Vol 10, No 1&2 (2017) : JAN - AUG

Genomics and Genetics

The Official Journal of Genetics Society of Thailand

www.tci-thaijo.org/index.php/gst

ISSN 2539-6129

Genomics and Genetics
The Official Journal of Genetics Society of Thailand

Contact office: Division of Molecular Medicine, 4th floor, Siriraj Medical Research Building, 2 Wang Lang road, Siriraj, Bangkoknoi, Bangkok 10700, Thailand.
Phone: (+66) 2 419-2754-59 **Fax:** (+66) 2 411-0169
E-mail: gesoth@gmail.com
E-Journal: http://www.tci-thaijo.org/index.php/gst/index

Advisory Board

Kiichi Fukui, Professor
Graduate School of Engineering, Osaka University and Faculty of Medicine, Tottori University, Japan

Pa-thai Yenchitsomanus, Professor
Faculty of Medicine Siriraj Hospital, Mahidol University

Budsaba Rerkamnuaychoke, Associate Professor
Faculty of Medicine Ramathibodi Hospital, Mahidol University

Thawornchai Limjindaporn, Associate Professor
Faculty of Medicine Siriraj Hospital, Mahidol University

Editor in Chief

Arinthip Thamchaipenet, Associate Professor
Faculty of Science, Kasetsart University

Associate Editor

Kornsorn Srikulnath, Assistant Professor
Faculty of Science, Kasetsart University

Krieng Kanchanawatee
Faculty of Science, Chulalongkorn University

Rachaneekorn Tammachote, Assistant Professor
Faculty of Science, Chulalongkorn University

Sompid Samipak, Assistant Professor
Faculty of Science, Kasetsart University

Piyada Juntawong
Faculty of Science, Kasetsart University

Teerasak E-kobon
Faculty of Science, Kasetsart University

Editorial Board

Theeraphap Chareonviriyaphap, Professor
Faculty of Agriculture, Kasetsart University

Chanitra Thuwajit, Associate Professor
Faculty of Medicine Siriraj Hospital, Mahidol University

Chanwit Tribuddharat, Associate Professor
Faculty of Medicine Siriraj Hospital, Mahidol University

Daoroong Kangwanpong, Associate Professor
Faculty of Science, Chiang Mai University

Skorn Koonawootrittriron, Associate Professor
Faculty of Agriculture, Kasetsart University

Supachitra Chadchawan, Associate Professor
Faculty of Science, Chulalongkorn University

Supawadee Poompuang, Associate Professor
Faculty of Fisheries, Kasetsart University

Chatchawan Jantasuriyarat, Assistant Professor
Faculty of Science, Kasetsart University

Onruthai Pinyakong, Assistant Professor
Faculty of Science, Chulalongkorn University

Uraiwan Arunyawat, Assistant Professor
Faculty of Science, Kasetsart University

Sissades Tongsima
National Center for Genetic Engineering and Biotechnology, NSTDA

Wanna Thongnoppakhun
Faculty of Medicine Siriraj Hospital, Mahidol University

Journal Management

Nunghathai Sawasdee
Kiattichai Saetai
Chatchai Phunumkang
Wanvisa Glinsrisuk

Genomics and Genetics
The Official Journal of Genetics Society of Thailand

The journal of Genetics Society of Thailand http://www.tci-thaijo.org/index.php/gst

Volume 10, Number 1&2 January–April 2017

Contents

Research Articles

Newly developed automated-massively parallel sequencing presenting the accuracy of hepatitis C virus genotyping and application on resistance-associated variants detection.....1–6
Narathon Manee, Wasun Chantratita, Ekawat Pasomsub

Heat shock protein 70 gene polymorphism in KU–Phuphan black–bone chicken.....7–12
Piyamas Phongkaew, Panuwat Khumpeerawat

Expression Analysis of Rice Polygalacturonase cDNA Responding to Brown Planthopper [*Nilaparvata lugens* (Stål)].....13–20
Sugunya Suebsan, Supranee Sitthiphrom, Kanta Sangwijit, Mondhon Sanguansermsri, Somboon Anuntalabhochai1

Influence of single nucleotide polymorphisms in the *BCL11A*, *HBS1L-MYB* intergenic region, and *HBB* gene cluster on the fetal hemoglobin levels in Bangladeshi patients with β-thalassemia/hemoglobin E disease.....21–26
Jenjira Chalerm, Mahmood Ahmed Chowdhury, Chayanon Peerapittayamongkol, Kittiphong Paiboonsukwong, Suthat Fucharoen, Orapan Sripichai

Identification of genes involving in salt tolerance using GWAS data based on Na^+ content in local Thai rice leaves and Arabidopsis orthologous gene validation.....27–37
Monthira Suntiparpluacha, Nattapol Tammachote, Rachaneekorn Tammachote

Development of allele-specific SNP markers for *PPR10* gene at *Rf4* locus of Fertility Restorer Gene for Identification of Maintainer and Restorer lines.....38–45
Saengtong Ponjaroenkit, Kanokwan Janpen, Chotipa Sakulsingharoj and Varaporn Sangtong

Newly developed automated-massively parallel sequencing presenting the accuracy of hepatitis C virus genotyping and application on resistance-associated variants detection

Narathon Manee[1], Wasun Chantratita[1, 2], Ekawat Pasomsub[1, 2]*

[1]Department of Clinical Pathology, Faculty of Medicine, Ramathibodi Hospital Mahidol University, Bangkok 10400, Thailand
[2]Virology laboratory and Center for Medical Genomics, Faculty of Medicine, Ramathibodi Hospital Mahidol University, Bangkok 10400, Thailand

**Corresponding author: ekawat.pas@mahidol.ac.th*

ABSTRACT

Most hepatitis C virus (HCV) infection becomes chronic diseases and predictably develops into liver cirrhosis or hepatocellular carcinoma. HCV can be classified into 7 genotypes (GT) and multiple subtypes (ST). HCV treatment guidelines focus on determining the accuracy of HCV genotyping and subtyping, before starting treatment in all HCV-infected individuals. Efficient HCV genotyping and subtyping have become essential tools for indicating the optimal treatment. Current routine assays for HCV genotyping are easy and quick turnaround, but apparently subtyping is limited to identification. This study evaluated the accuracy of HCV genotyping in a newly developed automated-massively parallel sequencing (MPS) technique of previously identified HCV-infected specimens. The results indicated that by using the reference method, MPS and a bioinformatic pipeline, the genotype and subtype had identical accuracy of 95.65%. Interestingly, MPS could detect the remaining 4.35% as mixed-HCV type infected specimens, while the reference method could not. This result indicated that the MPS technique was highly accurate in identifying the HCV genotype and mixed-type infection. In addition, MPS can detect resistance-associated variants (RAVs) simultaneously with the HCV genotyping, especially in HCV GT1-infected specimens. Since HCV GT1 has a highly frequent presence of pre-existing RAVs, with resistance associated to direct-acting antivirals (DAAs), this study found that up to 60% of HCV GT1 sample were RAVs. This discovery of important RAVs in the patient infected with HCV GT1 could be beneficial in guiding the therapy decisions while avoiding the use of some DAAs. Thereby, the opportunity for patients to access the best treatment is enhanced.

Keywords: HCV genotyping; mixed-HCV infection; massively parallel sequencing; Sanger sequencing

INTRODUCTION

Hepatitis C virus (HCV) is a positive single-stranded RNA virus in the *Hepacivirus* genus of the *Flaviviridae* family (Chevaliez *et al*., 2006; Li *et al*., 2015). Multiple strains, 7 genotypes (GT) and approximately 67 subtypes (ST) of HCV have been distributed differently in various regions worldwide (Smith *et al.*, 2014). No vaccine is currently available to prevent HCV infection, it can be cured with the antiviral drug, which result in cure rates approximately 90%. The treatment goal is sustained virological response (SVR; undetectable HCV RNA three or six months post therapy) (WHO, 2016). The initial treatment of chronic hepatitis C was pegylated interferon (Peg-IFN) combined with ribavirin (RBV). Later, direct-acting antivirals (DAAs in four groups: NS3 inhibitors, NS5A inhibitors, NS5B Nucleoside and Non-Nucleoside Polymerase Inhibitors) was developed and utilized with/without pegylated interferon and Ribavirin (WHO, 2016). The treatment of HCV-infected patients is different by HCV genotyping (GTs 2-6). Moreover, treatment guidance of HCV genotype 1 was specific to a and b subtypes (Chung *et al*., 2015; EASL, 2015; Thai Association for the Study of the Liver, 2016). Thereby, it is necessary to determine the strains of HCV (HCV genotyping) in all patients prior starting the treatment, which benefits the treatment plan that includes duration time to treat, dose of drugs, suitable regimens and prediction of treatment response (Chung *et al*., 2015; EASL, 2015; WHO, 2016). Besides, a single patient may be infected by more than one HCV strain (mixed-type infection). Mixed-type infection was observed frequently in about

14–39 percent of persons sharing a hypodermic needle (Cunningham *et al*., 2015). Methods with low sensitivity are unable to determine all strains that infect an individual patient, as a different HCV-minority genome will be undetected and in turn, causes an inappropriate treatment plan by reducing the response rate to therapy, or treatment failure from undetected, non-sensitive minority strains (Cunningham *et al.*, 2015).

In an era where DAAs are used increasingly, variants associated with drug resistance (resistance-associated variants: RAVs) were found and frequently presented in patients infected with genotype 1. RAVs were detected on three DAA-targeted genes of the HCV genome, and observed frequently in the NS3 and NS5A regions, but were rarely found in the NS5B region. RAVs were identified in patients before starting therapy with an estimated 8.6 and 1.4 percent in subtypes 1a and1b, respectively. Moreover, HCV RAVs have been found in most patients infected with genotype 1 after treatment failure with a first generation NS3/NS4A protease inhibitor (Ahmed and Felmlee, 2015). The Q80K polymorphism is a variant that is common and occurs naturally at *NS3* gene position of the HCV subtype 1a (19−48%) (Poveda *et al*., 2014), which significantly reduces the response rate of Simeprevir treatment (second-generation NS3/4A protease inhibitor), when compared to patients infected with HCV subtype 1a that does not appear with Q80K (Wyles, 2013; Wyles and Gutierrez, 2014; Ahmed and Felmlee, 2015; Lenz *et al*., 2015; Sarrazin *et al*., 2015). The HCV treatment guide from the World Health Organization (WHO), European Association for the Study of the liver (EASL), America American Association for the Study of Liver Diseases and the Infectious Diseases Society of America (AASLD/IDSA) advised against the use of Simeprevir when Q80K is presented in naive patients infected with HCV subtype 1a, as well as those re-starting treatment after previous treatment failure (retreatment) (Au *et al*., 2015; EASL, 2015; WHO, 2016). Approximately 10−15 percent of patients infected with HCV genotype 1 was detected with variants related to drug resistance to Daclatasvir, Ledipasvir, Elbasvir or Ombitasvir (NS5A inhibitors) (AASLD-IDSA, 2016a; 2016b). These RAVs are important, found commonly in patients infected with HCV subtype 1a and occur at the M28, Q30, L31, and Y93 positions in the HCV *NS5A* gene. The effects of these variants reduced the response rate of NS5A inhibitors treatment by more than 5 times when compared with patients without these variants (Wyles and Gutierrez, 2014; AASLD-IDSA, 2016a; 2016b). The EASL and AASLD/IDSA recommend avoiding the use of some NS5A inhibitors in patients infected with genotype 1 if they appear to have RAVs. The AASLD/IDSA recommends detecting these RAVs before re-starting treatment in patients experiencing treatment failure (AASLD-IDSA, 2016b).

The 5'UTR (untranslated region) and the core region are common for HCV genotyping. However, there are limitations in the accuracy of results that they do not cover the identification of all subtypes (Avo *et al*., 2013; Liu *et al*., 2015; Quer *et al*., 2015). The *NS5B* gene is capable of identifying all strains with high precision, and used commonly in HCV epidemiology. It is therefore considered as the gold standard for HCV genotyping (Enache and Enache, 2008; Gryadunov *et al.*, 2010). The VERSANT HCV Genotype 2.0 Line probe assay (LiPA) and the Abbott m2000 RealTime HCV Genotype II are popular methods used for HCV genotyping in clinical diagnostic laboratories. However, both approaches have the disadvantage of inaccuracy in genotype separation when compared to the reference method, and they also are weak in identifying the HCV subtype level (Avo *et al*., 2013; Liu *et al*., 2015; Quer *et al*., 2015).

The first-generation sequencing or Sanger sequencing is a reference method for HCV genotyping, performing by DNA synthetic HCV RNA at the *NS5B* gene position. However, Sanger sequencing has the disadvantage of sequencing only one sample per run and is therefore inappropriate for use in a clinical laboratory. Second-generation sequencing; also known as massively parallel sequencing (MPS) technology, is a newly developed technology with high throughput and high sensitivity (Cunningham *et al.*, 2015). MPS was developed recently as a new model that worked under the automatic system, in contrast to the older version. Furthermore, this newly developed automated-MPS added the ability of detecting HCV RAVs, which can be performed simultaneously with HCV genotyping. This study evaluated the performance of the newly developed automated-MPS in determining accurate GTs and STs including evaluation of an additional application performance for identifying RAVs often found in the HCV GT1.

MATERIALS AND METHODS

Samples

This retrospective study was conducted in 23 HCV-infected patients, who were previously identified as HCV genotype by using the Versant-Line Probe Assay (HCV genotyping LiPA2.0; Siemens Healthcare Diagnostics, Eragny, France). All of the HCV-infected patients were collected randomly between March and

August 2016 without patient history-associated demographic and clinical information. The inclusion criteria comprised tested samples with HCV viral load of more than 1,000 copies/mL, with sufficient plasma volume for using the newly developed automated-MPS, and Sanger sequencing to confirming the HCV genotyping result. This study was approved by the Committee on Human Rights Related to Research Involving Human Subjects, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand and performed based on the principles of the Declaration of Helsinki (ID 09-58-43).

The newly developed automated-massively parallel sequencing

The MPS-based *Sentosa*® SQ HCV Genotyping Assay was performed on the Ion Torrent PGM platform (Vela Operations Singapore Pte Ltd, The Kendall, 50 Science Park Road, Singapore). It worked under an automatic system as follows: HCV RNA extraction from 530 µl of samples and PCR set-up were performed on the *Sentosa*® robotic SX101 instrument with primers specific to the *NS5B*, *NS5A* and *NS3* genes. The off-board RT-PCR was administered on the Veriti® Dx 96-Well Thermal Cycler. After that, the reaction plate was transferred back onto the *Sentosa*® SX101 worktable in order to continue the library preparation, which generated 200-nucleotide fragments of a DNA amplicon, followed by adaptors ligated to the fragments. The final step of the library preparation was presented with a DNA library pool containing 15 samples and one HCV System Control. The DNA library pool was carried out on the *Sentosa*® ST401 instrument by clonal amplification in a scheme of PCR emulsion. Sequencing was executed on a *Sentosa*® SQ301 Sequencer based on the measurement changes of pH from release of hydrogen ions, due to nucleotide incorporation, before converting into an electrical voltage signal. *Sentosa*® SQ Suite software performed the primary analysis continuously on the raw sequencing data generated by the *Sentosa*® SQ301 Server. Then, the data was transferred to a *Sentosa*® SQ Reporter for the secondary analysis and a generation report (comprising GTs calling concomitantly with RAVs calling, which is found often in GT1). The report contained a full audit trail relating to the actions within the run, and also released output files representing reads (uBAM), mapped reads (BAM), and variant calls (VCF).

Data Analysis

Bioinformatic programs were used to analyze a very large number of sequenced reads from sequencing. All of the sequenced reads were mapped to all complete HCV genomes from NCBI and LANL database by using Ion Torrent TMAP 4.0.5. A sequenced read which could specify a map to the NS5B region of HCV genomes (mapped NS5B specific read) was analyzed for HCV genotyping, while all of the sequenced reads that could specify a map to *NS3*, *NS5A* and *NS5B* regions of complete HCV genomes were analyzed for the HCV resistance-associated variants occurring in the sample. A MIRA Assembler 3.4.1 was used to assemble mapped NS5B specific reads for HCV genotyping into NS5B contigs, which were then aligned to all *NS5B* sequences from the *NS5B* database by using BLAST. Top hit strain was chosen if the *NS5B* alignment score showed ≥ 80% similarity, and subsequent hits were chosen if their alignment score was ≥ 0.98 of the top hit. A phylogenetic Tree was used to determine the genotype.

When detecting HCV RAVs, all mapped specific reads with *NS3*, *NS5A* and *NS5B* regions of complete HCV genomes, which only indicated HCV genotype 1, were assembled to NS3, NS5A and NS5B contigs by using the MIRA Assembler 3.4.1. This step was carried out only if the read inferred genotype 1a and 1b, and then the assembled contigs aligned to reference genome of genotype 1a and 1b with the NCBI and LANL database, respectively, by using BLAST. Moreover, the mapped reads were aligned to the assembled contigs in order to obtain the variants (mutation) frequency, as in some cases, different mutations may be assembled into separate contigs. Finally, the mutations were called from reads rather than contigs. Variants were called per Codon and output in terms of AA change. However, in partially detecting RAVs, the phenotype of resistance-association drugs and therapy guidance decisions had not been discussed. Therefore, the geno2pheno web service (http://hcv.geno2pheno.org/index.php) was used to determine the resistance-association variant for each drug (Geno2pheno hcv, 2011).

Sanger sequencing

HCV RNA was extracted from 400 µL of EDTA-plasma sample. The PCR product was generated by RT-PCR amplification of the HCV RNA *NS5B* region, and sequenced directly by Sanger sequencing (Biosystems® 3500 Series Genetic Analyzer: Integrated DNA Technologies, Singapore). The consensus sequences were submitted to geno2pheno (HCV) web service for HCV genotype and subtype analyses.

Reference sequences

All HCV genome sequences available in the National Center for Biotechnology Information (NCBI) and Los Alamos National Laboratory (LANL) database were used as the reference sequences for the *NS5B*, *NS5A* and *NS3* genes for HCV genotyping. The GenBank accession numbers NC004102 and EU256045 were used as references for the variant calling of ST1a and ST1b, respectively, in partial RAVs detection.

Statistical analysis

Kappa statistics were used to evaluate inter-rater agreement (inter-method) between massively automated parallel sequencing and the reference method. The Kappa value ranged from 0 to 1.00, but generally, a Kappa of more than 0.80 is considered a satisfactory agreement (McHugh, 2012; Tan *et al*., 2015). A *p*-value of < 0.05 was considered statistically significant. Kappa statistics were performed using SPSS Software version 20.

RESULTS AND DISCUSSION

This study classified evaluation of the result into 2 parts; HCV genotyping and partial RAVs detection, both of which were obtained from the newly developed automated-MPS. Detail in the Table 1 shows the HCV genotyping result of MPS compared with Sanger sequencing as a reference method. Both approaches show identical results of HCV GTs and STs as 95.65% (22/23). The remaining 4.35% (1/23) was interesting as MPS could identify mixed-type HCV infection of GT 6n and 3b, while the reference method was unable to detect the GT 6n, which infected as another mixed-type in this sample (Sample no. 23). Sample number 3 interestingly had a low viral load of 219 IU/ml, but MPS as well as the reference method could accurately determine the genotype and subtype. This data indicated that MPS had a high accuracy in determining strains of HCV, which is consistent with the reference method (Quer *et al*., 2015). Furthermore, this data showed that MPS had high sensitivity in detecting infection with multiple HCV strains in a patient (Cunningham *et al*., 2015; Quer *et al*., 2015). Besides, the Kappa coefficient was 1.00 combined with a *p*-value < 0.001, which showed the statistical significance of perfect HCV genotyping agreement between MPS and the reference method. Thus, Kappa statistic indicated that automated-MPS was equivalent to the reference method in being reliable and valid in determining the HCV genotype and subtype.

Table 1 HCV genotype obtained from two methods; newly developed automated- massively parallel sequencing versus the reference method.

Sample No.	Viral load (IU/ml)	Genotyping result	
		NS5B Direct Sanger sequencing (reference method)	*Sentosa*® SQ HCV Genotyping Assay (newly developed-MPS)
1	3,022,691	1a	1a
2	4,393,604	1a	1a
3	219	1b	1b
4	1,075,521	1b	1b
5	2,793,458	1b	1b
6	5,330,000	3a	3a
7	156,973	3b	3b
8	11,051,463	3a	3a
9	11,051	3a	3a
10	1,826,305	3a	3a
11	825,356	6f	6f
12	2,502,365	6i	6i
13	3,795,542	6i	6i
14	13,537	6e	6e
15	6,805,737	6n	6n
16	655,832	6i	6i
17	511,188	6i	6i
18	278,112	6f	6f
19	10,147	3a	3a
20	80,517	6n	6n
21	4,599,512	6v	6v
22	6,998,899	6e	6e
23	676,280	3b	6n + 3b

In the partial RAVs detection, Table 2 shows the important mutation list detected in HCV GT1 samples obtained by additional MPS application, and predictions of phenotypic resistance to HCV direct-acting antivirals (DAAs) by the Geno2pheno web-service. This study observed RAVs as 60% (3/5) of the HCV GT1 samples; where there were Q30H combined with Y93H in one of the ST1a infected specimens (sample no.1). Likewise, the Y93H also was presented in a ST1b infected specimen (sample no.5). The Q30H and Y93H played an important clinical role by resisting NS5A inhibitors (Daclatasvir; Ledipasvir; Ombitasvir and Elbasvir). Moreover, the Q80K mutation was found in a ST1a infected specimen (sample no.2), which had clinical importance as resistant to Simeprevir (NS3/NS4A protease inhibitor). These detected RAVs in HCV GT1-infected patients are currently important mutations, and the EASL and AASLD/IDSA suggest to avoid the use of some DAAs contained in treatment regimens. This is of concern because these drugs (Boceprevir; Sofosbuvir; Simeprevir; Daclatasvir; Ledipasvir; Ombitasvir, etc.) are commonly used in Thailand to treat patients infected with hepatitis virus (Thai Association for the Study of the Liver, 2016). Furthermore, it was evident that all samples infected with HCV ST1a showed RAVs, while only one of three samples infected with HCV ST1b showed them, thus suggests that samples infected with HCV ST1a had a greater incidence of RAVs than ST1b, which is consistent with other studies (Poveda *et al*., 2014; Wyles and Gutierrez, 2014). In conclusion, the newly developed automated-MPS promises to be a reliable tool for subtype and genotype determinations, and also has an improvement in sensitivity for detecting mixed HCV infections. However, this study was limited by the small sample size in the mixed HCV-infection risk group. Therefore, large sample sizes may better demonstrate the capabilities of the MPS to identify different HCV strains in mixed-infected in individual patients. In additional, MPS might be unable to detect mixed-type infection when the minor strains contain <1% of the total viral population (Cunningham *et al*., 2015). A significant advantage of using MPS is its ability to determine the HCV subtype and genotype and also identify RAVs simultaneously, which benefits clinical decision making and better treatment plans with DAA-containing regimens, especially for patients infected with the HCV genotype 1. This newly developed method has an automatic workflow that helps to save time, reduce manual steps and save cost in the long term when compared to traditional methods. It also has the advantage of its ability to sequence more than one sample, and more than one gene per run, thus, it is appropriate to use in a clinical diagnostic laboratory.

