Omonrice 13: 12-18 (2005)

# TAG GENES CONTROLLING GRAIN PROTEIN CONTENT (GPC) USING MICROSATELLITE MARKERS IN RICE (*Oryza sativa* L.)

**Nguyen thi Lang, Bui Chi Buu**

**ABSTRACT**

*This study was undertaken with a view to tag gene(s) controlling grain protein content (GPC) using molecular markers in rice. Genotype IR64 with low protein (7.5%) was crossed with genotype Nang thom Cho Dao with high protein (10.62%). Then 149 individuals from BC3F2 population were derived. The parental genotypes and $BC_3F_2$ population were analyzed with RM234. Locus RM234 showed association between the marker and target gene for GPC. This was further confirmed through selective genotyping. The co-segregation data on the molecular marker and protein content on 149 $BC_3F_2$ was analyzed by mean of a single marker linear regression approach. It showed that this marker linked to QTL accounted for 18.1% of the variation of GPC between two parents. Significant regression suggested linkage between RM234 and a QTL for GPC on chromosome 7.*

**Key words**: Grain protein content (GPC), Microsatellite, QTL analysis

## INTRODUCTION

Protein content in rice grain (*Oryza sativa* L.) is an important trait for health of people whose main food in daily life is rice (Chunhai Shi et al. 1999). In order to improve the efficiency of breeding for rice nutrient quality, understanding variation of gene expression in different environments is necessary. Studies have shown that protein content of rice is a quantitative trait (Singh & Singh 1982; Shenoy et al. 1991). HilleRisLambers et al. (1973) pointed out that the heritability for protein content was 0.130–0.372 due to genotype x environment (GxE) interaction effect. Results by Shenoy et al. (1991) indicated that heritability of protein content was about 58.8%. Lang et al. (2005) found that protein content was controlled by genetic effects with heritability of 25.9%. Through inheritance studies carried out earlier, we noticed that difference in GPC between two parents was due to two major genes, even though a number of QTLs *per se* may control the trait. In recent years, the potential of molecular marker-assisted selection in plant breeding has been demonstrated in several crops (Huang et al. 1997; Ichikawa et al.1997; Reddy et al. 1997, Lang et al 1999, 2000, 2001, 2004). However, its utility in the hands of plant breeders has not demonstrated yet in protein, even though molecular markers associated with dozens of genes controlling several traits of economic importance have been developed in this crop.

The improvement in grain protein content (GPC) and its composition in protein has been a major concern of plant breeders. It has been difficult to achieve for effective selection criteria and because selection is expensive and time-consuming. In this repots, development of one or more molecular markers to be used for indirect selection for protein content / composition should be a convenient alternative.

## MATERIALS AND METHODS

Two genotypes differing for GPC, namely Nang thom Cho Dao (high GPC) and IR64 (low GPC), and a set of 96 $BC_2F_2$ derived from these parents were used in the present study. The $BC_2F_2$ population was developed at Cuu Long Delta Rice Research Insitute (CLRRI).

The low protein genotype was crossed to the high protein one. Subsequent $F_1$ was self-crossed and backcrossed to the recurrent to produce the $F_2$ and backcross generations,

respectively. The parents, $F_1$, $F_2$, $BC_1$ and $BC_3$ (recurrent parent is IR64). The promising $BC_1$ plants selected for desirable phenotypic traits were backcrossed to IR64 to generate $BC_2$ plants, and so on. 149 $BC_3F_2$ were evaluated due to protein analysis. Transgressive segregants were observed for high and low protein. The individuals of $BC_3F_2$ family were screened using microsatellites.

*Screening parents and $BC_3F_2$'s individuals for GPC*

Protein content in seeds harvested on individual plants was determined with the emphasis on dry weight basis. Nitrogen content in each grain sample was noticed by using Technicon autoanalyser (Warner and Jones 1970) then converted into protein value as following formula:

Protein content (%) =Nitrogen content (%) x 5.7

*CTAB extraction DNA*

Leaf tissues (4-6 gr) were grinded in liquid nitrogen, and transferred into 50 ml tube (10-15 ml leaf power). The finer the grind gives the higher the yield, but attentions should be paid to prevent the tissues from not thaw.

