorer Rf4 for wild-abortive cytoplasmic male sterility encodes a PPR protein that functions in reduction of WA352 transcripts. *Mol Plant* 7:1497–1500.
- Waza SA, Jaiswal HK (2016) Identification of elite grain quality restorers and maintainers for wa cms lines of rice (*Oryza sativa* L.). *SABRAO J Breed Genet* 48: 145–153.