Table 2. List of important mutations obtained by additional application of newly developed-MPS, including predictions of phenotypic resistance to HCV direct-acting antivirals (DAAs).

Sample No.	Genotyping result	Mutation result	
		RAVs	Resistance associated to
1	1a	Q30H, Y93H	NS5A inhibitors
2	1a	Q80K	NS3/NS4A inhibitor
3	1b	-	
4	1b	-	
5	1b	Y93H	NS5A inhibitors

ACKNOWLEDGEMENTS

This study thanks the virology laboratory and Center for Medical Genomics in the Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand for valuable advice and materials**.**

REFERENCES

AASLD-IDSA panel (2016a) Initial treatment of HCV infection. Hepatitis C guidance: Recommendations for Testing, Managing, and Treating Hepatitis C, http://www.hcvguidelines.org/full-report/initial-treatment-hcv-infection (November 2016)

AASLD-IDSA panel (2016b) Retreatment of a person with chronic HCV infection in whom prior therapy has failed. Hepatitis C guidance: Recommendations for Testing, Managing, and Treating Hepatitis C, http://www.hcvguidelines.org/full-report/retreatment-persons-whom-prior-therapy-has-failed (November 2016)

Ahmed A, Felmlee DJ (2015) Mechanisms of hepatitis C viral resistance to direct acting antivirals. Viruses 7: 6716–6729.

Au TH, Destache CJ, Vivekanandan R (2015) Hepatitis C therapy: Looking toward interferon-sparing regimens. J Am Pharm Assoc 55: e72–e86.

Avo AP, Agua Doce I, Andrade A, Padua E, (2013) Hepatitis C virus subtyping based on sequencing of the C/E1 and NS5B genomic regions in comparison to a commercially available line probe assay. J Med Virol 85: 815–22.

Chevaliez S, Pawlotsky JM (2006) HCV genome and life cycle. In: Tan SL (ed) Genomes and molecular biology. Hepatitis C Viruses, Horizon Bioscience, pp. 5–47.

Chung RT, Davis GL, Jensen DM, Masur H, Saag MS, Thomas DL (2015) Hepatitis C guidance: AASLD-IDSA recommendations for testing, managing, and treating adults infected with hepatitis C virus. Hepatology 62: 932–954.

Cunningham EB, Applegate TL, Lloyd AR, Dore GJ, Grebely J (2015) Mixed HCV infection and reinfection in people who inject drugs-impact on therapy. Nat Rev Gastroenterol Hepatol 12: 218–30.

EASL (2015) EASL Recommendations on Treatment of Hepatitis C 2015. J Hepatol 63: 199–236.

Enache EL, Enache LS (2008) Versant HCV genotype 2.0 assay (LiPA) in hepatitis C virus genotype determination. Rev Romana Med Lab 12: 47–53.

Geno2pheno hcv [Internet]. Germany: GENAFOR; 2011 [updated 2017 Feb 24; cited 2017 Mar 13]. Available from: http://hcv.geno2pheno.org/index.php.

Gryadunov D, Nicot F, Dubois M, Mikhailovich V, Zasedatelev A, Izopet J (2010) Hepatitis C virus genotyping using an oligonucleotide microarray based on the NS5B sequence. J Clin Microbiol 48: 3910–3917.

Lenz O, Verbinnen T, Fevery B, Tambuyzer L, Vijgen L, Peeters M, Buelens A, Ceulemans H, Beumont M, Picchio G, *et al.* (2015) Virology analyses of HCV isolates from genotype 1-infected patients treated with simeprevir plus peginterferon/ribavirin in Phase IIb/III studies. J Hepatol 62: 1008–1014.

Li Y, Yamane D, Masaki T, Lemon SM (2015) The yin and yang of hepatitis c: Synthesis and decay of hepatitis c virus RNA. Nat Rev Microbiol 13: 544–558.

Liu CH, Liang CC, Liu CJ, Lin CL, Su TH, Yang HC, Chen PJ, Chen DS, Kao JH (2015) Comparison of Abbott realtime HCV genotype II with Versant line probe assay 2.0 for hepatitis C virus genotyping. J Clin Microbiol 53: 1754–1757.

McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Med 22: 276–282, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/?report=printable (November 2016)

Poveda E, Wyles DL, Mena Á, Pedreira JD, Castro-Iglesias Á, Cachay E (2014) Update on hepatitis C virus resistance to direct-acting antiviral agents. Antiviral Res 108:181–191.

Quer J, Gregori J, Rodriguez F, Buti M, Madejon A, Perez del Pulgar S, Garcia Cehic D, Casillas R, Blasi M, Homs M, *et al.* (2015) High-resolution hepatitis C virus subtyping using NS5B deep sequencing and phylogeny, an alternative to current methods. J Clin Microbiol 53: 219–26.

Sarrazin C, Lathouwers E, Peeters M, Daems B, Buelens A, Witek J, Wyckmans Y, Fevery B, Verbinnen T, Ghys A, *et al.* (2015) Prevalence of the hepatitis C virus NS3 polymorphism Q80K in genotype 1 patients in the European region. Antiviral Res 116: 10–16.

Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, Simmonds P (2014) Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology 59: 318–27.

Tan CL, Hassali MA, Saleem F, Shafie AA, Aljadhey H, Gan VB (2015) Development, test-retest reliability and validity of the Pharmacy Value-Added Services Questionnaire (PVASQ). Pharm Pract 13: 598

Thai Association for the Study of the Liver (2016) Thailand Practice Guideline for Management of Chronic Hepatitis C 2016. Thai Association for the Study of Liver, pp. 7–32.

World Health Organization (2014) Guidelines for the screening, sare and treatment of persons with hepatitis C infection.

Wyles DL (2013) Antiviral resistance and the future landscape of hepatitis C virus infection therapy. J Infect Dis 207: S33–S39.

Wyles DL, Gutierrez JA (2014) Importance of HCV genotype 1 subtypes for drug resistance and response to therapy. J Viral Hepat 21: 229–40.

Heat shock protein 70 gene polymorphism in KU–Phuphan black–bone chicken

Piyamas Phongkaew*, Panuwat Khumpeerawat

Department of Agriculture and Resources, Faculty of Natural Resources and Agro–Industry, Kasetsart University, Chalermphrakiat Sakonnakhon Province Campus, Sakon Nakhon, Thailand.

**Corresponding author: csnpmk@ku.ac.th*

ABSTRACT

The objective of this study was to investigate the polymorphism of heat shock protein 70 (*hsp*70) gene in the second generation of KU–Phuphan black-bone chicken populations comparing with crossbreed native chicken and commercial layer chicken populations. Thirty, twenty–five, twenty–three and twenty–three samples of the second generation of KU–Phuphan black–bone female chickens (FB) and male chickens (MB), crossbreed native chickens (BR), and the layer commercial line (LA) were genotyped *hsp*70 gene by the PCR–RFLP technique. The *hsp*70 gene showed two loci, locus *C* generated by *Eae*I and locus *M* created by *Mme*I. The locus *C* consisted of two alleles (C_1 and C_2) providing three genotypes of C_1C_1, C_1C_2 and C_2C_2. The allele and genotype frequencies of KU–Phuphan black-bone chicken populations were not different from other Thai indigenous breeds. The locus *M* consisted of two alleles (M_1 and M_2) whereas the two genotypes of M_1M_2 and M_2M_2 were observed. The allele M_1 showed low allele frequency and was only found in the MB population. The allele M_2 showed monomorphism in FB, BR and LA populations. The male KU–Phuphan black-bone chickens exhibited different M_1M_2 genotype from other detected populations but the allele frequency of MB was less than other Thai indigenous chickens. The genotype frequency of the candidate thermotolerance genotype C_1C_2/M_2M_2 of both KU–Phuphan black-bone chicken populations was higher than other genotypes. According to the results, KU–Phuphan black-bone chickens showed variation in genotypes, especially for locus *C*, and contained high genotype frequency of the candidate thermotolerance genotype. Therefore, the locus *C* and the candidate genotype of *hsp*70 gene could be used as a genetic marker related with heat tolerance for selection and the breeding program of KU–Phuphan black-bone chickens to create heat tolerance chickens.

Keywords: heat shock protein 70 gene; heat stress; KU–Phuphan black–bone chicken; PCR–RFLP

INTRODUCTION

Tropical climate affects animal production. High ambient temperature is a major impact on performance of animals especially for poultry because they have no sweat glands, rapid metabolism, and high body temperature (Al-Aqil and Zulkifli, 2009). When high ambient temperatures are coupled with high humidity, the combination can become critical because the heat increment of animals is difficult to release from their bodies (Gaviol *et al*., 2008) and also can cause heat stress in chickens. Heat stress causes high mortality, low productivity, decline of feed intake and growth rate (Cooper and Washburn, 1998; Akaboot *et al*., 2010; Boonkum *et al*., 2014). Therefore, animal breeders attended to select heat tolerance poultry lines for solving these problems. At present, genetic markers are used for selection and breeding programs to detect the genes related with economic traits with marker assisted selection (MAS).

KU–Phuphan black–bone chicken is purebred black–bone chicken that was selected from native Mongolia black–bone chicken. KU–Phuphan black–bone chicken was selected under the project of the Faculty of Natural Resources and Agro Industry, Kasetsart University, Thailand. The selective breeding created the chicken to look different from other black–bone chickens such as the large body size, high growth rate, high egg production and disease resistance. Moreover, they were selected to have white color feathers and black color meat, skin, bones, tibias and combs. The black color is due to the melanin that is a polymer complex of the amino acid tyrosine. Natural melanin shows biochemical activities, including photoprotection, antioxidation, free radical sequestration and immunomodulatory effects (Hung *et al*., 2002). To improve the selective breeding program, the developed KU–Phuphan black–bone chickens should be detected the variation of the functional genes

related with the important commercial traits. The heat shock protein 70 (*hsp*70) gene is a candidate gene that is related with the heat tolerance trait. The *hsp*70 is the most widely studied member of the *hsp* family. This gene plays important roles for the survival and production performance of chickens, especially in high ambient temperature. Beckham *et al*. (2008) reported that lack of *hsp*70 gene led to a marked decrease in the heat resistance of cells. Tamzil *et al*. (2013) revealed about the polymorphisms of the *hsp*70 gene resulting in different *hsp*70 genotypes that are associated with differing heat tolerance levels in chickens. In addition, the association between polymorphisms of the *hsp*70 gene and heat tolerance has been regarded as a marker for selecting heat-resistant breeds of chickens (Mazzi *et al*., 2003; Gaviol *et al*., 2008). However, there is no evidence that polymorphisms of the *hsp*70 gene can be used effectively to select heat–tolerant KU–Phuphan black–bone chickens. Our study aims to observe the polymorphism of *hsp*70 of the second generation of KU–Phuphan black–bone chickens comparing with crossbreed native chickens and layer commercial chickens using PCR–RFLP.

MATERIALS AND METHODS

Chickens and sample collection

Thirty, twenty–five, twenty–three and twenty–three samples of the second generation of KU–Phuphan black–bone female chickens (FB) and male chickens (MB), crossbreed native chickens (BR), and the layer commercial line (LA) were studied. Blood samples (0.3 μL) were collected from the wing veins by syringes containing100 μL EDTA (0.5 M, pH 8.0) and then transferred to the laboratory to extract the genomic DNA. Animal care and all experimental procedures were approved by the Animal Experiment Committee, Kasetsart University.

Genomic DNA extraction

Genomic DNA was obtained from 0.1 μL of whole blood using Genomic DNA isolation kit (Invitrogen®, California, USA). The quality of genomic DNA was determined using 1% agarose gel and the DNA concentration was calculated using the VC 100bp Plus DNA Ladder marker (Vivantis, Selengor Darul Ehsan, Malaysia).

PCR amplification of the chicken *hsp*70 gene

The reactions were assembled in a 25 μl reaction volume containing: 100 ng genomic DNA, 0.2 μM of each primer, 1x buffer (including 2 mM $MgCl_2$), 100 μM dNTPs (dATP, dTTP, dCTP and dGTP) and 1 unit of *Taq* DNA polymerase (Invitrogen®, California, USA). PCR amplification was performed using primers designed by Mazzi *et al*. (2003): F: 5′–AACCGCACCACACCCAGCTATG–3′ and R: 5′–CTGGGAGTCGTTGAAGTAAGCG–3′. The cycling protocol was 5 min at 94ºC, 34 cycles of denaturing at 94ºC for 30 sec, annealing at 62ºC for 30 sec and extending at 72ºC for 30 sec, with a final extension at 72ºC for 5 min. The amplification of *hsp*70 was displayed in 2% agarose gel.

Restriction with *Eae*I and *Mme*I

The pools contained 1 μg PCR product and were digested with the restriction enzyme *Eae*I and *Mme*I (NEB®, New England, USA) for 60 and 40 min, respectively, at 37°C and then followed by 20 min at 65°C. The samples were separated by electrophoresis in a 3% agarose gel for 40 min, at 100 V. Genotypes of *hsp* 70 were determined at locus *C* for *Eae*I and locus *M* for *Mme*I as following Duangduen *et al*. (2008) and Tunim *et al*. (2010).

Statistical analysis

Genotype frequencies, allele frequencies and Hardy–Weinberg equilibrium (HWE) based on the Chi–square (χ^2) were estimated using Genepop (v 1.2) (Raymond and Rousset, 1995). Observed heterozygosity (H_o), expected heterozygosity (H_e) and inbreeding coefficient (F_{IS}) were performed using FSTAT (v 2.9.3.2) (Goudet, 2002).

RESULTS AND DISCUSSION

Heat-shock proteins (HSP) play important roles in the folding, translocation, and refolding/degradation of proteins can affect the heat resistant in living organisms. According to Zhang *et al*. (2002), the polymorphism regulatory and coding regions of the *hsp*70 provided the different heat tolerance capability trait in chickens. In this study, the polymorphisms of *hsp*70 were detected in four chicken populations; KU–Phuphan black–bone female chickens (FB) and male chickens (MB), crossbreed native chickens (BR) and layer commercial chickens (LA).

Genotype and allele frequencies

At locus *C*, three populations (FB, MB, LA) were under Hardy–Weinberg equilibrium, except BR population (P<0.05). The departure from HWE might due to many reasons such as the population derived from various groups of native chickens leading to be subpopulations, exceeding of homozygotes which

explained by non-random mating and founder event (inbreeding) (Frankham *et al*., 2002). The BR population could strongly explained the heterozygote deficit because they showed low value of observed heterozygosity. A 360 bp amplified fragment of *hsp*70 was obtained for all breeds. Explained by Mazzi *et al*. (2003), this amplified fragment contains a point mutation of *hsp*70 and also the mutation created different patterns related with the tolerance of heat. In our study, the different restriction patterns of the *hsp70* were observed after digestion with the enzymes *Eae*I and *Mme*I. The *Eae*I digestion generated a polymorphic site called the locus *C* which contains two alleles: C_1 (229 bp) and C_2 (240 bp) (Figure 1).

The allele frequencies of all populations were not different. Allele C_1 had frequencies in the range of 0.43 to 0.59 and allele C_2 had frequencies in the range of 0.41 to 0.57. The allele frequencies of our study were similar to the other Thai indigenous chickens (Pradu Hangdam C_1 = 0.45 and C_2 = 0.55, Lueng Hangkhao C_1 = 0.54 and C_2 = 0.46, and Dang C_1 = 0.55 and C_2 = 0.45) reported by Tunim *et al*. (2010). The locus *C* established three genotypes, C_1C_1, C_1C_2 and C_2C_2. The data was shown in Table 1. Genotype C_1C_2 was found for almost all individual chickens, except in BR population. The genotype frequency of C_1C_2 was 0.26, 0.39, 0.40 and 0.5 for BR, LA, MB, FB populations, respectively. The high genotype frequencies were found in both female and male KU–Phuphan black–bone chicken. This implied the KU–Phuphan black–bone chicken populations showed high polymorphism of *hsp*70. Mazzi *et al*. (2003) suggested that the polymorphism of *hsp*70 might be used by breeders to produce more heat tolerant chickens. In addition, Tamzil *et al*. (2013) found that the polymorphism of *hsp70* exhibited in native chickens had the highest heat resistance as compared to commercial breeds. According to our results, KU–Phuphan black–bone chickens might adapt in heat stress environment better than commercial chickens. Although, many researchers studied on the identification of *hsp*70 gene polymorphism and they found polymorphic sites that located upstream of the coding region (Mazzi *et al*., 2003; John *et al*., 2012; Abdolalizadeh *et al*., 2015; Gan *et al*., 2015). For our study, while the locus *C* was a polymorphic site, the locus *M* showed monomorphism for FB, BR and LA populations, containing one allele of M_2 (360 bp) (Figure 2). Only the MB chickens displayed both allele M_1 (229 bp) and allele M_2. The allele distributions of *EaeI* and *MmeI* cut of four chicken lines were displayed in Figure 3 and Figure 4, respectively. However in our study, allele M_1 was observed that only 4% of the population was not effective for the selection program whereas other Thai indigenous chickens were found 10–32% of populations (Tunim *et al*., 2010). This result might be caused from the selection process. Many alleles can be lost in the selection process due to random genetic drift, which is a consequence of finite populations and is proportional in the rate of inbreeding (Wolc *et al*., 2015). Therefore, the locus *C* seems effective to use as selection marker better than locus *M*.

Considering the genotype combination of locus *C* and locus *M*, our result found four possible genotypes of C_2C_2/M_1M_2, C_1C_1/M_2M_2, C_1C_2/M_2M_2 and C_2C_2/M_2M_2. Duangduen *et al*. (2008) detected effect of *hsp*70 genotype on thermotolerance in the native and broiler chickens. They reported that chickens containing C_1C_2/M_2M_2 genotype showed better heat tolerance performance than other chickens consisting

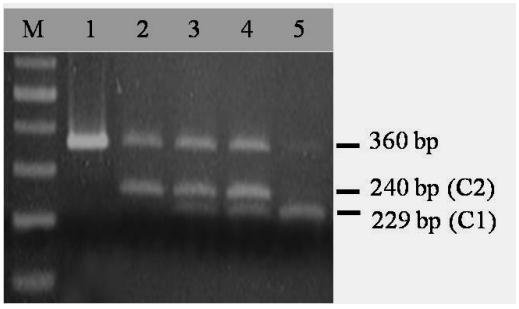


Figure 1 Alleles distribution of *hsp*70 gene from PCR–RFLP cut by *Eae*I in 3% agarose gel. M is 100 bp ladder standard DNA marker, 1 is uncut 360 bp amplified *hsp*70 gene, 2 is 240 bp of C_2 allele, 3 and 4 are 229 bp and 240 bp of allele C_1/C_2, respectively, and 5 is 229 bp of C_1 allele**.**

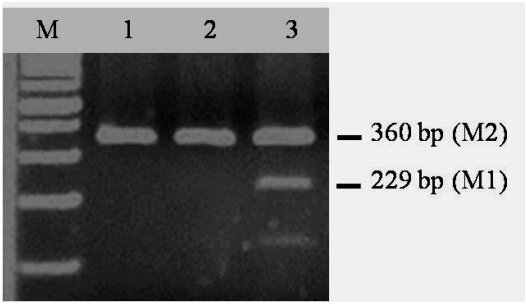


Figure 2 Alleles distribution of *hsp*70 gene from PCR–RFLP cut by *Mme*I in 3% agarose gel. M is 100 bp ladder standard DNA marker, 1 is uncut 360 bp amplified *hsp*70 gene, 2 is 360 bp of M_2 allele, 3 is 229 bp and 360 bp of allele M_1/M_2, respectively.

Table 1 Allelic frequencies, genotypic frequencies and Hardy-Weinberg Equilibrium (HWE) test of *hsp*70 gene observed for each breed [n = 30 for female KU–Phuphan black bone chickens (FB), n = 25 for male KU–Phuphan black bone chickens (MB), n = 23 for crossbreed native chickens (BR) and n = 23 for commercial layer chickens (LA)] using Genepop (v 1.2) (Raymond and Rousset, 1995).