20 ml 2X CTAB extraction buffer (preheated to 65°C) were added to the leaf power, they were mixed thoroughly and incubated at 65°C for 30 min (up to 1 hour). Tubles were removed from water bath and let cool briefly. 20 ml 24:1 $CHCl_3$: isoamyl alcohol were added, then incubated at room temperature with moderate shaking for 20min. they were centrifuged at 3000 rpm for 30 min at 4°C. Supernatant (upper phase) was transferred into 50ml tube with filtering through micracloth. One vol (20ml) isopropyl alcohol was added, then mixed by inversion. They were incubated at room temperature or colder (-20 °C) for 1 hour to overnight, then centrifuged at 3000 rpm for 30 min at 4°C. Supernatant was drained away from pellet, then pellet was washed with 70% EtOH. Pellets were in air dry and dissolved nucleic acids in 5 ml TE. Tubles were placed at 65°C to help dissolution. 1/10 vol 3M Na Acetate (500 µl) and 2 vol Absolute EtaOH (10-11ml) were added, then they were incubate at –20°C for one hour (or longer) at this step it is possible to leave the samples overnight at –20°C

Centrifuge at 3000rpm for 30 min was done, then wash with 70% EtOH, air dry, dissolve in 500 µl TE, transfer microfuge tube, rinse 50 ml tube with 200 µl TE, collect wash rest of solution.

*Amplification of microsatellites and detection of their polymorphism*

The nucleotide sequences of the primer pairs used for each microsatellite are identified. PCR amplification were performed in 10mM Tris-HCl (pH8.3) 50mMKCl, 1.5mM $MgCl_2$, 0.5 or 1 unit of *Taq* polymerese, 4 nmole of dNTP, 10 pmole of primer with 10ng of genomic DNA per 20µl using a thermal cycler. Thrity five PCR cycles were performed with 30s of denaturation at 94°C, 30s of anaealing at 55°C and 1 min of polymerization at 72°C. Polymorphism in the PCR products was detected by ethidium bromide staining after electrophoresis in 3% agarose gel.
Amplification products were obtained by using RM234 primers, developed by Cornell University, USA.
Forward primer: 5'-
ACAGTATCCAAGGCCCTGG-3'
Reverse primer: 5'-
CACGTGAGACAAAGACGGAG-3'

*Construction of SSR map and assignment of linked groups to chromosomes*

A set of SSR markers, which were present in the test population, was identified to construct the SSR map and to assign the linked groups in chromosomes. Linkage groups were ordered by using MAPMARKER (Lander et al 1987). Linkage groups were reconfirmed by using the "GROUP". Map units (cM) were derived by using the Kosambi function (Kosambi 1944).
*QTL analysis*

Single-marker QTL analysis using linear regression was applied (Tinker 1996). The marker alleles *hp* and *lp* were coded 1 and 0, respectively for conducting regression analysis.

To identy major gene(s) and QTLs for protein, two different analysis were performed. The data were analyzed by using MAPMARKER to locate genes for GPC. For QTLs, interval analysis was conducted with MAPMARKER/QTL. The threshold for declaring a QTL for salinity tolerance was at LOD> 3.0. Interval mapping developed by Lander and Botstein (1989) is able to define the most likely position of a QTL and precisely estimaste the phenotypic effect of the QTL if it does not lie exactly at the marker locus. Log of odds peaks for each significant QTL was used to position the QTL on the SSR map. The proportion of the total phenotypic variation explained by each QTL was calculated as $R^2$ value ($R^2$ = ratio of the sum of square explained by the QTL to the total sum of square)

## RESULTS AND DISCUSSION

### GPC evaluation in BC population

One hundred fourty nine $BC_3F_2$ were developed from the cross IR64 / Nang thom Cho dao (NTCD). The $BC_2F_2$ individuals were raised in a replicated trial at CLRRI. Their GPCs were measured by Technicon autoanalyser. Nang Thom Cho Dao's GPC obtained 10.62% significantly different from IR64 (7.5%). The GPC in the $BC_3F_2$ ranged from 6.5% to 11.5%. A frequency distribution curve was formed (Fig. 1) and chi-square to test the goodness of "t to normal distribution" was gained. The data suggested a very high probability for a good "t" (P=0.001) to the normal distribution. Based on an earlier study of inheritance of GPC, it was predominantly controlled by partial dominance gene action with additive effect (Lang ang Buu 2005). The present study based on a normal distribution in $BC_3F_2$ confirmed that GPC maybe controlled by polygenes located at different loci.