Frequency		Chicken lines			
		FB	**MB**	**BR**	**LA**
Allele frequency	C_1	0.48	0.56	0.43	0.59
	C_2	0.52	0.44	0.57	0.41
	M_1	0.00	0.02	0.00	0.00
	M_2	1.00	0.98	1.00	1.00
Genotype frequency	C_1C_1	0.23	0.36	0.30	0.39
	C_1C_2	0.50	0.40	0.26	0.39
	C_2C_2	0.27	0.24	0.44	0.22
	M_1M_2	0.00	0.04	0.00	0.00
	M_2M_2	1.00	0.96	1.00	1.00
Combination genotype frequency	C_2C_2/M_1M_2	0.00	0.04	0.00	0.00
	C_1C_1/M_2M_2	0.23	0.36	0.30	0.39
	C_1C_2/M_2M_2	0.50	0.40	0.26	0.39
	C_2C_2/M_2M_2	0.27	0.20	0.44	0.22
HWE	*p*-value	1.00	0.41	0.03*	0.40

*$p < 0.05$

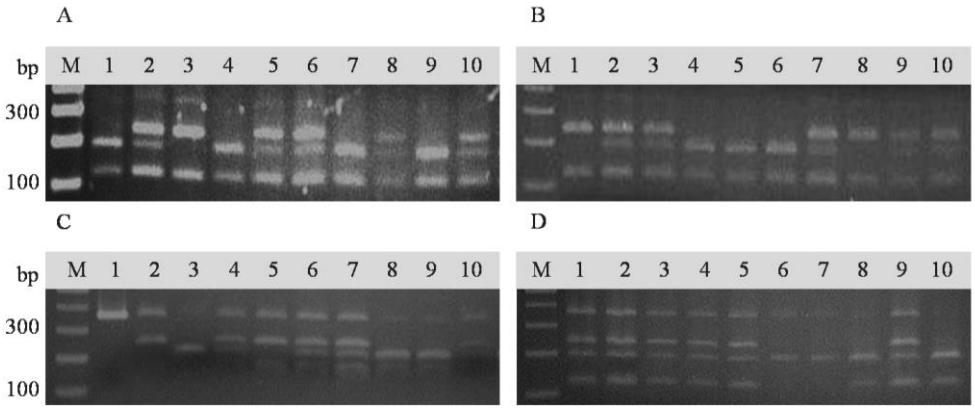


Figure 3 Some genotypes of *hsp*70 gene of four chicken lines derived from PCR–RFLP cut by *Eae*I in 3% agarose gel. A is a male KU–Phuphan black bone chicken, B is a female KU–Phuphan black bone chicken, C is a crossbreed native chicken, and D is a commercial layer line. M is 100 bp ladder standard DNA marker. 1–10 are the samples of each chicken population.

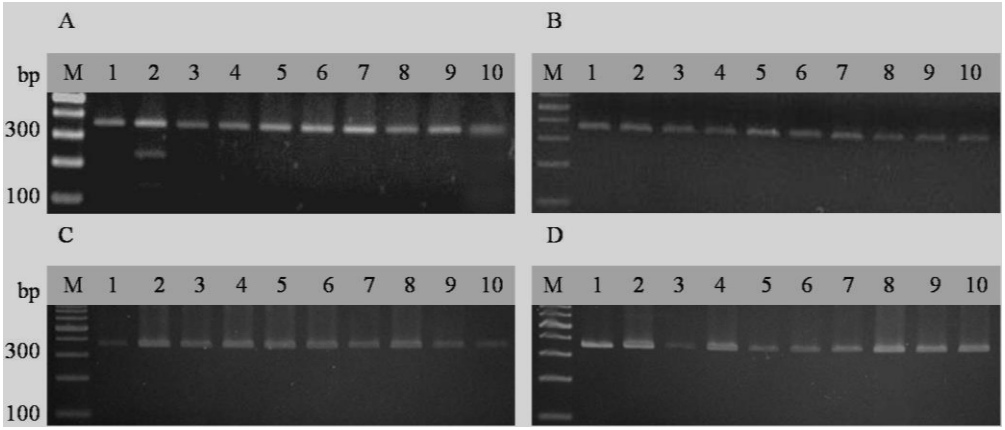


Figure 4 Some genotypes of *hsp*70 gene of four chicken lines derived from PCR–RFLP cut by *Mme*I in 3% agarose gel. A is a male KU–Phuphan black bone chicken, B is a female KU–Phuphan black bone chicken, C is a crossbreed native chicken, and D is a commercial layer line. M is 100 bp ladder standard DNA marker. 1–10 are the samples of each chicken population.

with other genotypes. In addition, chickens presented the genotype of C_2C_2/M_1M_2 was sensitive to heat stress (Duangduen *et al*., 2008) which should be considered for the breeding program. In our study, female and male KU–Phuphan black–bone chicken populations showed the high C_1C_2/M_2M_2 genotype frequency (FB was 0.5 and MB was 0.4). This combination genotype showed the higher genotype frequency than other three genotypes (Table 1). The female and male KU–Phuphan black–bone chickens exhibited the higher genotype frequency of C_1C_2/M_2M_2 than various strains of Thai native chickens (Pradu Hangdam = 0.48, Lueng Hangkhao = 0.34, Dang = 0.30 and Chee = 0.28) reported by Tunim *et al*. (2010). These results indicated that KU–Phuphan black-bone chicken populations showed good performance at high ambient temperatures to tolerate from heat stress. Moreover, the polymorphic locus *C* and C_1C_2/M_2M_2 genotype of *hsp*70 could be used as candidate markers to select the heat tolerant chickens.

Genetic diversity of *hsp*70

Locus *M* showed monomorphism for most of the populations. A reduction of heterozygosity could occur during the selection and inbreeding process (Akaboot *et al*., 2012), and resulted in a loss of alleles of *hsp*70 that could affect to the adaptation to heat stress of chickens. Therefore, the utilization from locus *M* was carefully considered for our chicken populations, especially for KU–Phuphan black–bone chickens. The observed (H_o) and the expected (H_e) heterozygosities were detected from all populations for locus *C* but locus *M* was determined only the MB population. At locus *M*, the H_o and the H_e of the MB population were 0.04. At locus *C*, the H_o varied from 0.21 to 0.50 and H_e ranged from 0.48 to 0.51 (Table 2). The H_o of BR was lower than other breeds (H_o=0.26). The highest value of H_o and H_e were found in female KU–Phuphan black–bone chicken population which were 0.50 and 0.51, respectively. The expected heterozygosity at locus *C* of all chicken populations was not different from the H_e of the other Thai native breeds (Pradu Hangdam = 0.500, Lueng Hangkhao = 0.502 and Dang = 0.501) reported by Tunim *et al.* (2010) (H_e = 0.398 to 0.502) and also was not different from the H_e of Pradu Hangdam (H_e = 0.49) reported by Akaboot *et al*. (2012). However, although FB and MB showed high value of the H_e but the inbreeding coefficient within population (F_{IS}) was detected. All populations remaining F_{IS} values were positive, suggesting that inbreeding was common in almost all chicken populations. The F_{IS} was 0.016, 0.208, 0.486 and 0.214 for FB, MB, BR and LA populations, respectively. Although the F_{IS} of both KU–Phuphan black–bone chicken populations was detected the low level of inbreeding, a need for genetic intervention, for examples, exchange of breeding cocks among flocks to avoid the possible negative impact of inbreeding depression on survival, reproduction and productivity, should be considered in the KU–Phuphan black–bone chickens.

Table 2 Observed and expected heterozygosity, and Inbreeding coefficient (F_{IS}) of *hsp*70 gene for each breed [n = 30 for female KU-Phuphan black bone chickens (FB), n = 25 for male KU-Phuphan black bone chickens (MB), n = 23 for crossbreed native chickens (BR) and n = 23 for commercial layer chickens (LA)] using FSTAT (v 2.9.3.2) (Goudet, 2002).

Loci	Heterozygosity	Chicken lines			
		FB	MB	BR	LA
C	Observed	0.50	0.40	0.26	0.39
	Expected	0.51	0.50	0.48	0.48
M	Observed	0.00	0.04	0.00	0.00
	Expected	NA	0.04	NA	NA
F_{IS}		0.016	0.208	0.486	0.214

NA is Not Available.

In conclusion, Chen *et al*. (2016) found that the use of molecular breeding method to identify genetic markers related to thermotolerance could allow the possibility of direct gene selection. According to our study, both the KU–Phuphan black–bone chickens exhibited genotype related with heat tolerance. The polymorphisms in the *hsp70* gene might be used by the breeders to produce chickens that are more tolerant to high temperature.

ACKNOWLEDGEMENTS

The author would like to thank the animal farm and KU–Phuphan black–bone chicken project of Faculty of Natural Resources and Agro-Industry for

supplying blood samples. This research supported by Faculty of Natural Resources and Agro-Industry, Kasetsart University, Chalermphrakiat Sakonnakhon Province Campus, Thailand.

REFERENCES

Abdolalizadeh N, Noshary A, Eila N (2015) Identification of single nucleotide polymorphisms of Hsp70 gene in a commercial broiler strain. Res Opin Anim Vet Sci 5: 265–269.

Akaboot P, Dungjinda M, Phasuk Y (2010) Genetic comparison of functional genes in Red jungle fowl, Thai native chicken and commercial chicken. Proceeding the 14th AAAP Animal Science Congress. August 23–27, 2010. Pingtung, Taiwan, ROC.

Akaboot P, Duangjinda M, Phasuk T, Kaenchan C, Chinchiyanond W (2012) Genetic characterization of Red Junglefowl (*Gallus gallus*), Thai indigenous chicken (*Gallus domesticus*), and two commercial lines using selective functional genes compared to microsatellite markers. Genet Mol Res 11: 1881–1890

Al-Aqil A, Zulkifli I (2009) Changes in heat shock protein 70 expression and blood characteristics in transported broiler chickens as affected by housing and early age feed restriction. Poult Sci 88: 1358–1364.

Beckham JT, Wilmink GJ, Mackanos MA, Takahashi K (2008) Role of *HSP*70 in cellular thermotolerance. Lasers Surg Med 40: 704–715.

Boonkum W, Duangjinda M, Laopaiboon B, Vongpralub T (2014) Effects of heat stress on genetic parameters and egg production in Pradu Hang Dam Thai native chickens. Khon Kaen Agr J 42: 319–328. (in Thai)

Chen ZY, Zhang WW, Gan JK, Kong LN, Zhang XQ, Zhang DX, Luo QB (2016) Genetic effect of an A/G polymorphism in the *HSP70* gene on thermotolerance in chicken. Genet Mol Res 15: gmr.15028271.

Cooper MA, Washburn KW (1998) The relationship of body temperature to weight gain, feed consummation, and feed utilization in broilers under heat stress. Poult Sci 77: 237–242.

Duangduen C, Duangjinda M, Katavatin S, Aeagwanich W (2008) Effect of hsp70 Genotypes on Thermotolerance in Thai Native Chicken and Broilers. Proceedings of the 4th Symposium on Thermotolerance in Domestic Animals. Khon Kaen, pp 241–245.

Frankham R, Ballou JD, Briscoe DA (2002) Introduction to Conservation Genetics. Cambridge University Press, Cambridge, UK.

Gan JK, Jiang LY, Kong LN, Zhang XQ, Luo QB (2015) Analysis of genetic diversity of the heat shock protein 70 gene on the basis of abundant sequence polymorphisms in chicken breeds. Genet Mol Res 14: 1538–1545.

Gaviol HCT, Gasparino E, Prioli AJ, Soares MAM (2008) Genetic evaluation of the HSP70 protein in the japanese quail (*Coturnix japonica*). Genet Mol Res 7(1): 133–139.

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3) http://www2.unil.ch/popgen/softwares/fstat.htm. (September 2016).

Hung YC, Sava VM, Makan SY, Chen THJ, Hong MY, Huang GS (2002) Antioxidant activity of melanins derived from tea: comparison between different oxidative states. Food Chem 78: 233–240.

John L, Panicker VP, George SA (2012) Polymorphism analysis of Hsp70 gene in chicken and duck. Indian J Poult Sci 47: 129–135.

Mazzi CM, Ferro JA, Ferro MIT, Savino VJM, Coelho AAD, Macari M (2003) Polymorphism analysis of the hsp70 stress gene in Broiler chickens (*Gallus gallus*) of different breeds. Genet Mol Biol 26: 275–281.

Raymond M, Rousset F (1995) GENEPOP (version 1.2) population genetics software for exact tests and ecumenicism. J Hered 86: 248–249.

Tamzil MH, Noor RR, Hardjosworo PS, Manalu W, Sumantri C (2013) Acute heat stress responses of three lines of chickens with different heat shock protein (*HSP*) –70 genotypes. Int J Poult Sci 12: 264–272.

Tunim S, Duangjinda M, Katavatin S (2010) Study of *HSP70* gene polymorphism in various strains of Thai indigenous chickens. Khon Kaen Agr J 38 Suppl: 71–75. (in Thai)

Wolc A, Zhao HH, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, Stricker C, Habier D, Fernando RL, *et al*. (2015) Response and inbreeding from a genomic selection experiment in layer chickens. Genet Select Evol (47): 59. https://doi.org/10.1186/s12711-015-0133-5

Zhang X, Du H, Li J (2002) Single nucleotide polymorphism of chicken heat shock protein 70 gene. 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Expression Analysis of Rice Polygalacturonase cDNA Responding to Brown Planthopper [*Nilaparvata lugens* (Stål)]

Sugunya Suebsan[1*], Supranee Sitthiphrom[3], Kanta Sangwijit[2], Mondhon Sanguansermsri[4], Somboon Anuntalabhochai[1,2]

[1]*Biology Department, School of Science, University of Phayao, Phayao, Thailand*
[2]*Biotechnology Unit, University of Phayao, Phayao, Thailand*
[3]*Faculty of Science and Technology, Loei Rajabhat University, Loei, Thailand*
[4]*Pharmaceutical Sciences Department, School of Pharmaceutical Sciences, University of Phayao, Phayao, Thailand*

[*] *Corresponding author: spitakrattana@gmail.com*

ABSTRACT

A cDNA (*OsKPG*) encoding polygalacturonase (PG) from rice (*Oryza sativa* cv. KDML105) was cloned and sequenced. The cDNA full length was 1,103 bp and carried 277 deduced amino acids with 4 highly conserved domains among PG family. The expression of *OsKPG* was investigated under brown planthopper [*Nilaparvata lugens* (Stål)] attack. The transcription levels of *OsKPG* under different phytohormone applications including ethylene, abscisic acid (ABA), 6-benzyladenine (6-BA) and 2, 4-dichlorophenoxyacetic acid (2, 4-D) were also analyzed in leaves. Semi-quantitative RT-PCR revealed that the expression of *OsKPG* was high upon the brown planthopper attack, or under the application of ethylene or 6-BA treatments. This is the first evidence demonstrating that *OsKPG* may play a role in response against insect herbivore infestation through ethylene and cytokinin signaling pathways in rice.

Keywords: polygalacturonase; *Oryza sativa* cv. KDML105; insect herbivore; ethylene; cytokinin

INTRODUCTION

Plants and insects have interacted for more than 350 million years, leading plants to develop an elegant defense system that has the ability to recognize foreign molecules, signals from damaged cells, and activate the plants immune response against the herbivores (War *et al.*, 2012). The signaling pathways that enable plants to mount defenses against insect herbivores are known to be complex. Some defenses are constitutive while others are inducible. Constitutive defenses are present and offer continuous protection against herbivores such as furanocoumarin, saponin and cardenolide (Wittstock and Gershenzon, 2002). Some intracellular signals in wounded tissues induced by herbivores are jasmonic acid (JA); salicylic acid (SA) and ethylene (Fürstenberg-Hägg *et al*., 2013).

Ethylene is known as a major signal in plants that can induce defense-related genes, leading to produced defense-related proteins. For example, ethylene can induce chitinases accumulation in rice (Rakwal *et al*., 2004). The activity of peroxidase in barley is increased by ethylene application as well as infestation by 2 aphids; *Schizaphis graminum* (biotype C) and *Rhopalophum padi*. It has been suggested that ethylene is involved in the oxidative responses of barley plants induced by infestation (Argandoña *et al*., 2001). Ethylene can also regulate the cysteine proteinase (*mir-1*) expression in maize (*Zea mays* L.) genotype Mp708, leading to defense against insect herbivore attack (Harfouche *et al.*, 2006).

Polygalacturonases (PGs) have been known to play a major role in many processes of plants such as seed germination (Sitrit *et al*., 1999), cell elongation and flower development (Xiao *et al*., 2014); senescence and abscission of leaves (Lee *et al*., 2001) and fruit ripening (Gayathri and Nair, 2015). Evidence has shown that plant endogenous PG induced by mechanical damage could degrade pectin in plant cell walls to release oligogalacturonides (OGAs) (Orozco-Cardenas and Ryan, 1999). The OGAs have been known to act as elicitor molecules that trigger a variety of plant responses against pathogens and insects (Shibuya and Minami, 2001). These responses included the accumulation of phytoalexins (Davis *et al*., 1986), glucanase, and chitinase (Davis and Hahlbrock, 1987; Broekaert and Pneumas, 1988), the production of reactive oxygen species (ROS) (Galletti *et al*., 2008) and nitric oxide (Rasul *et al*., 2012). This evidence suggests that PG may be a key molecule to activate defense responses during herbivore and pathogen attacks (Bergey *et al*., 1999). In addition, PG is one of several enzymes induced by ethylene. For examples, the expression of PG gene in cucumber fruit (*CUPG1*) is induced by water stress (water loss after harvesting)

and exogenous ethylene, but not by the application of abscisic acid (ABA) (Kubo *et al*., 2000); PG gene isolated from papaya (*cpPG*), which controls the process of pulp softening during papaya ripening, is strongly induced during ripening and highly ethylene-dependent (Fabi *et al*., 2006). Moreover, the PG in ethylene-stimulated abscission of tomato pedicel shows an abundant accumulation in the cortical and vascular tissues in the abscission zone at 8 hours after ethylene treatment (Qi *et al*., 2014).

Despite much information on PG, reports about the relationship between phytohormones and PG in the herbivore-defense mechanism are rare. This study is the first report to show that *OsKPG* is involved in herbivore attack, probably, under ethylene and cytokinin signaling in rice.

MATERIALS AND METHODS

Plant material

Seeds of rice (*Oryza sativa* cv. KDML105) were submerged in water for 2 days and germinated in a moisture chamber. After 5 days, the seedlings were transferred into plastic pots (9 cm in diameter) containing soil and grown in a greenhouse under natural light conditions. The rice seedlings at 10 day old were used for *OsKPG* cloning and expression analysis under insect attack. The brown planthopper [*Nilaparvata lugens* (Stål)] (BPH) used for infestation were collected from rice fields in Phayao province, Thailand and fed on tillers of a susceptible rice variety "Taichung Native 1" (TN1) for mass propagation. The 40 day seedlings were chosen for expression analysis under phytohormone treatments.

Cloning of a full-length *OsKPG* cDNA from rice

First-strand cDNA was synthesized using Superscript III Reverse transcriptase (Invitrogen, USA) following the manufacturer's instructions. The first-strand cDNA was used as template for PCR amplification. The partial cDNA sequence encoding the OsKPG protein was obtained using degenerate forward primer which was designed to correspond to the highly conserved N terminal region of the PG proteins. The forward primer, OsKPGFD, was designed to match a sequence coding for domain I of PG family. A poly-T oligonucleotide, oligo $(dT)_{18}$ was used as a reverse primer. The amplification conditions were an initial denaturation at 94ºC for 2 min, followed by 35 cycles at 94ºC for 45 s, annealing at 55ºC for 30 s, extension at 72ºC for 1 min, and a final extension at 72ºC for 10 min. The DNA fragments were ligated into the pTZ57R vector (Fermentas, Germany). DNA sequencing was performed using the dideoxynucleotide chain termination method (Sanger *et al*., 1977) using an automated sequencer (ABI). The amino acid sequence similarity was analyzed using the BLAST program (Altschul *et al*., 1990). The 5' terminal of the *OsKPG* cDNA sequences were obtained using 5' Full Core Set (Takara, Japan) following the manufacturer's instructions. The primers used for partial cDNA cloning and 5' RACE PCR are shown in Table 1.

Sequence analysis

The complete sequence of *OsKPG* cDNA was translated into peptide sequence. The peptide sequence was aligned with 11 PG protein sequences from 10 plant species [*Oryza sativa japonica* (XP_015623329), *Oryza sativa indica* (EAY87758), *Setaria italic* (XP_004954128), *Brachypodium distachyon* (XP_003570492), *Zea mays* (NP_001140630), *Triticum urartu* (EMS61824.1), *Sorghum bicolor* (XP_002452925), *Phoenix dactylifera* (XP_008784338), *Ananas comosus* (OAY70358), *Aegilops tauschii* (EMT11866) and *Lycopersicon esculentum* (Bergey *et al.*, 1999)] using Clustal W (Goujon *et al*., 2010) and the initial phylogenetic tree was then bootstrapped 1,000 times using the UPGMA method, which was included in the MEGA5 software package (Tamura *et al*., 2011).

Table 1 The sequences of PCR primer sets used for partial cDNA cloning and 5′ RACE PCR

Primers	Sequences
OsKPGFD	5′-GCNCCNAAYACNGAYGGNATHCCN-3′
Oligo $(dT)_{18}$	5′-TTTTTTTTTTTTTTTTTT- 3′
OsKPG_deFD	5′-GCNCCNAAYACNGAYGGNATH- 3′
OsKPGRT-p	5′-ATCGTGGATGTAGCC-Ⓟ- 3′
OsKPG_A1	5′-CACAATGTTACGAGATGGGC- 3′
OsKPG_A2	5′-ACCGCTATGGCATCATCACC- 3′
OsKPG_S1	5′-CGTTCATTGGTCAGTGCTGG- 3′
OsKPGS2	5′-GCAGTGAGATGTCTGGTGGG- 3′
OsKPG_3'FD	5′-CAAGACCTCCTCTTTTGCAGC- 3′
OsKPG_3'RV	5′-ATGAGAATGGGAATGCCACCC- 3′

Rice infestation procedure

The 10 day old seedlings were thinned to 10 plants per pot. All pots were transferred into the 50 x 40 x 50 (W x L x H) cm cages covered with a nylon net (32 holes/cm^2). To provide suitable humidity for insect survival, the experiment was conducted at a temperature of 28ºC to 30ºC and kept relatively high humidity at 70% to 80%. The rice seedlings were infested with 2nd – 3rd instar of brown planthopper at a density of 8 to 10 per seedling. The seedling without insect infestation was used as a control. Five seedlings were harvested after 5, 10 and 15 days of herbivore infestation and stored at -80ºC.

Phytohormone treatment

The fourth leaf blades from 40 day old rice plants were chopped to 2 to 3 cm. The chopped leaf blades were incubated in 3 mM MES buffer, pH 5.8, supplemented with various phytohormones including 1 mM ethephon, 100 μM of abscisic acid (ABA), 100 μM 6-benzyladenine (6-BA) and 100 μM 2,4-dichlorophenoxyacetic acid (2,4-D) to examine the effect of these phytohormones (Pitakrattananukool *et al*., 2012). Three leaves from each experiment were collected at 0, 6, 12 and 24 hours after incubation. All experiments were performed under continuous lighting.

Semi-quantitative reverse transcriptase PCR reaction and analysis

Total RNA was extracted from the rice tissues using Trizol reagent (Invitrogen, USA) and first-strand cDNA was synthesized. Twenty ng of first-strand cDNA was used as the PCR template. The amplified PCR product (237 bp) was obtained using *OsKPG* specific primers (Table 2). The RT-PCR reaction was performed by initiating for template denaturation for 2 min at 94ºC, followed by 28 cycles of denaturation for 30 s at 94ºC, annealing for 30 s at 62ºC and extension for 30 s at 72ºC. Rice *β-Actin* transcripts were used as the internal standards. The primers used for RT-PCR are shown in Table 2. Agarose electrophoresis was performed to visualize the PCR products. To monitor the relative level of gene expression, the intensities of the band were analyzed using Scion Image software (Scion, Frederick, MD). The level of *OsKPG* expression was presented as the ratio of expression of *OsKPG* to *β-Actin*.

Table 2 The sequences of PCR primer sets used for RT-PCR

Primers	Sequences
PGRT-FD	5′-TCACGGCGAGGTTTGAGAATAAAGACTGCCAT-3′
PGRT-RV	5′-ACTGCTGCCATGAGCACGGACTGGCACCCGGACA-3′
βActin-FD	5′ - CAAGGCCAATCGTGAGAAG - 3′
βActin-RV	5′ - AGCAATGCCAGGGAACATA - 3′

RESULTS

Cloning of full-length *OsKPG* cDNA from rice (KDML105)

The partial sequence (705 bp) of *PG* cDNA from rice (*Oryza sativa* cv. KDML105) was cloned. Then the full length of *PG* cDNA was amplified using 5′ RACE PCR technique. The 1,103 bp full-length cDNA was cloned into pTZ57R and was analyzed. This sequence contained a 26 bp 5′ noncoding region, 243 bp 3′ noncoding region and translated into 277 amino acids. The four conserved domains found in higher plant and fungi PG sequences were observed in OsKPG sequence (domain I-IV). Domain I (NTD) analyzed as substrate-binding region. Domain II (G/QDD) had a carboxylate group in the three aspartic acids that may be a component of the catalytic site. Domain III (G/SHG) contained the histidine residue that is thought to participate to the catalytic reaction. Domain IV (RIK) was thought to be as a substrate-binding region (Bussink *et al*., 1991; Rao *et al*., 1996; Palanivelu, 2006). Sequence analysis is shown in Figure 1.

Phylogenetic analysis of the OsKPG

The deduced amino acid sequence of OsKPG was analyzed by alignment among members of PG sequences from 10 plant species using Clustal W and phylogenetic tree was constructed by UPGMA method. Comparative analysis revealed the OsKPG displayed high similarity to PGs members, *Oryza sativa japonica* (XP_015623329.1) and *indica* (EAY87758.1) at 99%, and to other monocot plant species such as *Setaria italic* (XP_004954128.1), *Brachypodium distachyon* (XP_003570492.1), *Zea mays* (NP_001140630.1), *Triticum urartu* (EMS61824.1) and *Sorghum bicolor* (XP_002452925.1) at 89, 88, 87, 86 and 86% respectively. OsKPG showed low homology at 28.4% identity to PG from dicot specifically one induced by wounding in *Lycopersicon esculentum* (Bergey *et al*., 1999) (Figure 2).

*ATACAAGACCTCCTCTTTTGCAGCTG***ATG**TGGTCCAAGGACATTATTGTTGCAAATATAA
M W S K D I I V A N I T
CATTGAAGAATTCACCTTTCTGGCACTTCCACCCATATGATTGCACGAATATAACTGTTT
L K N S P F W H F H P Y D C T N I T V S
CTAATGTTACTATCTTAGCTCCTATTTCTAGTGCTCCAAACACAGATGGCATAGATCCAG
N V T I L A P I S S A P N T D G I D P D
I
ATTCTTGTCAGGATGTGCTTATTGAGAATTGCTACATTTCAGTTGGTGATGATGCCATAG
S C Q D V L I E N C Y I S V G D D A I A
II
CGGTAAAGAGTGGGTGGGATCAATATGGGATTGCATATGGGCGCCCATCTCGTAACATTG
V K S G W D Q Y G I A Y G R P S R N I V
TGATACGCAATGTAATGGCTCGTTCATTGGTCAGTGCTGGAATTTCAATTGGCAGTGAGA
I R N V M A R S L V S A G I S I G S E M
III
TGTCTGGTGGGATTGCAAATGTTACAGTGGAGGATGTCCGCATTTGGGAGTCACGGCGAG
S G G I A N V T V E D V R I W E S R R G
GTTTGAGAATAAAGACTGCCATAGGAAGAGGGGGCTACATCCGCGATATCTCCTATCGCA
L R I K T A I G R G G Y I H D I S Y R N
IV
ACATAACCTTCGACAATGTCCGTGCTGGTATTGTGATAAAGGTTGACTACAATGAGCACG
I T F D N V R A G I V I K V D Y N E H A
CTGATGATGGGTATGACCGGGATGCCTTTCCAGACATCACAAACATATCATTCAAGGAAA
D D G Y D R D A F P D I T N I S F K E I
TACATGGGCGAGGTGTCCGGGTGCCAGTCCGTGCTCATGGCAGCAGTGACATTCCCATCA
H G R G V R V P V R A H G S S D I P I K
AGGACATCAGCTTTCAGGACATGTCTATCGGCATCAGCTACAAGAAGAAACATATTTTCC
D I S F Q G M S I G I S Y K K K H I F Q
AGTGTTCCTTCATTGAGGGGCGTGTCATCGGGTCAGTGTTTCCAAAACCATGCGAGAATT
C S F I E G R V I G S V F P K P C E N L
TGGATCTCTACAATGAGCAAGGGCAGCTTGTTAAGCGTGCAGCAATGGTAAACAGCACGG
D L Y N E Q G Q L V K R A A M V N S T E
AAGTTGATTATGACATA**TGA***TGATATAGGTGGCTAAGGCTAACAGCAGAACATGGGTGGC*
V D Y D I stop
ATTCCCATCCTCATTTTGTTCTTTTAGTCTTTCTGTGTTTTTTGAGATTAGAACTGTAT
ATACATAGGAGATTCACCTCTTCATAGGTTTGAATGCAGTGGCTAATTGTATCTCTCATT
TCATTTGTCCACCAGCATTTTGTTCTTGCTGCCAAACATGTAATACAACCTTGCAATCTT
AGAAATACACTTGGCCATTCACC

Figure 1 Nucleotide sequence and deduced amino acid sequence of *OsKPG*. Bold letters indicate start and stop codons, while the 5′ and 3′ UTR are indicated in italics. The predicted amino acid sequence is shown below the nucleotide sequence in single-letter code. The four functional domains of PG are underlined and Roman numerals indicate the substrate-binding domains (I, IV) and catalytic domains (II, III).

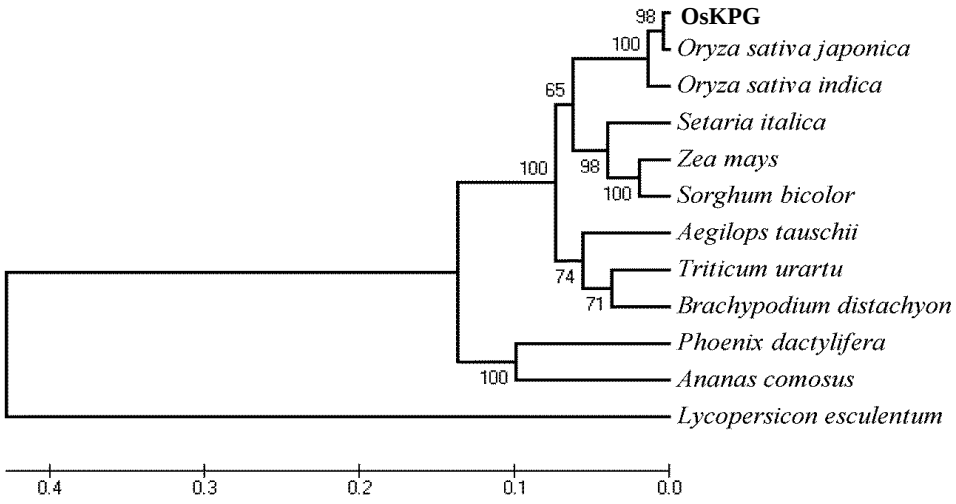


Figure 2 The phylogenetic analysis of the relationship among OsKPG amino acid sequence and amino acid sequences of PG from 10 different plant species. The numbers at an internal node show the bootstrapping value.

Expression profile of the *OsKPG* under brown planthopper infestation

To examine *OsKPG* expression under insect infestation, 2nd - 3rd instar of *Nilaparvata lugens* (Stål) was introduced to 10 day old rice seedlings in the cage, as mentioned in material and methods, for 15 days. Then, the seedlings were collected to evaluate the *OsKPG* mRNA intensities (Figure 3A). RT-PCR revealed that the *OsKPG* expression in rice seedlings attacked with brown planthopper had increased in day 10 and 15 by 1.3 and 1.5 fold respectively when compared to the seedlings without insect attack (Figure 3B).

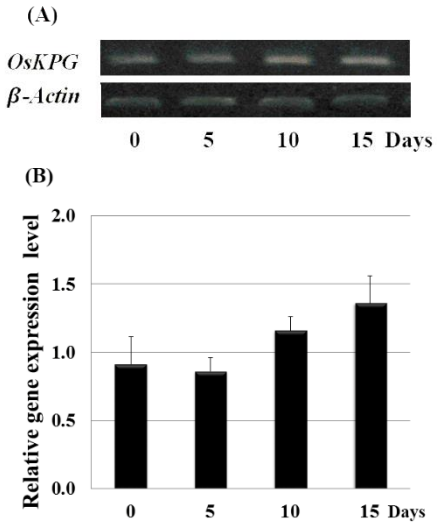


Figure 3 The expression of *OsKPG* under herbivore infestation. The expression of *OsKPG* after infestation with brown planthopper for 5, 10 and 15 days (A). The relative expression of *OsKPG* under herbivore infestation, presented as the ratio of expression of *OsKPG/β-Actin* (B). *β-Actin* was used as the internal control.

Expression profile of the *OsKPG* under Phytohormone treatments

The *OsKPG* transcription responded to 4 phytohormones, i.e. ethylene, ABA, 6-BA and 2,4-D was determined in rice leaves. The RT-PCR reactions revealed that the expression of *OsKPG* responded to ethylene and 6-BA showing a 2.0 and 4.0 fold upregulation in rice leaves exposed to 1 mM of ethephon and 100 µM of 6-BA compared with untreated leaves at the beginning of experiment, respectively (Figure 4).

DISCUSSION

Polygalacturonase (*PGs*) was known as a super-family gene for pectin dehydration. There were many *PGs* isolated from numerous organisms. In the case of plants, plants utilize PGs in many processes such as the growth and development processes as well as defending mechanism (Bergey *et al*., 1999). In this work *PG* cDNA from rice *O. sativa* (KDML105) named *OsKPG* was isolated and analyzed. Comparisons of deduced amino acid sequences of full-length *OsKPG* cDNA and 11 plant PG members revealed that *OsKPG* displayed the highest identity (99%) to *O. sativa japonica* and *indica* groups and 98 – 86% identity to PG from other monocot plant species: *S. italic*, *B. distachyon*, *Z. mays*, *T. urartu* and *S. bicolor*. The four functional conserved domains (domain I-IV) of all PG members were also present, suggesting that *OsKPG* belonged to the PG family (Figure 2).

The expression of *OsKPG* in rice seedlings was analyzed after being infested by brown planthopper (Figure 3). RT-PCR revealed that the expression level of *OsKPG* increased up to 1.5 fold than that of intact seedling. Bergey *et al*. (1999) supported this with evidence that the activity of PG was induced in tomato leaves under wounding and herbivore attacks. This led to an increase in endogenous oligogalacturonide elicitor that may be involved in the local and systemic activation of defense responses against herbivores. Therefore, this evidence indicated that *OsKPG* was one of the defense response genes against herbivores in rice leaves.

Meanwhile, the expression of *OsKPG* under phytohormone treatments was investigated. The results revealed that *OsKPG* was upregulated in rice leaves exposed to ethylene and 6-BA to the factor of 2 and 4 respectively. It was not responsive to ABA and 2,4-D application (Figure 4). So far, ethylene and jasmonic acid were well known as the regulators of herbivore and pathogen responsive genes. Ethylene is a main mediator of the signal transduction pathway leading to defense against insect herbivores in maize and other plants (Harfouche *et al*., 2006; von Dahl and Baldwin, 2007; Louis *et al*., 2015). There have been reports that the expression of ethylene biosynthesis gene (*OsACS*) in rice strongly increases after rice is wounded or infested by brown planthopper (Lu *et al*., 2014). The stability of ethylene biosynthesis proteins was under regulation of the increase of cytokinins (Vogel et al., 1998). Moreover, there have been reports that both natural and synthetic cytokinin applications in cotton leaves resulted in increased ethylene production (Suttle, 1986) and some cytokinins showed response to wounding and insect oral secretion applications. These cytokinins were *isopentenyl adenine* (IP), *isopentenyl*

adenosine (IPR), cis-zeatin riboside (cZR), trans-zeatin riboside O-glucoside (tZROG), cis-zeatin riboside O-glucoside (cZROG) and trans-zeatin N7- glucoside (tZ7G). When insect attacked plant cell, herbivore-associated molecular patterns (HAMPs) presented in insect saliva were deposited to the plant cell. These compounds included glucose oxidase, alkaline phosphatase and other proteinaceous elicitors which activated plant defenses through a complex signaling network (Cheng *et al*., 2013). This suggests cytokinin is an integral component of wounding and HAMPs triggered responses in many plant species (Schäfer *et al*., 2015).

Our results clearly demonstrated that the expression of rice *OsKPG* responded to the attack of brown planthopper and to the exposure of ethylene or cytokinin treatment. Hereby, we propose that *OsKPG* may play a role in defense response against insect herbivores in rice under ethylene and cytokinin signal transduction.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support provided by Loei Rajabhat University and Research group development program, School of Science, University of Phayao, Thailand.

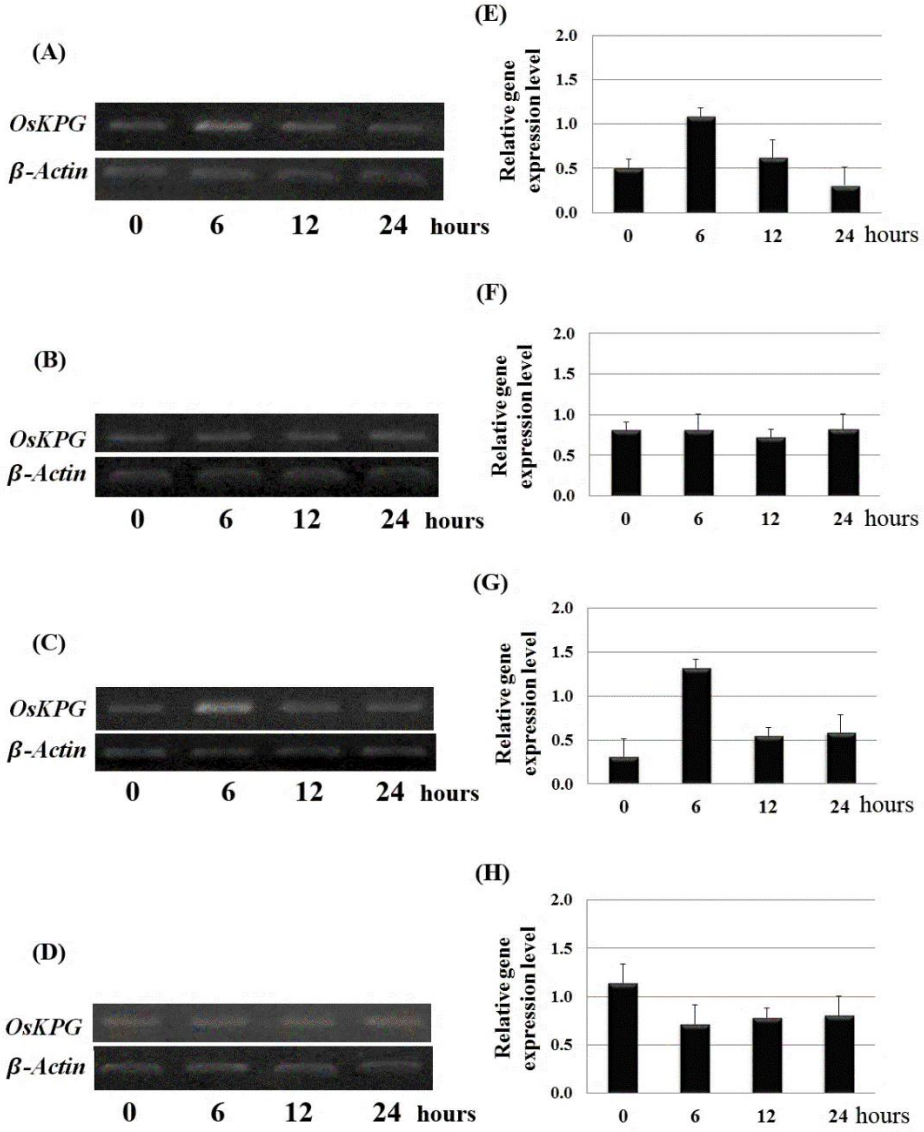


Figure 4 The expression of *OsKPG* under phytohormone treatments. The expression of *OsKPG* in rice leaves incubated in 3 mM MES buffer pH 5.8, supplemented with ethephon (A) and ABA (B), 6-BA (C) and 2,4D (D) for 0, 6, 12, and 24 hours. The relative expression level of *OsKPG* in those rice tissues is presented as the ratio of expression of *OsKPG* /β-Actin (E-H). *β-Actin* was used as an internal control.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–411.

Argandoña VH, Chaman M, Cardemil L, Muñoz O, Zúñiga GE, Corcuera LJ (2001) Ethylene production and peroxidase activity in aphid-infested barley. J Chem Ecol 27: 53–68.

Bergey D, Orozco-Cardenas M, De Moura DS, Ryan CA (1999) A wound- and systemin-inducible polygalacturonase in tomato leaves. Proc Natl Acad Sci USA 96: 1756–1760.

Broekaert WF, Pneumas WJ (1988) Pectic polysaccharides elicit chitinase accumulation in tobacco. Physiol Plant 74: 740–744.

Bussink HJD, Buxton FP, Visser J (1991) Expression and sequence comparison of the *Aspergillus niger* and *Aspergillus tubigensis* genes encoding polygalacturonase II. Curr Genet 19: 467–474.

Cheng X, Zhu L, He G (2013) Towards understanding of molecular interactions between rice and the brown planthopper. Mol Plant 6: 621–634.

Davis KR, Darvill AG, Albersheim P, Dell A (1986) Host–pathogen interactions. XXIX. Oligogalacturonides released from sodium polypectate by endopolygalacturonic acid lyase are elicitors of phytoalexins in soybean. Plant Physiol 80: 568–577.

Davis KR, Hahlbrock K (1987) Induction of defense responses in cultured parsley cells by plant cell wall fragments. Plant Physiol 84: 1286–1290.

Fabi JP, Cordenunsi BR, Seymour GB, Lajolo FM, do Nascimento JR (2006) Molecular cloning and characterization of a ripening-induced polygalacturonase related to papaya fruit softening. Plant Physiol Biochem 47: 1075–1081.

Fürstenberg-Hägg J, Zagrobelny M, Bak S (2013) Plant defense against insect herbivores. Int J Mol Sci 14: 10242–10297.

Galletti R, Denoux C, Gambetta S, Dewdney J, Ausubel FM, De Lorenzo G, Ferrari S (2008) The AtrbohD-mediated oxidative burst elicited by oligogalacturonides in *Arabidopsis* is dispensable for the activation of defense responses effective against *Botrytis cinerea*. Plant Physiol 148: 1695–1706.

Gayathri T, Nair AS (2015) Purification and characterization of polygalacturonase from ripened fruits of *Musa acuminata* cultivar from Kerala (*Musa acuminata* cv. Palayankodan). Food Measure 9: 233–239.

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R (2010) A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 38: W695–W699.

Harfouche AL, Shivaji R, Stocker R, Williams PW, Luthe DS (2006) Ethylene signaling mediates a maize defense response to insect herbivory. Mol Plant Microbe Interact 19: 189–199.

Kubo Y, Xue Y, Nakatsuka A, Mathooko FM, Inaba A, Nakamura R (2000) Expression of a water stress-induced polygalacturonase gene in harvested cucumber fruit. J Jpn Soc Hort Sci 69: 273–279.

Lee RH, Wang CH, Huang LT, Chen SG (2001) Leaf senescence in rice plants: cloning and characterization of senescence up-regulated genes. Short communication in J Exp Bot 52: 1117–1121.

Louis J, Basu S, Varsani S, Castano-Duque L, Jiang V, Williams WP, Felton GW, Luthe DS (2015) Ethylene contributes to maize insect resistance1-mediated maize defense against the phloem sap-sucking corn leaf aphid. Plant Physiol 169: 313–324.