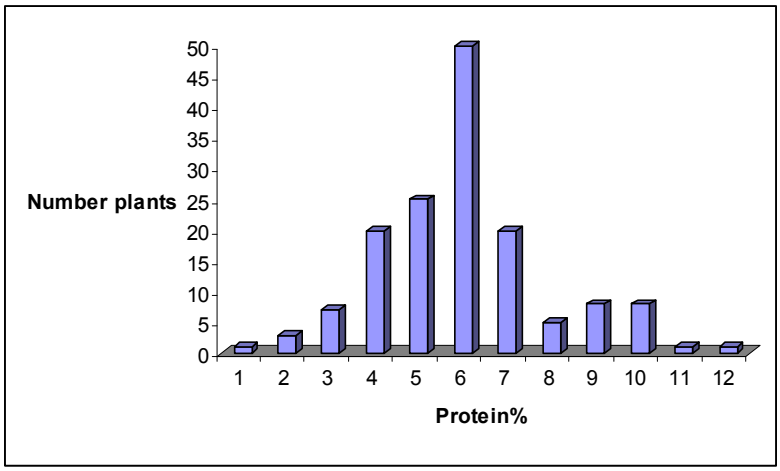


Figure 1: Frequency distribution of grain protein content in $BC_3F_2$ showing a goodness of "t" to the normal distribution

### Marker identification

From 20 SSR primer pairs used for detection of polymorphism between the two parental genotypes, we received scorable amplification products. Of these primers, 15 were detected as polymorphism between two parental genotypes. Using these primers, we conducted bulked segregant analysis on two pooled DNA samples, each consisting of 149 $BC_3F_2$ individuals and representing the two tails of the normal distribution.

### For the cross IR64 / NTCD

Polymorphic markers accounted for 92% of the SSRs surveyed in IR64 and NTCD. The $BC_3F_2$ population was also surveyed using the 15 SSRs. The expected segregation ratio would be 97.5% homozygote, 2.5% heterozygote. Resulting in an allele frequency

of 55.5% for IR64, 37.5% for NTCD was due to skewed allele frequencies at 44 out of 150 (31.3%) of marker loci. Almost markers have detected their linkages when DNA survey on the population, except RM501, RM 477.

**Construction of genetic linkage map**

Grouping of markers was carried out by a two point linkage analysis with a LOD score of 4 and a recombination fraction D of 0.3

For the $BC_3F_2$ of IR64 / NTCD, fifteen markers were used to construct a framework map with a total length of 76.1 cM on chromosome 7 (Figure 2)