Lu J, Li J, Ju H, Liu X, Erb M, Wang X, Lou Y (2014) Contrasting effects of ethylene biosynthesis on induced plant resistance against a chewing and a piercing-sucking herbivore in rice. Mol Plant 7: 1670–1682.

Orozco-Cardenas M, Ryan CA (1999) Hydrogen peroxide is generated systemically in plant leaves by wounding and systemin via the octadecanoid pathway. Proc Natl Acad Sci USA 96: 6553–6557.

Palanivelu P (2006) Polygalacturonase: Active site analyses and mechanism of action. Indian J Biotechnol 5: 148–162.

Pitakrattananukool S, Kawakatsu T, Anuntalabhochai S, Takaiwa F (2012) Overexpression of *OsRab7B3*, a small GTP-binding protein gene, enhances leaf senescence in transgenic rice. Biosci Biotechnol Biochem 76: 296–302.

Qi MF, Xu T, Chen WZ, Li TL (2014) Ultrastructural localization of polygalacturonase in ethylene-stimulated abscission of tomato pedicel explants. Sci World J 2014: 389896, doi: 10.1155/2014/389896.

Rakwal R, Yang G, Komatsu S (2004) Chitinase induced by jasmonic acid, methyl jasmonate, ethylene and protein phosphatase inhibitors in rice. Mol Biol Rep 31: 113–119.

Rao MN, Kembhavi AA (1996) Pant A. Implication of tryptophan and histidine in the active site of endo-polygalacturonase from *Aspergillus ustus*: elucidation of the reaction mechanism. Biochim Biophys Acta 1296: 167–173.

Rasul S, Dubreuil-Maurizi C, Lamotte O, Koen E, Poinssot B, Alcaraz G, Wendehenne D, Jeandroz S (2012) Nitric oxide production mediates

oligogalacturonide-triggered immunity and resistance to *Botrytis cinerea* in *Arabidopsis thaliana*. Plant Cell Environ 35: 1483–1499.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467.

Schäfer M, Meza-Canales ID, Navarro-Quezada A, Brütting C, Vanková R, Baldwin IT, Meldau S (2015) Cytokinin levels and signaling respond to wounding and the perception of herbivore elicitors in *Nicotiana attenuate.* J Integr Plant Biol 57: 198–212.

Shibuya N, Minami E (2001) Oligosaccharide signaling for defense responses in plant. Physiol Mol Plant Pathol 59: 223–233.

Sitrit Y, Hadfield A, Bennett AB, Bradford KJ, Downie AB (1999) Expression of polygalacturonase associated with tomato seed germination. Plant Physiol 121: 419–428.

Suttle JC (1986) Cytokinin-induced ethylene biosynthesis in nonsenescing cotton leaves. Plant Physiol 82: 930–935.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.

Vogel JP, Schuerman P, Woeste K, Brandstatter I, Kieber JJ (1998) Isolation and characterization of *Arabidopsis* mutants defective in the induction of ethylene biosynthesis by cytokinin. Genetics 149: 417–427.

von Dahl CC, Baldwin IT (2007) Deciphering the role of ethylene in plant–herbivore interactions. J Plant Growth Regul 26: 201–209.

War AR, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, Sharma HC (2012) Mechanisms of plant defense against insect herbivores. Plant Signal Behav 7: 1306–1320.

Wittstock U, Gershenzon J (2002) Constitutive plant toxins and their role in defense against herbivores and pathogens. Curr Opin Plant Biol 5: 300–307.

Xiao C., Somerville C, Anderson CT (2014) POLYGALACTURONASE INVOLVED IN EXPANSION1 functions in cell elongation and flower development in *Arabidopsis*. Plant Cell 26: 1018–1035.

Influence of single nucleotide polymorphisms in the *BCL11A*, *HBS1L-MYB* intergenic region, and *HBB* gene cluster on the fetal hemoglobin levels in Bangladeshi patients with β-thalassemia/hemoglobin E disease

Jenjira Chalerm[1, 2], Mahmood Ahmed Chowdhury[3, 4], Chayanon Peerapittayamongkol[2], Kittiphong Paiboonsukwong[1], Suthat Fucharoen[1], Orapan Sripichai[1]*

[1]*Thalassemia Research Center, Institute of Molecular Biosciences, Mahidol University, Nakorn Pathom, Thailand.*
[2]*Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand*
[3]*Department of Paediatrics, Chattagram Maa-Shishu O General Hospital, Chittagong, Bangladesh*
[4]*Thalassaemia Welfare Centre - Bangladesh, Chittagong, Bangladesh*

**Corresponding author: orapan.sri@mahidol.ac.th*

ABSTRACT

Thalassemia is considered a major health burden in the Southeast Asian and Indian populations. Patients with β-thalassemia display disease heterogeneity ranging from nearly asymptomatic to severe anemia with complications. Increased levels of fetal hemoglobin (HbF) can ameliorate the clinical severity of β-thalassemia patients. The HbF production is influenced by many quantitative trait loci (QTL). Three major HbF-QTLs are the *BCL11A* gene, the *HBS1L-MYB* intergenic region (HMIR), and the β-globin (*HBB*) gene cluster. Therefore, this study aimed to evaluate the influence of these genetic modifiers (*BCL11A*, rs766432; HMIR, rs4895411; and *HBB* cluster, rs2071348) on the HbF levels in Bangladeshi patients with β-thalassemia/HbE disease. The cohort study comprised of 90 patients with β-thalassemia/HbE disease from in and around the area of Chittagong, Southeast Bangladesh. The HbF levels ranged from 7.9% to 59.1%. The results showed that levels of HbF were primarily influenced by alleles of the *HBB* cluster (rs2071348), and to a lesser extent by rs766432 *HBL11A* gene and HMIP (rs4895441) loci. The rs2071348 SNP explained 12.5% of the variation in the HbF levels, while 3.6% and 3.9% of trait variation were explained by rs766432 and rs4895441, respectively. In a case-control model of the low and high HbF analysis, we found that genotypes *AC* and *AA* ($p = 2.0\times10^{-4}$) and the allele *C* ($p = 2.0\times10^{-4}$) of rs2071348, and genotypes *AG* and *GG* ($p = 0.02$) and the allele *G* ($p = 0.05$) of rs4895441were associated with a significantly higher frequency with high HbF. However, the rs766432 did not exhibit such features. Our results suggest these three major HbF-QTLs as the influencing phenotypic factors of β-thalassemia in Bangladeshi β-thalassemia/HbE patients.

Keywords: β-thalassemia; HbE; fetal hemoglobin; HbF; SNPs

INTRODUCTION

β-Thalassemia/HbE is a monogenic disorder caused by a compound heterozygous mutation of the β-globin gene. Patients with β-thalassemia/HbE co-inherit the β-thalassemia allele with the structural variant hemoglobin E (HbE; c.79G>A (p.Glu27Lys)). The condition presents with exceptional and variable clinical symptoms ranging from a mild asymptomatic anemia to a transfusion-dependent thalassemia (Sripichai *et al*., 2008). The reduction of β-globin chain synthesis in β-thalassemia causes an excess of unbounded α-globin chains, which precipitate in erythroid cells, leading to ineffective erythropoiesis and a shortened lifespan of red blood cells (Fucharoen and Weatherall, 2016). Therefore, the degree of imbalance between the α-globin chains and the non-α-globin chains dictates the phenotypic severity of the disorder. Increased production of the γ-globin chains can thus modulate the β-thalassemia disease severity by reducing the pool of free α-chains (Weatherall, 2001). In particular, an increase in the levels of fetal hemoglobin (HbF; α2γ2) in many studies was shown to affect the clinical outcome of patients with β-thalassemia (Sripichai and Fucharoen, 2016). Remarkably, the diversity of HbF levels is strongly influenced by genetic factors.

Genome-wide association studies (GWAS) have identified quantitative trait loci (QTL) harboring single nucleotide polymorphisms (SNPs) showing strong associations with HbF levels. The major HbF-QTLs include the *BCL11A* gene on chromosome 2p16.2 (Menzel *et al*., 2007; Uda *et al*., 2008; Nuinoon *et al*., 2010), the *HBS1L-MYB* intergenic region (HMIR) on

chromosome 6q23.3 (Thein *et al*., 2007; Lettre *et al*., 2008; Nuinoon *et al*., 2010), and the β-like-globin gene cluster (*HBB* cluster) on chromosome 11p15.5 (Lettre *et al*., 2008; Thein *et al*., 2009; Nuinoon *et al*., 2010). Several genetic association studies in different populations have shown that the impact of each variant on the variation of HbF levels could be population-specific and is dependent on the genetic background and frequency of each HbF-associated allele (Menzel *et al*., 2007; Lettre *et al*., 2008; Galarneau *et al*., 2010; Thein, 2013)

However, the genetic basis of the β-thalassemia/HbE disease heterogeneity and the genotype-phenotype association are not entirely clear. Several HbF-QTLs previously identified in other populations have not yet been assessed in Bangladeshi individuals. Therefore, the aim of this study was to evaluate the associations between genetic variations in the major HbF-QTLs, namely *BCL11A*, HMIR and *HBB* cluster, and the HbF levels in Bangladeshi patients with β-thalassemia/HbE disease.

MATERIALS AND METHODS

Study population

A total of 90 unrelated β-thalassemia/HbE patients (38.9% were female) with different HbF levels were recruited. The patient ages were 1-43 years (mean age 11.2 ± 8.2 years; median age 9 years). The HbF levels were determined by high performance liquid chromatography (HPLC) using the VARIANT™ Hemoglobin Testing System (β-Thalassemia Short Program, Bio-Rad Laboratories, Hercules, CA, USA). The levels of HbF in this cohort ranged from 7.9% to 59.1%, the mean was 27.3% (SD 13.1) and the median was 24.6%. Forty-four subjects with HbF levels $\geq$ 25% were classified as high HbF whereas 46 subjects with HbF levels < 25% were categorized as low HbF. The study was approved by the Institutional Ethical Committee (IEC) in accordance with the Declaration of Helsinki; certificate of approval number 2010/010.0701. Written informed consent was obtained from all participants or their respective parents in the case of children.

Genotyping

Genomic DNA was extracted from peripheral blood using the Gentra Puregene blood kit (Qiagen Science, Germantown, MD, USA) following the manufacturer's instructions. The quality and quantity of DNA were measured by a Nano Drop spectrophotometer. The β-thalassemia and HbE mutations were confirmed by the GenoFlow Beta thalassemia array test kit (DiagCor Bioscience, Hong Kong, China). Three SNPs were chosen for this study based on the previous studies that reported genetic variants most strongly associated with increased HbF levels in β-thalassemia/HbE patients; *BCL11A* (rs766432), HMIP (rs4895441), and *HBB* cluster (rs2071348). SNPs were genotyped by a PCR-HRM assay using specific primers (oligonucleotide sequences available upon request). Allelic discrimination was carried out using SsoFast EvaGreen Supermix (Bio-rad Laboratories, Hercules, CA, USA) in a CFX Connect™ Real-Time PCR machine (Bio-rad Laboratories).

Statistical analysis

The genotypic distribution frequencies of all polymorphisms were analyzed for Hardy-Weinberg equilibrium (HWE). The association between the HbF levels and the SNPs was analyzed by linear regression under an additive genetic model and by a case-control model of low and high HbF levels (the cutoff value was set up by the median at 25% HbF). Odds ratio (OR) and 95% confidence intervals (CIs) were calculated for association of high and low HbF phenotypes with genotypic and allelic tests. All analyses were conducted using the statistical package R, version 2.5.1. A P-value below 0.05 was considered statistically significant.

RESULTS

The minor allele frequencies (MAF) and HWE observed for the three HbF-associated polymorphisms in the total sample were displayed in Table 1. The minor allele of the *HBB* cluster SNP rs2071348 was *C*, accounting for 34% of the total chromosomes. By contrast, the HMIP SNP, rs4895441 only had 10% of the minor allele *G*, with a *GG* homozygosity percentage of 1%. Meanwhile, the *C* allele of the *BCL11A* SNP rs766432 constituted 1.4% of the whole chromosomes examined, whereas homozygous *CC* was not present in this cohort. The genotype distributions of rs766432 and rs4895441were in agreement with the Hardy-Weinberg equilibrium (HWE; P > 0.05), but the rs2071348 was not in HWE (P = 0.0007).

The distribution of HbF levels within each genotype group of the three HbF- associated SNPs was examined, as shows in box-plots in Figure 1. β-Thalassemia/HbE patients who were homozygous for the major ancestral alleles, *AA* for rs766432, *AA* for rs4895441 and *AA* for rs2071348, showed lower HbF values when compared to those with the derived alleles. The *CC* and *AC* genotypes of rs2071348 in the *HBB* gene cluster showed the higher mean HbF levels

Table 1. HbF association results for the SNPs *BCL11A* rs766432, HMIR rs4895441 and *HBB* gene cluster rs2071348 in Bangladeshi patients with β-thalassemia/HbE disease.

Chr:position	Gene	SNP	Variant	MAF (allele)	P-HWE	β (SE)	Variance explained (%)	P
2:60492835	*BCL11A*	rs766432	A/C	0.14 (C)	0.463	6.23 (2.99)	3.63	0.040
6:135105435	HMIP	rs4895441	A/G	0.10 (G)	1.000	6.82 (3.17)	3.94	0.034
11:5242916	*HBB* cluster	rs2071348	A/C	0.34 (C)	0.0007	5.71 (1.54)	12.47	0.0004

Abbreviations: MAF: minor allele frequency; p-HWE: P-value for Hardy–Weinberg Equilibrium; HMIR: HBS1L-MYB intergenic region. The table includes the effect sizes of the minor allele (regression coefficient beta, β), standard error (SE) and P-values for the HbF levels using a linear regression model. Chromosome position (Chr:position) is according Ensembl.

(33.1±11.6% and 33.4±15.2%, respectively) than the *AA* genotype (22.8±11.2%). The *AC* genotype of *BCL11A* SNP rs766432 showed high HbF levels of 31.7±13.1 % as compared to the *AA* genotype (25.5±12.7%). Similarly, the *AG* genotype of HMIP SNP rs4895441 showed higher HbF levels (35.3±13.0%) in comparison to the *AA* genotype (25.7±12.6%). Regression analysis under an additive genetic model was carried out to examine the impact of SNPs in modifying levels of HbF in the patients. The results showed that rs766432, rs4895441 and rs2071348 polymorphisms displayed similar correlation coefficient values (6.23%, 6.82% and 5.71%, respectively) in increasing HbF level. The HbF phenotypic variation was explained by the three loci from 23.5% to 30.1%. These three SNPs explained more variation in the HbF levels than the cumulative sum of the phenotypic variance explained by the SNPs individually (7.3% compared to 6.8%), although the difference was not statistically significant. The rs2071348 SNP was more strongly associated with the HbF levels ($P = 4.0\times10^{-4}$) than rs766432 and rs4895441 (P = 0.04 and 0.03, respectively) and accounted for 12.5% of the trait variance. The effect of these three HbF-associated SNPs could explain 20% of phenotypic variation in the HbF levels.

To further analyze the distribution frequencies of genotypes and alleles, the β-thalassemia/HbE patients were divided into two groups, a high HbF group (n = 44; range 25.5% to 59.1%, mean HbF = 37.8±10.2%), and a low HbF group (n = 46; range 7.9% to 24.7%, mean HbF = 17.2±5.1%) according to the cutoff value of 25% HbF (Table 2). The mean HbF levels differ significantly between the two groups ($P = 1.2\times10^{-17}$). The case-control study showed significant associations between levels of HbF and the *HBB* cluster SNP rs2071348 (OR = 5.56; 95% CI, 2.27–14.29; $P = 2.0\times10^{-4}$) and HMIP SNP rs4895441 (OR = 4.35; 95% CI, 1.41–16.67; P = 0.02) (Table 2). We found that the *C* allele of rs2071348 and the *G* allele of rs4895441 had a significantly higher frequency in the high HbF group than the low HbF group ($P = 7.0\times10^{-4}$ and 0.05, respectively). Although, the linear regression models revealed a statistical significant association between *BCL11A* SNP rs766432 and levels of HbF, this SNP was not found to correlate with the HbF levels in the case-control model of low and high HbF analysis.

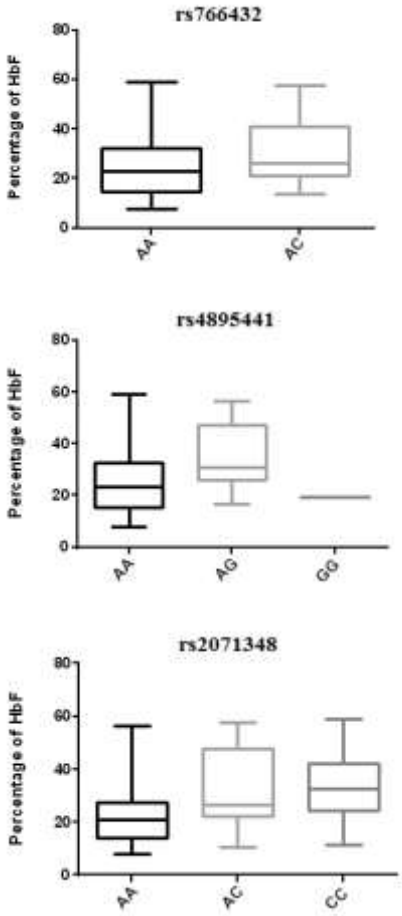


Figure 1 Box plots showing the distribution of HbF levels within genotypes of the SNPs *BCL11A* rs766432, HMIR rs4895441 and *HBB* gene cluster rs2071348 in Bangladeshi patients with β-thalassemia/HbE disease. Each rectangle represents the data between the 25th and 75th quartiles, and the bar within each rectangle is the median value for HbF.

Table 2. Distribution frequencies of genotypes and alleles for the SNPs *BCL11A* rs766432, HMIR rs4895441 and *HBB* gene cluster rs2071348 in the high and low HbF groups of Bangladeshi β-thalassemia/HbE patients.

SNPs	HbF Group	Genotypes		P	OR (95% CI)	Alternative allele	Reference allele	P	OR (95% CI)
rs766432		AC+CC	AA			C	A		
	High	14	30	0.545	1.32 (0.53-3.33)	14	74	0.549	1.32 (0.53-3.34)
	Low	12	34			8	80		
rs4895441		AG+GG	AA			G	A		
	High	13	31	0.017	4.35 (1.41-16.67)	13	75	0.047	3.05 (1.08-10.14)
	Low	4	42			5	87		
rs2071357		AC+CC	AA			C	A		
	High	28	16	0.0002	5.56 (2.27-14.29)	44	44	0.0007	2.64 (1.54-4.79)
	Low	11	35			17	75		

Abbreviations: SNPs: single nucleotide polymorphisms; HbF: fetal hemoglobin; OR (95% CI): odds ratio (95% confidence interval).

DISCUSSION

The reactivation of γ-globin expression in β-thalassemia patients is a response to the reduced level of β-globin synthesis due to the β-thalassemia mutation. The high levels of HbF production contribute to the amelioration of the β-thalassemia clinical manifestations and disease severity. In this study of 90 Bangladeshi β-thalassemia/HbE patient cohort, the range of HbF levels was 7.9–59.1%. According to a systematic review, this level was in accordance with previous studies where HbF levels were found to be 7–70% (Sripichai and Fucharoen, 2016). Genetic variants in three modifier loci; the *BCL11A* gene, the HBS1L-MYB intergenic region and the β-globin like gene cluster, have been proved to be the major factors that can regulate the expression of HbF in different ethnic populations. Similarly, SNPs rs766432 in the *BCL11A* gene, rs4895441 in the HMIR and rs2071348 in the HBB cluster were identified to be significantly associated with the HbF levels in our cohort. This is the first report on the association between these modifier loci and HbF values in Bangladeshi population.

The rs2071348 *A/C* polymorphism, located at the intergenic region between the *HBPB* and the *HBD* genes in the *HBB* cluster, has shown the strongest association with the HbF levels in the cohort of 618 Thai patients with β-thalassemia/HbE (Nuinoon *et al*., 2010). Another study in Indonesian patients also showed that rs2071348 was associated with HbF levels (Nuinoon *et al*., 2010). High-risk genotypes and alleles of SNP rs2071348 (*CC* and *AC*, allele *C*) were determined to be susceptible to having high HbF levels in this study, which demonstrated that there was no difference in racial difference for this variant. In addition, the *C* allele of rs2071348 was significantly more frequent in mildly affected β-thalassemia/HbE patients than in severely affected patients, suggesting that this SNP may help to predict the disease severity (Sherva *et al*., 2010).

Studies in populations in Europe and Asia showed that several polymorphisms in the *HBS1L-MYB* intergenic region had a strong relationship with high HbF levels in both β-hemoglobinopathy patients and normal individuals (Thein *et al*., 2007; Sherva *et al*., 2010; Makani *et al*., 2011). In this study, the rs4895441 SNP was selected to evaluate the effect on HbF levels, and the significant effect was observed by the *G* allele. We found the rs4895441 SNP had significant correlation with the HbF levels in both regression analysis and the case-control study, replicating previous findings in Thai and Indonesian patients. The frequency of the *G* allele in our study was 10%, which is lower than approximately 18% in Thai and Indonesian. However, this study is carried out in a small cohort of Bangladeshi β-thalassemia/HbE patients. Disruption in the intergenic region between the *HBS1L* and *MYB* leads to elevated HbF expression (Farrell *et al*., 2011; Suzuki *et al*., 2013) through the reduction of transcription factor bindings that affect long-range interactions with *MYB*, resulting in reduced *MYB* expression (Stadhouders *et al*., 2014). MYB is an erythroid transcription factor, whose down-regulation in adult erythroid cell was reported to result in increased γ-globin expression with reduced cell expansion and accelerated erythroid differentiation (Jiang *et al*., 2006; Soza-Ried *et al*., 2010), suggesting that levels of MYB might affect HbF production through its influence on the cell cycle. HBS1L is a G-protein/elongation factor involved in a variety of cellular processes. But overexpression of *HBS1L* in K562 erythroid cells did not affect the γ-globin expression (Jiang *et al*., 2006).

Our data in β-thalassemia/HbE patients showed nominal statistically significant associations

between *BCL11A* SNP rs766432 and the HbF levels. rs766432 is known to correlate with the levels of HbF in various cases of β-hemoglobinopathies in Thai, Chinese, Middle East, European, and African populations (Sedgewick *et al*., 2008; Nuinoon *et al*., 2010; Akinsheye *et al*., 2012; Pule *et al*., 2015). Thus, our finding is in line with previous studies, confirming that the rs766432 SNP has a significant correlation with increasing HbF levels among thalassemia patients. However, case-control models that assess the correlation of variants in patients with HbF levels between high and low HbF (the considered cutoff point of 25%) did not show significant association between HbF and rs766432. The absence of significant association between rs766432 and the HbF levels may be explained by the low population frequency of this variant, with a 1.4% frequency for the minor *C*-allele in the cohort. BCL11A is a regulator responsible for γ-globin to β-globin gene switching (Sankaran *et al*., 2008). BCL11A interacts with other repressor factors to form a repressor complex, thereby silencing the γ-globin gene in adult erythroid cells. The rs766432 SNP is located within an erythroid-specific enhancer of *BCL11A* and likely to influence *BCL11A* expression (Bauer *et al*., 2013).

In conclusion, our results suggest that the level of HbF in Bangladeshi β-thalassemia/HbE patients is strongly associated with the *HBB* gene cluster rs2071348, but also with *BCL11A* rs766432 and HMIR rs4895441. The effect of these three HbF-associated SNPs explained for 20% of the variance in HbF levels in this cohort. A larger sample size is needed and the mechanistic function(s) of these modifier loci should be explored in future studies.

ACKNOWLEDGEMENTS

This work was supported by Mahidol University and the Research Chair Grant, National Science and Technology Development Agency (NSTDA), Thailand. Jenjira Chalerm received the Siriraj Graduate Scholarship, Faculty of Medicine Siriraj Hospital, Mahidol University.

REFERENCES

Akinsheye I, Solovieff N, Ngo D, Malek A, Sebastiani P, Steinberg MH, Chui DH (2012) Fetal hemoglobin in sickle cell anemia: molecular characterization of the unusually high fetal hemoglobin phenotype in African Americans. Am J Hematol 87: 217–219.

Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L, *et al*. (2013) An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science 342: 253–257.

Farrell JJ, Sherva RM, Chen Z, Luo H, Chu BF, Ha SY, Li CK, Lee ACW, Li RCH, Li CK, *et al*. (2011) A 3-bp deletion in the HBS1L-MYB intergenic region on chromosome 6q23 is associated with HbF expression. Blood 117: 4935–4945.

Fucharoen S, Weatherall DJ (2016) Progress toward the control and management of the thalassemias. Hematol Oncol Clin North Am 30: 359–371.

Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G (2010). Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. Nat Genet 42: 1049–1051.

Jiang J, Best S, Menzel S, Silver N, Lai MI, Surdulescu GL, Spector TD, Thein SL (2006) cMYB is involved in the regulation of fetal hemoglobin production in adults. Blood 108: 1077–1083.

Lettre G, Sankaran VG, Bezerra MA, Arau'jo AS, Uda M, Sanna S, Cao A, Schlessinger D, Costa FF, Hirschhorn JN, *et al*. (2008) DNA polymorphism at the BCL11A, HBS1L-MYB, and β-globin loci associates with fetal haemoglobin levels and pain crises in sickle cell disease. Proc Natl Acad Sci USA 105: 11869–11874.

Makani J, Menzel S, Nkya S, Cox SE, Drasar E, Soka D, Komba AN, Mgaya J, Rooks H, Vasavda N, *et al*. (2011) Genetics of fetal hemoglobin in Tanzanian and British patients with sickle cell anemia. Blood 117: 1390–1392.

Menzel S, Garner C, Gut I, Matsuda F, Yamaguchi M, Heath S, Foglio M, Zelenika D, Boland A, Rooks H, *et al*. (2007) A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nat Genet 39: 1197–1199.

Nuinoon M, Makarasara W, Mushiroda T, Setianingsih I, Wahidiyat PA, Sripichai O, Kumasaka N, Takahashi A, Svasti S, Munkongdee T, *et al*. (2010) A genome-wide association identified the common genetic variants influence disease severity in β-thalassemia/ hemoglobin E. Hum Genet 127: 303–314.

Pule GD, Ngo Bitoungui VJ, Chetcha CB, Kengne AP, Antonarakis S, Wonkam A (2015) Association between variants at BCL11A erythroid-specific enhancer and fetal hemoglobin levels among sickle cell disease patients in cameroon: implications for future therapeutic interventions. Omics 19: 627–631.

Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH (2008) Human fetal hemoglobin expression

is regulated by the developmental stage-specific repressor BCL11A. Science 322: 1839–1842.

Sedgewick AE, Timofeev N, Sebastiani P, So JC, Ma ES, Chan LC, Fucharoen G, Fucharoen S, Barbosa CG, Vardarajan BN, *et al.* (2008) BCL11A is a major HbF quantitative trait locus in three different populations with beta-hemoglobinopathies. Blood Cells Mol Dis 41: 255–258.

Sherva R, Sripichai O, Abel K, Ma Q, Whitacre J, Angkachatchai V, Makarasara W, Winichagoon P, Svasti S, Fucharoen S, *et al.* (2010) Genetic modifiers of Hb E/beta0 thalassemia identified by a two-stage genome-wide association study. BMC Med Genet 11: 51–59.

Soza-Ried C, Hess I, Netuschil N, Schorpp M, Boehm T (2010) Essential role of c-myb in definitive hematopoiesis is evolutionarily conserved. Proc Natl Acad Sci USA 107: 17304–17308.

Sripichai O, Fucharoen S (2016) Fetal hemoglobin regulation in β-thalassemia: heterogeneity, modifiers and therapeutic approaches. Expert Rev Hematol 9: 1129–1137.

Sripichai O, Makarasara W, Munkongdee T, Kumkhaek C, Nuchprayoon I, Chuansumrit A, Chuncharunee S, Chantrakoon N, Boonmongkol P, Winichagoon P, *et al.* (2008) A scoring system for the classification of beta-thalassemia/Hb E disease severity. Am J Hematol 83: 482–484.

Stadhouders R, Aktuna S, Thongjuea S, Aghajanirefah A, Pourfarzad F, van Ijcken W, Lenhard B, Rooks H, Best S, Menzel S, *et al.* (2014) HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. J Clin Invest 124: 1699–1710.

Suzuki M, Yamazaki H, Mukai HY, Motohashi H, Shi L, Tanabe O, Engel JD, Yamamoto M (2013) Disruption of the Hbs1l-Myb locus causes hereditary persistence of fetal hemoglobin in a mouse model. Mol Cell Biol 33: 1687–1695.

Thein SL, Menzel S, Peng X, Best S, Jiang J, Close J, Silver N, Gerovasilli A, Ping C, Yamaguchi M, *et al.* (2007) Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. Proc Natl Acad Sci USA 104: 11346–11351.

Thein SL, Menzel S (2009) Discovering the genetics underlying foetal haemoglobin production in adults. Br J Haematol 145: 455–467.

Thein SL (2013) Genetic association studies in β-hemoglobinopathies. Hematology Am Soc Hematol Educ Program 2013: 354–361.

Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, *et al.* (2008) Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta thalassemia. Proc Natl Acad Sci USA 105: 1620–1625.

Weatherall DJ (2001) Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. Nat Rev Genet 2: 245–255.

Identification of genes involving in salt tolerance using GWAS data based on Na^+ content in local Thai rice leaves and Arabidopsis orthologous gene validation

Ratchata Chokwiwatkul[1], Natthaya Tantipirom[1], Noppkhun Khunpolwatatna[1], Apichart Imyim[2], Duangjai Suriya-aroonroj[3], Teerapong Buaboocha[1,4,5], Monnat Pongpanich[5,6], Luca Comai[7], Supachittra Chadchawan[1,5]*

[1]Center of Excellence in Environment and Plant Physiology, Department. of Botany, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand.
[2]Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand.
[3]Nakhon Ratchasima Rice Research Center, Bureau of Rice Research and Development, Nakhon Ratchasima 30110, Thailand.
[4]Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand.
[5]Omic Science Center, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand.
[6]Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand.
[7]Department of Plant Biology, UC Davis Genome Center, UC Davis, Davis, CA 95616, USA.

**Corresponding author: s_chadchawan@hotmail.com; Supachitra.C@chula.ac.th*

ABSTRACT

High salinity is one of the most abiotic stresses that can adversely affect plant growth, development and productivity in rice for rice grown in the Northeastern part of Thailand. To overcome this problem and improve crop yield under salt stress conditions, it is important to improve salt tolerance in crops. This research aims to validate if the SNPs predicted by Genome-Wide Association Study (GWAS) is consistent with the QTL for salt tolerant trait that was previously reported. GWAS based on Na^+ content in leaf tissues of local Thai rice varieties after six days of salt stress revealed the involvement of SNPs from 10 loci, located on chromosomes 1, 2, 5, 10, 11 and 12. Five loci are in the reported salt tolerant QTL. To determine if these loci contributing to salt tolerant phenotype, the orthologous genes in Arabidopsis were used as the representatives. Homozygous Arabidopsis mutant lines, each of which contains the T-DNA in the orthologous genes of the predicted loci, were analyzed in comparison with wild type for the physiological responses, fresh weight, dry weight, chlorophyll *a*, chlorophyll *b* and carotenoids contents, after 7 days of 250 mM NaCl treatment. All of the mutant lines had the significant lower levels of chlorophyll *a* and chlorophyll *b* contents in normal condition, while under salt stress condition, all of them showed the significant difference in stability index of photosynthetic pigment contents. These indicate that all of these genes are involved in salt tolerance, which implies the impact of the predicted loci from GWAS in salt tolerant ability of rice. Our data show the consistency between GWAS and QTL studies and suggested that GWAS can be used to identify the precise loci that are responsible for the trait of interest.

Keywords: genome-wide association study (GWAS); Na^+/K^+ ratio; salt stress; QTL; Arabidopsis mutant lines

INTRODUCTION

Rice (*Oryza sativa* L.) is one of the world's most crucial food crops. Rice has previously been reported to be susceptible to salt stress at both seedling and reproductive stages, which leads to grain yield reduction (Zeng *et al.*, 2001; Moradi and Ismil, 2007). It affects physiological processes, for instance, Na^+ uptake, exclusion, and ion imbalance (Na^+/K^+ ratio) (Siringam *et al.*, 2011). Salinity causes two main problems; osmotic stress which reduces the capacity of the plant to absorb water and decreases soil water potential (Zhu *et al.*, 1997), and ionic stress which is induced by an excessive concentration of salt, especially sodium chloride (NaCl) from the soil (Jouyban, 2012). Excess of Na^+ in plant cells causes damage in membrane systems and organelles, resulting in plant growth reduction and abnormal development (Davenprot *et al.*, 2005; Quintero *et al.*, 2007). Moreover, high Na^+ concentration inhibits uptake of K^+

ions which is an essential element for growth, development and enzyme activities. Eventually, these culminate in lower productivity and may even lead to death (Luan, 2009; Rubio *et al.,* 1995).

Salt tolerance is a quantitative trait, controlled by multiple loci in the genome. Several studies have reported quantitative trait loci (QTL) responsible for this trait (Dixit *et al.,* 2015; Koyama *et al.,* 2001; Dell'Acqua *et al.,* 2015). Moreover, genome wide association study (GWAS) was also studied for salt tolerance trait in rice (Kumar *et al.,* 2015). However, the regions in the genome responsible for this trait identified by these two approaches were not consistent.

This research aims to identify the SNPs involving in Na^+ content in rice using GWAS approach, and validate the involvement of the genes in salt tolerance if they were in the regions of salt tolerant QTL that were previously reported. For validation of gene functions in salt tolerance, the Arabidopsis's orthologous genes of the predicted rice loci were analyzed. The homozygous mutant lines containing T-DNA insertion at the orthologous genes were evaluated for the phenotypes grown in normal and salt stress conditions in comparison with the wild type. The gene mutations that result in significant difference between the wild type and the mutant line under salt stress condition would be considered as the salt tolerant-related genes.

MATERIALS AND METHODS

Plant Materials and growing conditions

For the GWAS experiment, 75 local Thai rice varieties (Table 1) were grown in 4-inch pots, containing 800 g of soil for 14 days in natural condition. Four replications of each variety per condition were used in this experiment. Then, seedlings were separated into 2 groups, one was treated with 115 mM NaCl solution, which caused soil salinity to 8-9 $dS.m^{-1}$, and the other was treated with filtered water with the same volume. After 6 days of the treatment, leaf tissues were collected and weighed and dried until used for Na^+ content measurement.

Arabidopsis T-DNA insertion lines (Alonso *et al.,* 2003) were used for the validation of the orthologous gene involvement in salt tolerance. The mutant lines containing the T-DNA insertion at the orthologous genes of the predicted rice loci from GWAS results were ordered from the Arabidopsis Biological Resource Center (ABRC).

Arabidopsis mutant line seeds were sterilized with 7.85% w/w of sodium hyperchlorite and 0.01% of Tween20 for 12 minutes. Then, seeds were washed with sterile water for 5 times. Sterile Arabidopsis seeds were placed on Murashige and Skoog (MS) media plates (Murashige and Skoog, 1962) and incubated in 4°C for 48 hours. After that, the seeds were grown on 18°C for 7 days. Then, Arabidopsis seedlings were transferred to soil (peat moss: lite: vermiculite, 3: 1: 1, respectively) and grew for another 7 days before leaf tissue collection for homozygosity determination. The seeds from the homozygous plants were collected for the physiological response analysis. If it was the heterozygous plant, the seeds were collected and replanted. Then, the screening was done again until the homozygous line was obtained.

Genome Wide Association Study (GWAS)

Rice genomic DNA was extracted from leaf tissues and subject to high-throughput sequencing. Then, SNP calling was done as followed. The Illumina paired-end reads were mapped to the rice reference genome, IRGSP-1.0.21 (Kawahara *et al.,* 2013) using Burrows Wheeler Aligner software (BWA version 0.7.10) (Li and Durbin, 2009). Variants from rice varieties were called using Genome Analysis Toolkit (GATK) (McKenna *et al.,* 2010). SNPs were filtered using the following criteria: minimum coverage in order for a position to be called homozygous was 5; minimum coverage of each of the two observed major basecalls to be called heterozygous was 5; and minimum percentage of each of the two observed major basecalls in order to be called heterozygous is 20.

Association analysis was performed with the following condition. SNPs with more than 40% missing data were removed from the analysis. Ungenotyped SNPs were then inferred by genotype imputation method using Beagle (Browning and Browning, 2016). Principle Component Analysis based on SNPs in all accessions was conducted using the EIGENSOFT package (Patterson *et al.,* 2006). Association analysis was performed using the mixed model implemented in the software GEMMA (Zhou and Stephens, 2012). The significance threshold, set at p values from Wald test less than 10^{-13}, was used in this research.

Screening of homozygous T-DNA insertion lines of Arabidopsis

Total genomic DNA of Arabidopsis was extracted from leaf tissues according to Chabi et al (2015). Then, the homozygous insertion at the target genes were screened using the specific primer set, LP+RP and BP+RP. LP and RP are the primers specific

to the sequence of the target gene at the 5'end and 3'end of the T-DNA insertion, which will be about 900 bp apart from each other. BP primer is specific to the sequence of T-DNA fragment. If there is the T-DNA insertion at the target gene, the positive band (about 400 bp) from the amplification with LP and BP primers will be detected, if no insertion, no amplification will be obtained. Moreover, the amplification with LP and RP will be detected (about 900 bp) in the case of no T-DNA insertion. If T-DNA is inserted, LP and RP will not be able to amplify any fragment as the length of amplification is not appropriate with the amplification cycle (Figure 1). The wild type (WT) Arabidopsis, ecotype Columbia-0 (Col-0), was used as the control.

Table 1 Information of rice varieties used in the GWAS experiment

No.	Rice varieties	GS No.	Type of rice	Salt tolerance level*	Source of rice variety
1	IR29	-	Lowland rice	Salt sensitive standard variety	Philippines
2	POKKALI	-	Lowland rice	Salt tolerant standard variety	India
3	RD 73	-	Lowland rice	The latest salt tolerant cultivar from Rice Department, Thailand	Thailand
4	LPT123	868	Lowland rice	2	Thailand
5	GWIAN HAK	13201	Lowland rice	1	Thailand
6	JAO RAHK HAENG	12494	Lowland rice	2	Thailand
7	CHIANG PHATTHALUNG	21964	Lowland rice	2	Thailand
8	LEB NOK	3979	Lowland rice	2	Thailand
9	SETTI	5678	Lowland rice	2	Thailand
10	NIAW DAM LAI	21240	Lowland rice	1	Thailand
11	LEUANG TIA	6448	Lowland rice	4	Thailand
12	LEUANG GAEW	6812	Lowland rice	4	Thailand
13	LEUANG BAI JAEK	14155	Lowland rice	2	Thailand
14	LEUANG BAI LOD	14154	Lowland rice	4	Thailand
15	LEUANG KWAI LAH	5551	Lowland rice	2	Thailand
16	LEUANG KWAI LAH	7285	Lowland rice	4	Thailand
17	LEUANG HUAN	7303	Lowland rice	4	Thailand
18	LEUANG PRATEW	5681	Lowland rice	2	Thailand
19	LEUANG PLAH GIM	7293	Lowland rice	2	Thailand
20	GAEN JAN	22653	Lowland rice	3	Thailand
21	JAE GAN	12507	Lowland rice	3	Thailand
22	DAENG NAH	22850	Lowland rice	1	Thailand
23	RD10	4790	Lowland rice	3	Thailand
24	RD12	21628	Lowland rice	1	Thailand
25	RD17	3999	Deep water rice	3	Thailand
26	RD19	4000	Deep water rice	2	Thailand
27	RD21	4791	Lowland rice	2	Thailand
28	RD25	4793	Lowland rice	3	Thailand
29	RD27	7125	Lowland rice	3	Thailand
30	RD31	24533	Lowland rice	2	Thailand
31	GAM FEUANG	4490	Lowland rice	1	Thailand
32	PRACHIN BURI 1	23406	Deep water rice	2	Thailand
33	KHAO SAMER	3370	Deep water rice	2	Thailand
34	KHAO GAEW	6152	Lowland rice	3	Thailand
35	KHAO TAENG MO	3810	Lowland rice	3	Thailand
36	KHAO KOD	12503	Lowland rice	1	Thailand
37	KHAO TAH JEUA	3808	Lowland rice	2	Thailand
38	KHAO TAH CHEUA	12270	Lowland rice	2	Thailand

Table 1 (continued)

No.	Rice varieties	GS No.	Type of rice	Salt tolerance level*	Source of rice variety
39	KHAO BAHN POD	19877	Lowland rice	1	Thailand
40	KHAO PUANG	9362	Lowland rice	3	Thailand
41	KHAO LUANG	7282	Lowland rice	2	Thailand
42	KHAO LUANG	5533	Lowland rice	2	Thailand
43	KHAO' HAO	5625	Lowland rice	1	Thailand
44	KHITOM KHAO	21708	Lowland rice	2	Thailand
45	KAN NAH	6442	Lowland rice	2	Thailand
46	JAMPAH TAWNG	5211	Lowland rice	2	Thailand
47	CHAW PLI KHAO	9742	Lowland rice	2	Thailand
48	CHAW MA GAWK	2042	Deep water rice	2	Thailand
49	CHAI NAT 1	20712	Lowland rice	3	Thailand
50	CHUMPAE 60	16235	Lowland rice	1	Thailand
51	SEW MAE JAN	4001	Upland rice	3	Thailand
52	DAWK KHAH	12160	Lowland rice	3	Thailand
53	TAH JEUA	5545	Lowland rice	2	Thailand
54	TAH BAHN	21695	Lowland rice	3	Thailand
55	NAH KHAWAN	22379	Lowland rice	2	Thailand
56	NAHNG NUAN	3151	Lowland rice	2	Thailand
57	NAM SAGUI 19	3023	Lowland rice	2	Thailand
58	NAM SAGUI 19	15833	Lowland rice	2	Thailand
59	BEU SAW MI	23595	Upland rice	1	Thailand
60	BUN MAH	3031	Lowland rice	1	Thailand
61	PLAH KHAENG	3241	Lowland rice	2	Thailand
62	PLAI NGAHM PRACHIN BURI	20864	Deep water rice	3	Thailand
63	PUANG NGERN	2963	Lowland rice	1	Thailand
64	PUANG TAWNG	18442	Deep water rice	2	Thailand
65	PUANG NAHK	12266	Lowland rice	2	Thailand
66	PUANG HAHNG NAHK	23233	Lowland rice	3	Thailand
67	LOI HAH RUANG	6230	Lowland rice	3	Thailand
68	LAI MAHK	7025	Lowland rice	4	Thailand
69	LOOK DAENG PATTANI	21963	Lowland rice	3	Thailand
70	LOOK DAENG PATTANI	23303	Lowland rice	2	Thailand
71	SANG YOD	15101	Lowland rice	2	Thailand
72	SAHM RUANG	7288	Lowland rice	3	Thailand
73	SOON	22492	Lowland rice	3	Thailand
74	LUANG PRATAHN	6440	Lowland rice	3	Thailand
75	HANTRA 60	16232	Deep water rice	2	Thailand

*The salt tolerant level was reported based on hierarchical clustering with the seedling phenotypes of salt injury score, relative water content, cell membrane stability, Na^+, K^+ and Na^+/K^+ content in leaves under salt stress condition for 9 days. Four clusters were found and ranked 1 as the most salt sensitive and 4 as the most salt tolerant group.

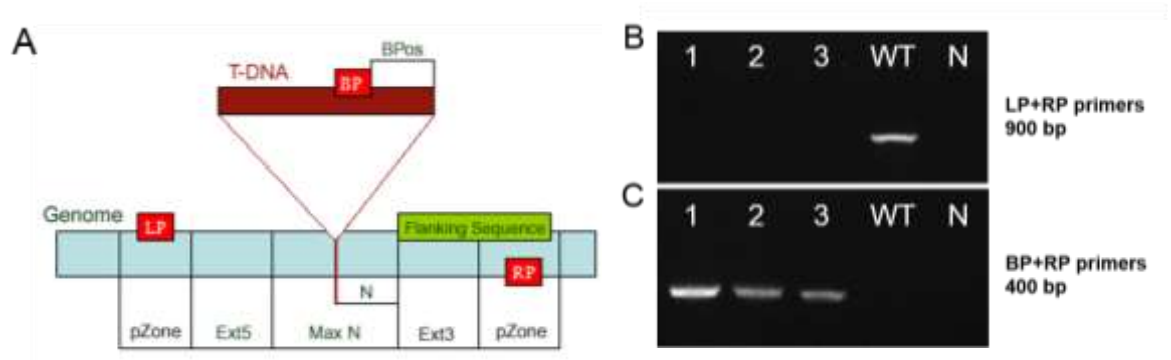


Figure 1 Example of homozygous mutant line screening. Homozygous plants (1, 2 and 3) were screened with the genomic DNA amplification using LP+RP primers and LB+RP primers. The locations of the primers were shown in A. The amplification of LP+RP and LB+RP primers was shown in B and C, respectively. The patterns of the DNA bands were different between wild type (WT) and homozygous plants. N was the negative control sample.

Physiological response comparison of mutant lines and WT

Experimental design

Randomized complete block design (RCBD) with 4 replications and 4 plants per replication was performed to compare physiological responses, determined by fresh weight, dry weight, chl *a*, chl *b* and carotenoid contents, between *Arabidopsis thaliana* ecotype Columbia-0 (Col-0) (WT) and homozygous Arabidopsis mutant lines under normal and salt stress conditions.

Growing condition for physiological studies

Homozygous Arabidopsis mutant seeds and WT seeds were sterilized with 7.85% w/w of sodium hyperchlorite and 0.01% of Tween20 for 12 minutes. Then, seeds were washed with sterilized water for 5 times. Then the seeds were placed on MS plates and incubated at 4°C for 48 hours. After that, they were germinated under 35μmol.m^{-1}.s^{-1} light intensity with 16/8 light/dark period at 22°C for 7 days. Then, Arabidopsis seedlings were transferred to grow in the mix of peat moss: perlite: vermiculite at the ratio of 3: 1: 1 and grew in the controlled condition room (closed system) under 70 μmol.m^{-1}.s^{-1} light intensity with 16/8 light/dark period at 23°C for 14 days. Total 44 plants were grown in each tray at the size of 50 x 40 cm^2 and 3 liters of water was supplied at the beginning of the transfer and the level of water was maintained until the end of the experiment.

Twenty-one-day-old seedlings were separated into two groups, one was for normal grown condition and the other was for salt stress grown condition. For normal grown condition, the plants were maintained as mention above, while under the salt stress condition, 3 liters of 250mM NaCl solution was supplied instead of water. The level of water or solution was maintained for 7 days by addition of filtered water. The soil conductivity in normal growing condition was 2 dS.m^{-1}, while in the salt stress condition was 9 dS.m^{-1}.

Collection of physiological parameters

After 7 days of treatment, the above ground tissues were collected and determined as fresh weight per plant. The tissues then were dried under 80°C until no weight change to determine dry weight per plant. Physiological parameters were collected at day 7 after treatment, e.g. fresh weight, dry weight, chl *a*, chl *b* and carotenoids contents from leaf tissue.

For photosynthetic pigment content measurement, the first fully-expanded leaf of each plant sample was collected and weighed. The pigments were extracted with 80% acetone at 4°C for 48 hours in darkness. The absorbance at 470, 646.8 and 663.2 nm was determined using spectrophotometer (Agilent 8453 UV-visible Spectroscopy System). The pigment contents, chlorophyll *a* (chl *a*), chlorophyll *b* (chl *b*) and carotenoids were calculated according to Wellburn (1994) as shown below:

Chlorophyll *a* content = $12.25A_{663.2} - 2.79A_{646.8}$
Chlorophyll *b* content = $21.5A_{646.8} - 5.1A_{663.2}$
Carotenoids content
= $[(1000A_{470} - 1.82$ Chl $a - 85.02$ Chl $b)]/ 198$

Data analysis

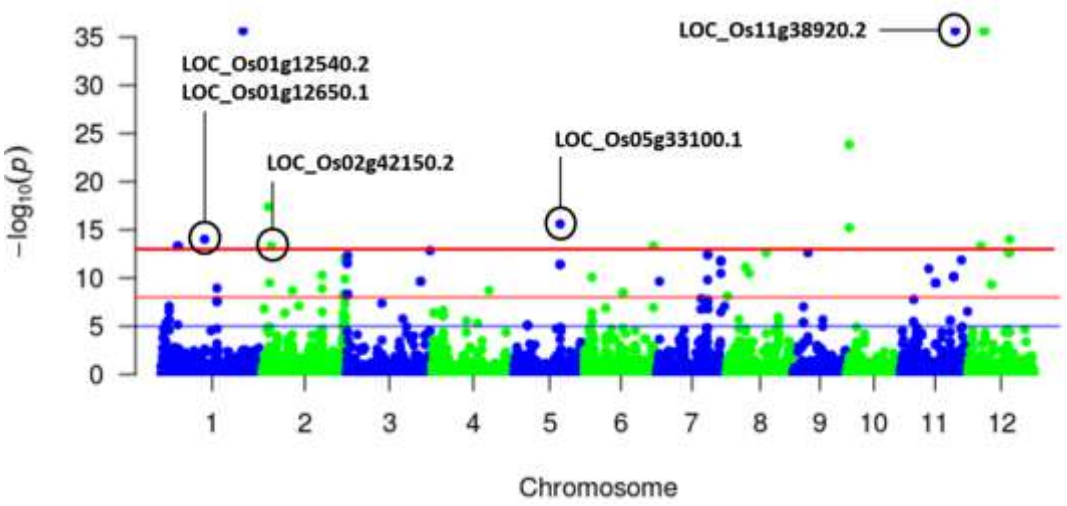
Analysis of variance of each parameter was performed and the means were compared with Duncan's multiple range tests at a significant level of 0.05 using SPSS statistic base software package. The cluster analysis of the SNPs from 10 loci was performed according to Ward's method using JMP10 software.

RESULTS AND DISCUSSION

Predicted loci from GWAS and identification of orthologous genes in Arabidopsis

Based on the threshold of *p* values from Wald test less than 10^{-13}, 10 loci on chromosome 1, 2, 5, 10, 11 and 12 was predicted to be involved with Na^+ content under salt stress condition. Among these, five of them, LOC_Os02g42150.2, LOC_Os01g12540.2, LOC_Os01g12650.1, LOC_Os05g33100.1, and LOC_Os11g38920.2, were located in the QTL regions previously reported (Figure 2 and Table 2). None of these were consistent with the predicted loci from GWAS performed by Kumar et al (2015). This should be due to the difference in rice populations used for the analyses. This research was done with local Thai rice varietie, while Kumar's was performed with Indian rice varieties. Moreover, the p-value threshold used by Kumar et al (2015) ($p<10^{-4}$) was higher than what we used in this experiment ($p<10^{-13}$). This is due to the number of SNPs obtained from each dataset. In our data, 77,063 SNPs were obtained for the association, while the 6,000 SNP chip was used in Kumar et al's experiment.

Some predicted loci from the GWAS were characterized for their functions. LOC_Os02g42150.2 encodes for OsWAK14, which was previously identified to be a positive regulator for rice blast resistance (Delteil *et al.*, 2016; Cyrol *et al.*, 2016). There has been no report on salt stress response in this gene. LOC_Os01g12540.2 encodes for a GTPase with unknown function. In Arabidopsis, RAB11 GTPases are required for salt tolerance (Asaoka *et al.*, 2013). Therefore, this gene may have the similar function in rice. LOC_Os01g12650.1 encodes for a reticulon domain containing protein. In Tibetan wild barley (*Hordeum spontaneum* L.), the reticulon domain-containing protein (F2DHE3) was up-regulated in the variety that had the higher Na^+ accumulation. This finding was suggested to be involved with salt adaptation in this species (Shen *et al.*, 2016). Moreover, reticulon family protein (RTNLB2) in Arabidopsis was also up-regulated under salt stress condition (Ma and Bohnert, 2007).



No	Rice Locus	Arabidopsis Locus	Gene Product Name
1	LOC_Os01g12540.2	AT5G39960	GTPase of unknown function domain containing protein
2	LOC_Os01g12650.1	AT4G11220	Reticulon domain containing protein
3	LOC_Os02g42150.2	AT1G21230	OsWAK14 - OsWAK receptor-like protein kinase
4	LOC_Os05g33100.1	AT3G23600	Endo-1,3;1,4-beta-D-glucanase precursor
5	LOC_Os11g38920.2	AT1G74520	HVA22

Figure 2 Manhattan plot of the Na^+ content in leaves after 6 days of salt stress. The selected loci were cut off with $P<10^{-13}$. The positions that were consistent with QTL regions were circled and the orthologous genes were listed in the table below.

Table 2 Loci identified by GWAS with Na^+ content in leaf tissues and the p-value threshold used was less than 10^{-13}.

No.	P-value	Rice chr.	Position of candidate SNPs on rice chromosome	Locus	Gene product name	Orthologous gene in Arabidopsis	Referents for salt tolerance QTL
1	1.42×10^{-14}	2	25349420	LOC_Os02g42150	OsWAK14 - OsWAK receptor-like protein kinase	AT1G21230	Dixit *et al.*, 2015
2	2.44×10^{-14}	1	6900271	LOC_Os01g12540	GTPase of unknown function domain containing protein	AT5G39960	Koyama *et al.*, 2001
3	2.44×10^{-14}	1	6969721	LOC_Os01g12650	Reticulon domain containing protein	AT4G11220	Koyama *et al.*, 2001
4	2.73×10^{-14}	5	19412122	LOC_Os05g33100	Endo-1,3;1,4-beta-D-glucanase precursor	AT3G23600	Dixit *et al.*, 2015
5	4.89×10^{-15}	12	17420950	LOC_Os12g29350	ATP binding protein	AT3G48770	-
6	3.42×10^{-16}	10	1023521	LOC_Os10g02644	Hypothetical protein	-	-
7	2.78×10^{-18}	2	2320436	LOC_Os02g38380	Transposon protein	-	-
8	5.94×10^{-36}	11	23175746	LOC_Os11g38920	HVA22	AT1G74520	Dell'Acqua *et al.*, 2015
9	5.94×10^{-36}	1	34769930	LOC_Os01g60110	E2F-related protein	AT1G09575	-
10	5.94×10^{-36}	12	6912259	LOC_Os12g12560	NADP-dependent oxidoreductase	AT3G03080	-

LOC_Os05g33100.1 encodes endo-1,3;1,4-beta-D-glucanase precursor. Endo-1,4-beta-D- glucanase plays a role in cell wall extensibility. It was reported to be involved in flooding, heavy metal, ozone and salt stress in various species (Gall *et al.*, 2015). The transcription factor OsMPS is induced by salt stress and regulates cell wall biosynthesis genes, including endo-1,4-beta-D- glucanase in rice (Lippold *et al.*, 2009; Schmidt *et al.*, 2013). LOC_Os11g38920.2 is orthologous to HVA22, which is induced by environmental stresses, including salt stress (Brands and Ho, 2002). HVA22 in Arabidopsis is regulated by abscisic acid (ABA) and salt stress (Chen *et al.*, 2002). There is a high potential that this gene in rice is involved in salt stress response.

Five loci that are located in the salt tolerant QTL were selected for further characterization by searching for the orthologous genes in Arabidopsis by using the data from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/) (Kahawara *et al.*, 2013). These orthologous genes are AT5G39960, AT4G11220, AT1G21230, AT3G23600 and AT1G74520 (Table 2).

Physiological responses under salt stress in homozygous Arabidopsis mutant lines

The homozygous mutant lines of Arabidopsis were evaluated for salt stress responses with WT. All of the mutant lines and wild type (WT) had the similar fresh weight and dry weight in both conditions (Figure 3). In normal condition, all mutants showed the significant lower levels of all photosynthetic pigment contents than wild type, except the mutant with insertion at AT1G74520 that had the similar level of carotenoids to WT. The lowest amount of chl *a* and chl *b* in normal grown plants was found in AT4G11220 and AT3G23600 mutant lines, which had about a quarter of chlorophyll content of WT. However, the carotenoid contents of these two mutant lines were similar to those of WT (Figure 4). The significantly lower carotenoid content in plants grown under normal condition was found in the AT5G39960 and AT1G21230 mutants, while other mutant lines showed the similar level of carotenoids (Figure 4).

To determine the effects of salt stress, stability index, calculated from the ratio of pigment content in stress grown plants to pigment content in normal grown plants, was used for the comparison (Figure 5). This parameter reveals the ability to maintain photosynthetic pigment levels under salt stress condition, which will reflect the ability of photosynthetic performance under salt stress. It has been shown that the salt tolerant genotypes had the higher ability to maintain chlorophyll content (Sevengor et al. 2011) For chl *a*, stability indices of all mutant lines showed no significant difference from stability index of chl *a* in WT. On the contrary, four mutant lines, AT5G39960, AT4G11220, AT3G23600 and AT1G74520, showed significantly lower stability index of chl *b* contents than that of WT. Although the AT1G21230 mutant had the similar stability indices of chl *a* and chl *b* content to WT, it had the significantly higher stability index of carotenoid. Another mutant line that had the significantly higher stability index of carotenoid than WT was AT5G39960 mutant. AT1G74520 mutant showed the similar stability index of carotenoid to WT, while the other two mutants, AT4G11220 and AT3G23600 had the significant lower carotenoid stability index than WT (Figure 5). These data indicate that mutations in these genes of Arabidopsis affect the stability index of photosynthetic pigments, which implies the importance of these genes in salt tolerance.

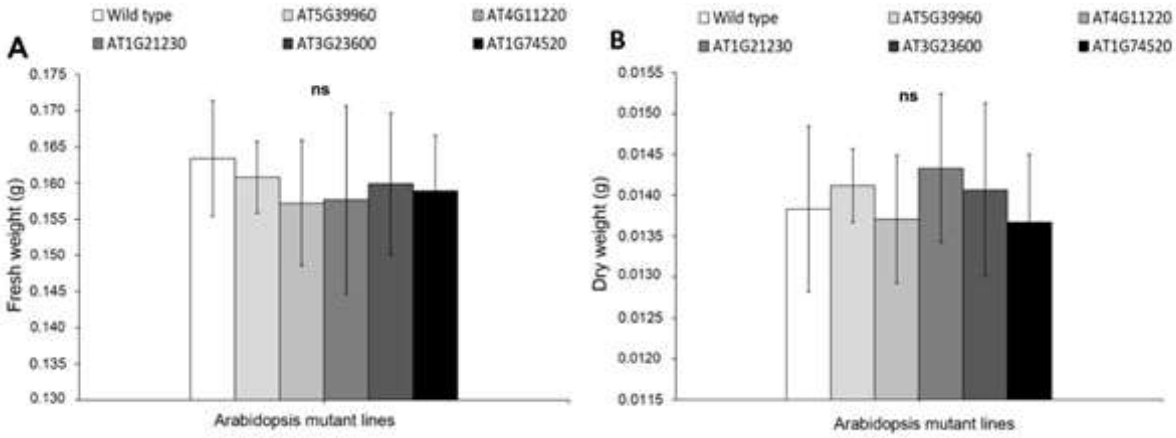


Figure 3 Fresh weight (A) and dry weight (B) of wild type and 5 mutants after grown in salt stress for 7 days

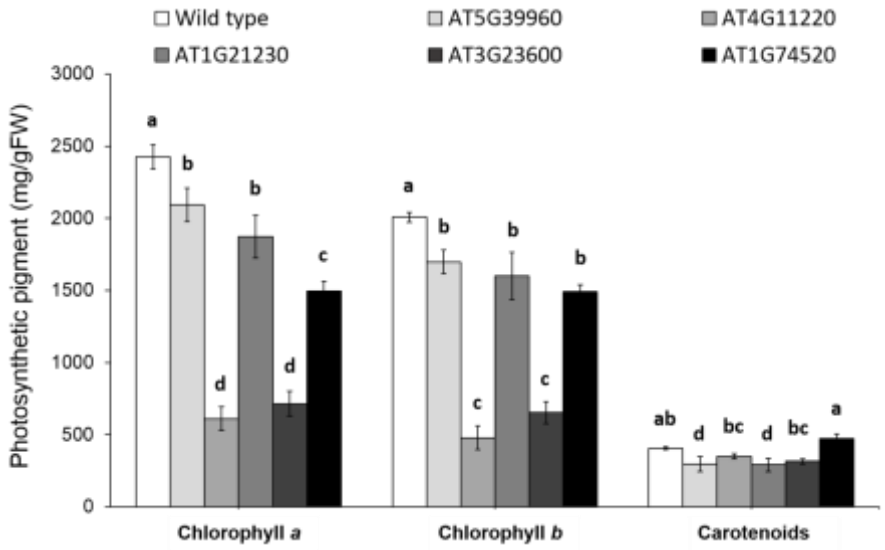


Figure 4 Photosynthetic pigments, chlorophyll *a*, chlorophyll *b*, and carotenoid contents in wild type and 5 mutants under normal conditions.

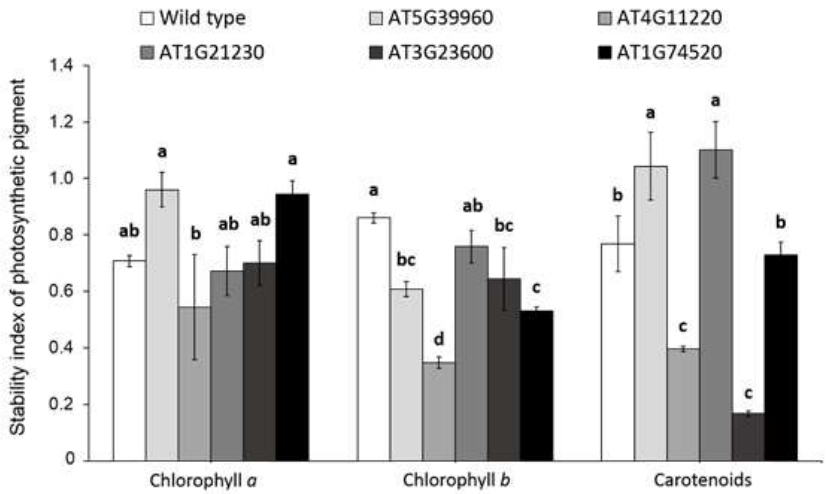


Figure 5 Stability indices of photosynthetic pigments, chlorophyll *a*, chlorophyll *b*, and carotenoid contents in wild type and 5 mutants after salt stress for 7 days

Salt stress causes the reduction of photosynthetic pigments (Munns and Tester, 2008) due to osmotic stress and ionic toxicity (Lakshmi *et al.*, 1996; Misra *et al.*, 1997). Therefore, the stability index of photosynthetic pigments of salt stress is lower than 1. In our experiment, most stability indices were less than 1, except the stability indices of carotenoids in AT1G21230 and AT5G39960 mutants. This may be because the level of carotenoid content in these mutants under normal condition is at the minimal level to be able to survive. Carotenoids are the accessory pigments that only function for light harvest, but also function in the ROS scavenging system to protect the photosynthesis system from photoinhibition (Taiz et al, 2015).

The hierarchy clustering of SNPs in 10 loci and the comparison with salt responsive phenotypes of 75 rice varieties.

When SNPs from 10 identified loci were used for hierarchy clustering, 75 varieties were clustered into 4 groups (Figure 6). IR29, which is the salt susceptible standard is in Group I, while Group II consists of three

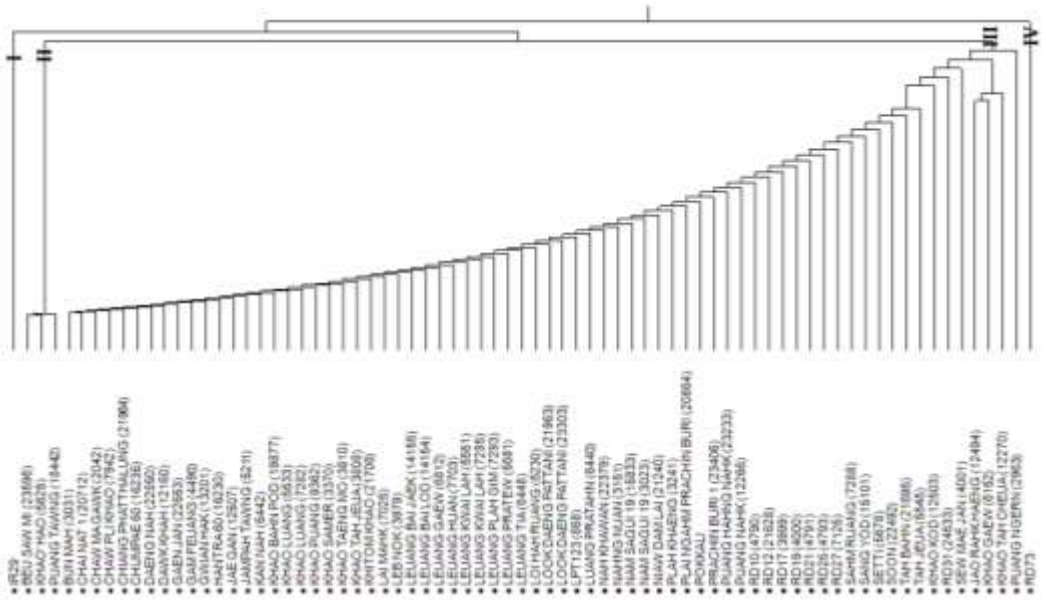


Figure 6 Hierarchy clustering of SNPs from 10 loci identified by GWAS with Na^+ content in leaves at seedling stage under salt stress condition

varieties, BEU SAW MI, KHAO' HAO, and PUANG TAWNG. BEU SAW MI and KHAO' HAO were classified in the most salt sensitive group according to the phenotype studies as shown in Table1, while PUANG TAWNG was classified in the second most salt susceptible group. The rest of the rice varieties in this study, except RD 73, was in Group III, including Pokkali, the salt tolerant standard cultivar. RD 73, which is the latest salt tolerant varieties developed by breeding program in Rice Department of Thailand is in Group IV.

REFFERENCES

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301: 653–657.

Asaoka R, Uemura T, Nishida S, Fujiwara T, Ueda T, Nakano A (2016) New insights into the role of Arabidopsis RABA1 GTPases in salinity stress tolerance. Plant Signal Behav 11: e1149676.

Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. Am J Hum Genet 98: 116–126.

Cayrol B, Delteil A, Gobbato E, Kroj T, Morelb JB (2016) Three wall-associated kinases required for rice basal immunity form protein complexes in the plasma membrane. Plant Signal Behav 11: e1149676.

Chabi K, Kefela T, Sagbadja H, Ahoton L, Saidou A, Moussa L, Baptistea L, Kotconia S, Gachomo E (2015) A simple and efficient genomic DNA extraction protocol for large scale genetic analyses of plant biological systems Plant Gene 1: 43–45.

Chen CN, Chu C, Zentella R, Pan Z, Tuan-Hua David (2002) AtHVA22 gene family in Arabidopsis:

phylogenetic relationship, ABA and stress regulation, and tissue-specific expression. Plant Mol Biol 49: 633–644.

Davenport R, James RA, Zakrisson-Plogander A, Tester M, Munns R (2005) Control of sodium transport in durum wheat. Plant Physiol 137: 807–818.

Dell'Acqua DM, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung H, Nelissen H, Baute J, *et al*. (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. Genome 16: 1–23.

Delteil A, Gobbato E, Cayrol B, Estevan J, Michel-Romiti C, Dievart A, Kroj T, Morel JB. 2016. Several wall-associated kinases participate positively and negatively in basal defense against rice blast fungus. BMC Plant Biol. 16:17.

Dixit S, Biswal AK, Min A, Henry A, Oane RH, Raorane ML, Longkumer T, Pabuayon IM, Mutte IM, Vardarajan AR, *et al*. (2015) Action of multiple intra-QTL genes concerted around a co-localized transcription factor underpins a large effect QTL. Sci Rep 5: 15183.

Gall HL, Philippe F, Domon JM, Gillet F, Pelloux J, Rayon C (2015) Cell wall metabolism in response to abiotic stress. Plants 4: 112–166.

Jouyban Z (2012) The effects of salt stress on plant growth. Technical Journal of Engineering and Applied Sciences 2: 7–10.

Kawahara Y, Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL, *et al*. (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice 6: 4.

Koyama ML, Levesley A, Koebner MD, Flowers TJ, Yeo AR (2001) Quantitative trait loci for component physiological traits determining salt tolerance in rice. Plant Physiol 125: 406–422.

Kumar V, Singh A, Amitha MSV, Krishnamurthy SL, Parida SK, Jain S, Tiwari KK, Kumar P, Rao AR, Sharma SK (2015) Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). DNA Res 22: 133–145.

Lakshmi A, Ramanjulu S, Veeranjaneyulu K, Sudhakar C (1996) Effect of NaCl on photosynthesis parameters in two cultivars of mulberry. Photosynthetica 32: 285–289.

Li H, Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25: 1754-60.

Lippold F, Sanchez DH, Musialak M, Schlereth A, Scheible WR, Hincha DK, Udvardi MK (2009) *AtMyb41* regulates transcriptional and metabolic responses to osmotic stress in Arabidopsis. Plant Physiol 149: 1761–1772.

Luan S, Lan W, Lee C (2009) Potassium nutrition, sodium toxicity and calcium signaling connections through the CBL-CIPK network. Curr Opin Plant Biol 12: 339–346.

Ma S, Bohnert HJ (2007) Integration of *Arabidopsis thaliana* stress-related transcript profiles, promoter structures, and cell-specific expression. Genome Biol 8: R49.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20: 1297-1303.

Misra AN, Sahu SM, Misra M, Singh P, Meera I, Das N, Kar M, Shau P (1997) Sodium chloride induced changes in leaf growth, and pigment and protein contents in two rice cultivars. Biol Plant 39: 257–262.

Moradi F, Ismail AM (2007) Responses of photosynthesis, chlorophyll fluorescence and ROS-scavenging stems to salt stress during seedling and reproductive stages in rice. Ann Bot 99: 1161–1173.

Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annu Rev Plant Biol 59: 651–681.

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

Patterson N, Price AL, Reich D (2006) Population structure and eigen analysis. PLoS Genetics 2: 2074–2093.

Quintero JM, Fournier JM, Benlloch M (2007) Na^+ accumulation in shoot is related to water transport in K^+-starved sunflower plants but not in plants with a normal K^+ status. J Plant Physiol 164: 60–67.

Rubio F, Gassmann W, Schroeder JI (1995) Sodium-driven potassium uptake by the plant potassium transporter HKT1 and mutations conferring salt tolerance. Science 270: 1660–1663.

Schmidt R, Schippers JHM, Mieulet D, Obata T, Fernie AR, Guiderdoni E, Mueller-Roeber B (2013) MULTIPASS, a rice R2R3-type MYB transcription factor, regulates adaptive growth by integrating multiple hormonal pathways.Plant J 76: 258–273.

Sevengor S, Yasar F, Kusvuran S, Ellialtioglu S (2011) The effect of salt stress on growth, chlorophyll content, lipid peroxidation and

antioxidative enzymes of pumpkin seedling. Afr J Agric Res 6: 4920–492.

Shen Q, Fu L, Dai F, Jiang L, Zhang G, Wu D (2016) Multi-omics analysis reveals molecular mechanisms of shoot adaption to salt stress in Tibetan wild barley. Genomics 17: 889–903.

Siringam K, Juntawong N, Cha-um S, Kirdmanee C (2011) Salt stress induced ion accumulation, ion homeostasis, membrane injury and sugar contents in salt-sensitive rice (*Oryza sativa* L. spp. *indica*) roots under iso-osmotic conditions. Afr J Biotechnol 10: 1340–1346.

Taiz L, Zeiger E, Moller IM, Murphy A (2015) Plant Physiology and Development. The sixth edition. Sinauer Associates.

Wellburn AR (1994) The spectral determination of chlorophyll *a* and *b*, as well as total carotenoid, using various solvents with spectrophotometers of different resolution. J Plant Physiol 144: 307–313.

Zeng L, Shannon MC, Lesch SM (2001) Timing of salinity stress affects rice growth and yield components. Agric Water Manag 48: 191–206.

Zhou X, Stephens M (2012). Genome-wide efficient mixed-model analysis for association studies. Nature Genetics. 44: 821-824.

Zhu JK, Hasegawa PM, Bressan RA, Bohnert PHJ (1997) Molecular aspects of osmotic stress in plants. CRC Crit Rev Plant Sci 16: 253–277.

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

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

**Corresponding author: saengton@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 Minghiu63 (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 F_1 pollen was stained with 2% IKI_2 solution. If F_1 had sterile pollen, male parents were identified as a maintainer line. In contrast, if F_1 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.edu/). 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 µl containing 20 ng of genomic DNA, 1X MYTAQ RED MIX (BIOLINE, UK), 0.125 µM of each inner primer, 0.5 µM of each outer primer. The PCR condition was as follows: 94 °C for 5 min, followed by 45 cycles of 94 °C for 30 sec, 60 °C for 15 sec and 72 °C for 30 sec, and final 72 °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.

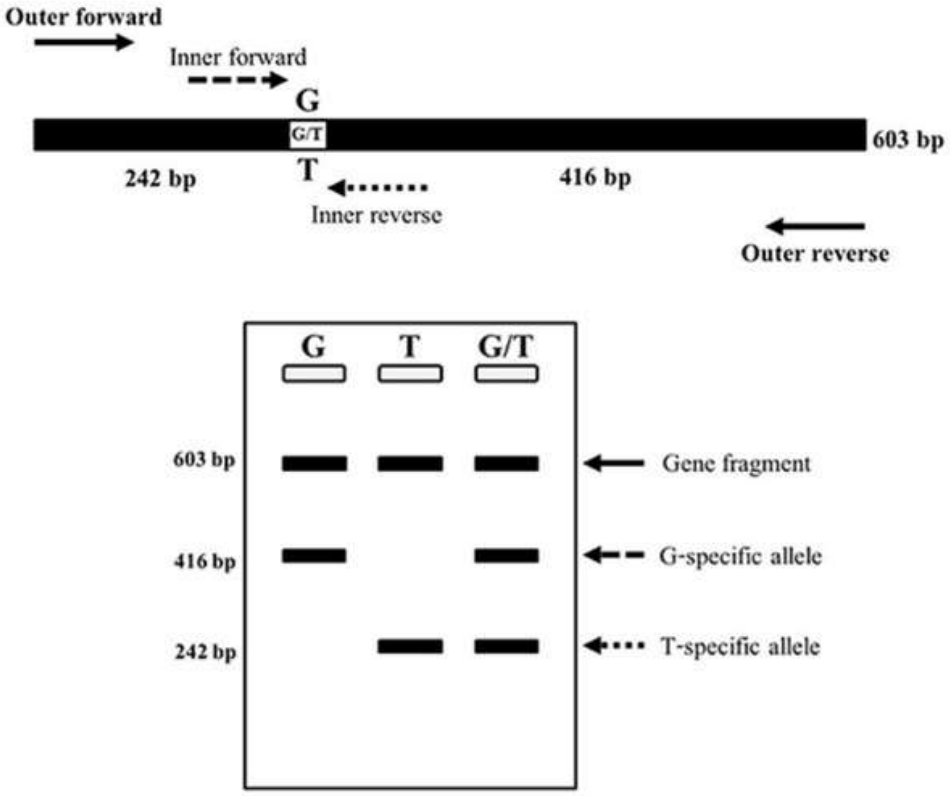


Figure 1. Diagram representation of tetra- primer ARMS-PCR for allele-specific SNP genotyping. Four primers are used: the two outer primers amplify a gene fragment that contains an SNP (white box). The inner primers are designed to amplify the G or T allele. The outer forward primer and inner reverse 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 DISSCUSSIONS

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)
PPR10 gene	PPR10-Forward	TGCTGCTGCACCTGTCAGC	-	2,200
	PPR10-Reverse	GCCGATTAGGGTAGTATCGGGG	-	
	Os10g0495100R	CCTCAGCCTTCTCCCATTTG	-	-
Tetra primers	PPR10- Inner forward	AGAAGGCTGAGGAGTTAATTTTTG	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 (G^{1392} to T^{1392}), 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)

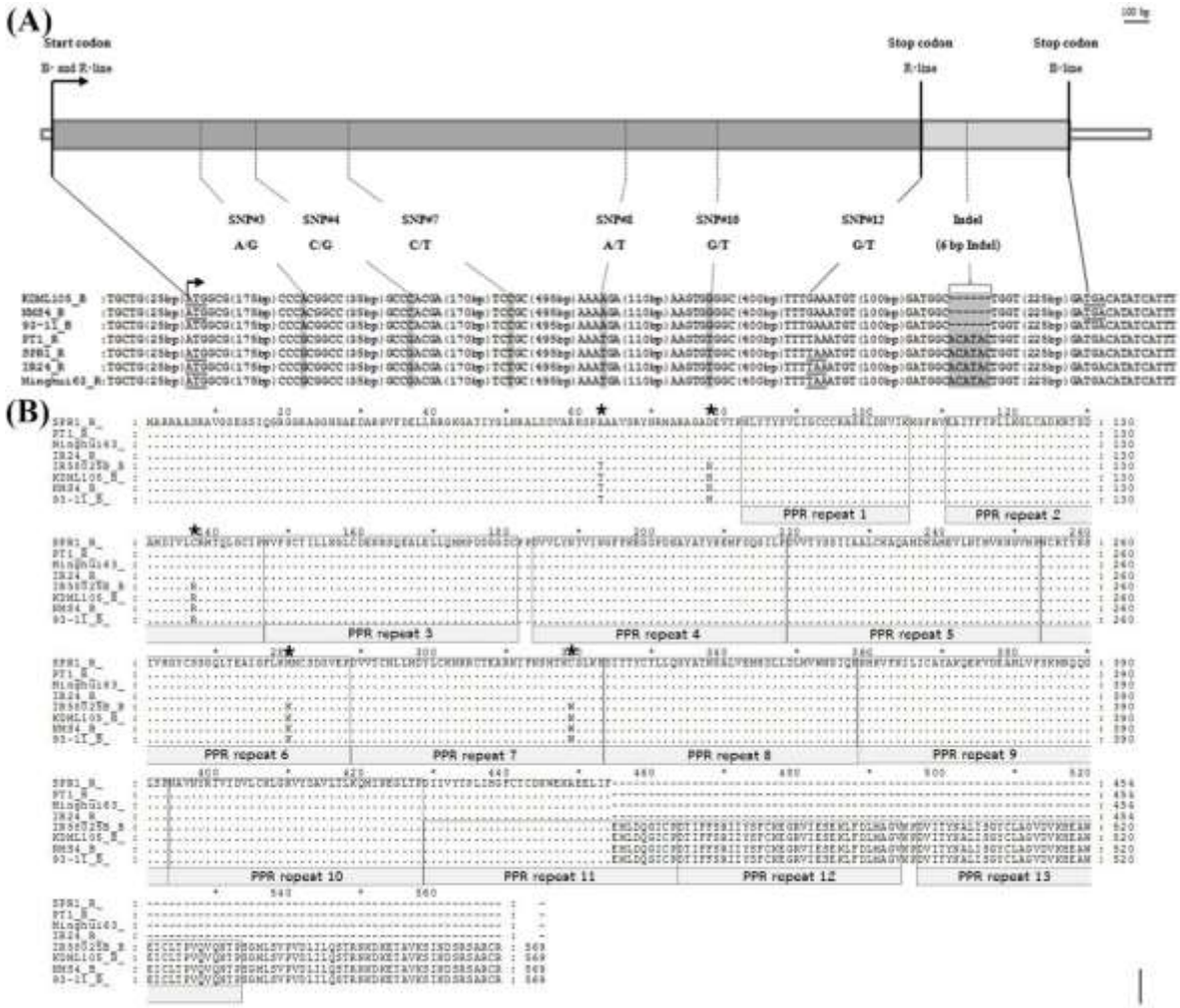


Figure 2. Nucleotide and amino acid sequence comparison of *PPR10* gene. The isolated *PPR10* genes were from KDML105, NMS4, SPR1, and PTT1 cultivars. The reported *PPR10* gene were from 93-11, IR24 and Minghui63. The KDML105, NMS4 and 93-11 were maintainers whereas SPR1, PTT1, IR24 and Minghui63 were 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 acids 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

Moreover, male sterile restore ability of seven rice cultivars were analyzed with *PPR10* gene as shown in Table 3. The result showed that G and T allele belong to maintainer and restorer lines, respectively. Therefore, this result suggested that $SNP_{1,392}$ could be associated with male sterile restoration ability and could be used as a gene marker for *PPR10* gene. As well as the SNP A/C at +474 in the CDS was converted to an allele-specific PCR marker for *PPR9* gene (Chen *et al.*, 2017).

Primer design and validation of allele specific SNP marker

The 1,170-1,780 bp region of IR24 *PPR10* gene sequence containing $SNP_{1,392}$ was used in tetra-primer ARMS-PCR primer design (Figure 3A) and the sequence of these four primers were shown in Table 1. PPR10-Outer forward and PPR10-Outer reverse primers were used to amplify gene fragment of 603 bp.

Table 3 Male sterile restore ability and *PPR10* gene of each rice cultivar.

Rice cultivar	Restore ability	SNP at position 1,392	Length (amino acids)
Nipponbare	Maintainer	G	569
93-11		G	569
KDML105		G	569
NMS4		G	569
IR24	Restorer	T	454
Minghui63		T	454
PTT1		T	454
SPR1		T	454

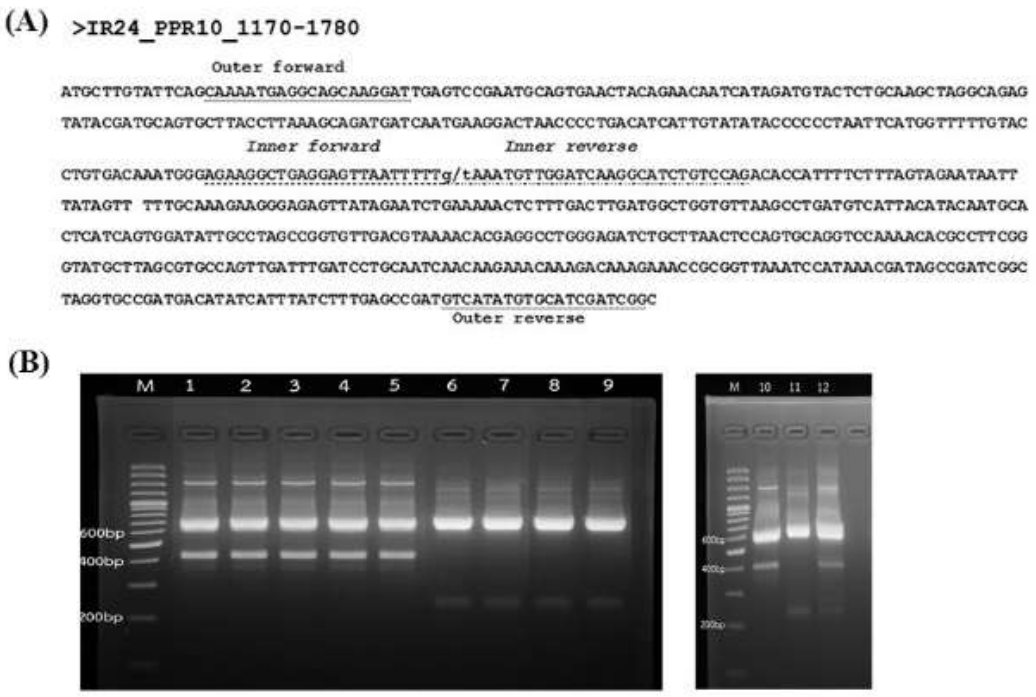


Figure 3. Primer design and validation of allele specific SNP marker. (A) The 1,170-1,780 bp region of IR24 *PPR10* gene sequence that was used in tetra primer-ARMS -PCR primer design, Outer forward, Outer reverse, Inner forward and Inner reverse primers. (B) Agarose gel electrophoresis for tetra-primer ARMS-PCR of five maintainer cultivars, four restorer cultivars and one hybrid of IR58025A x PTT1. Lane M was GeneRuler 100 bp DNA plus DNA ladder (Thermo Scientific). Lane 1-12 were PCR products from KDML105, JHN, NMS4, PG56, PTT1, SPR1, LDP, RD47, IR58025A, PTT1 and hybrid of IR58025A x PTT1, respectively. Lanes 1-5 were maintainers whereas lanes 6-9 were restorers. The G allele (maintainer allele) was 416 bp whereas T allele (restorer allele) was 242 bp.

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, Moukoumbi 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.

Genomics and Genetics
The Official Journal of Genetics Society of Thailand

Genomics and Genetics

The journal of Genetics Society of Thailand http://www.tci-thaijo.org/index.php/gst

Author Guidelines

Genomics and Genetics is published by the Genetics Society of Thailand.

Genomics and Genetics is listed in Thai-Journal Citation Index Center (TCI: http://www.kmutt.ac.th/jif/public_html/index.html) group 1.

The Journal considers contributions that present the results of original research in genetics, biochemistry, molecular biology and related scientific disciplines.

Although Genomics and Genetics is an official publication of the Genetics Society of Thailand, contributors are not required to be members of the Society.

It is a fundamental condition that submitted manuscripts have not been and will not be published elsewhere.

Categories of Contribution: Research Articles, Review Articles and Short Communications

Manuscripts considered in conformity with the scope of the journal, judged by the Editor in conjunction with the Editorial Board, are peer reviewed by two or more external reviewers. Acceptance by the Editor is based on the quality of the work as substantial contribution to the field and on the overall presentation of the manuscript.

Genomics and Genetics follows an Open-Access policy. Articles are made available in full content at http://www.tci-thaijo.org/index.php/gst. Back issues (Thai Journal of Genetics) dating to 2008 are available through this site.

The official abbreviation for Genomics and Genetics is Genomics Genet.

How to submit:

1. Manuscripts must be submitted by the Corresponding Author who will correspond/communicate with Editor for revision during the editorial process. Registration for submission is available at http://www.tci-thaijo.org/index.php/gst/user/register

2. The cover letter should be addressed to: Arinthip Thamchaipenet, Editor-in-Chief, Genomics and Genetics. It must be stated by the corresponding author that the data have not been published and are not under consideration elsewhere, and that all authors have approved the submission of the manuscript. The names of all co-authors, including institutional affiliations and e-mail addresses must be included. Furthermore, up to three reviewer names should be suggested with valid contact e-mail addresses.

3. Manuscripts can be written following the manuscript templates available at http://www.si.mahidol.ac.th/th/department/research/GaG_Format.doc

4. For further enquiry, please contact Arinthip Thamchaipenet, Editor-in-Chief, Genomics and Genetics at arinthip.t@ku.ac.th



สมาคมพันธุศาสตร์แห่งประเทศไทย
Genetics Society of Thailand

Genomics and Genetics

Contact Office :
Division of Molecular Medicine,
4th floor, Siriraj Medical Research Building, 2 Wang Lang Road,
Siriraj, Bangkoknoi, Bangkok, 10700, Thailand
Phone : (66-2) 419-2754-59 Fax : (66-2) 411-0169
E-mail : gesoth@gmail.com
E-Journal : http://www.tci-thaijo.org/index.php/gst
A peer-reviewed journal published by Genetics Society of Thailand