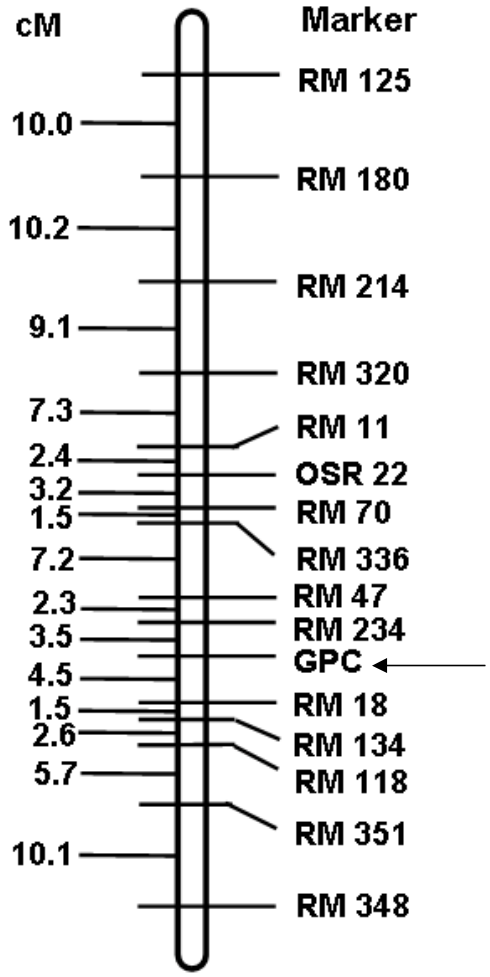


Figure 2: Molecular linkage map on chromosome 7 based on $BC_3F_2$ of IR64 / NTCD

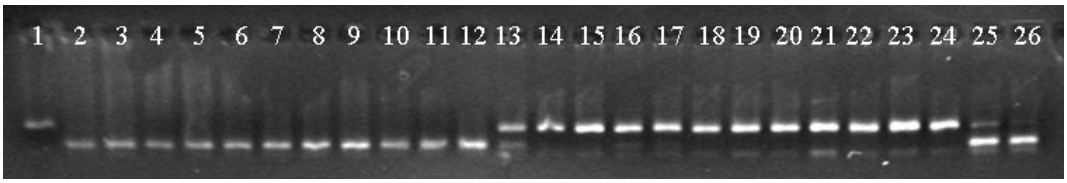


Figure 3: PCR products of segregating individuals amplified by using RM234. Lane 1: IR64 and Lane 2: NTCD, Lane 3-26: $BC_3F_2$

The solitary remaining RM234 *primer* exhibited amplification product with size band (163 bp) concerning to high and low GPC parents in the corresponding bulked segregant analysis (fig. 3). This suggested an association between this marker and GPC. To further confirm this association, we carried out selective genotyping (Lander and Botstein 1989) of individual $BC_3F_2$. The results revealed that out of two $BC_3F_2$s belonging to high protein pool, eight showed a profile similar to that of the high protein parent, out of ten $BC_3F_2$s belonging to low protein pool, twenty exhibited a profile similar to that of the low protein parent. This confirmed an association between the RM234 marker and GPC. Subsequently, all $BC_3F_2$s were genotyped by using the above SSRs, and the data on segregation of the marker was recorded for conducting

**QTL analysis and gene effects**

Since the GPC data of $BC_3F_2$ conformed with a normal distribution (Fig. 1), the data on genotypes of these $BC_3F_2$ at the locus RM234 were considered for QTL analysis using single-marker linear regression approach (Tinker 1996). The regression of protein content on the RM234 marker was highly significant indicating a linkage between the molecular marker and a QTL for protein content. The $R^2$ value of 0.1712 suggested that RM234 linked to QTL contributed to 17.12% of the total variation in protein content among the $BC_3F_2$ (Fig. 3). These results suggested that the marker RM234 may either be tightly linked to a QTL with a small effect or loosely linked to a QTL with a large effects (Prasad et al.1999). In rice, genes for protein content have been located on chromosome 7. Recently, through the cross IR64 / Khao Daw Mali 105, we also observed QTLs for grain protein content located on chromosome 7 and a major QTL accounting for 15.12% variation in GPC (Lang et al. 2005).

**CONCLUSION**

The genotype NTCD with high protein content (10.62%) was crossed with IR64 with lower GPC (7.5%) to develop 149 $BC_3F_2$s. The population exhibited a normal distribution for GPC. Of 20 SSR primers for DNA survey, 15 polymorphic markers gave scorable amplification products and distinguished high GPC from low one. We carried out bulked segregant analysis on 149 $BC_3F_2$s representing the two extremes of the distribution. At the locus RM234 on chromosome 7, we noticed association between the target gene(s) and DNA marker. This was further confirmed through selective genotyping.

**REFERENCES**

HilleRisLambers D, JN Rutger, CO Qualset, WJ Wiser.1973. Genetic and environmental variation in protein content rice (*Oryza sativa* L.). Euphytica 22: 264–273.

Huang N, ER Angeles, J Domingo, G Magpantay, S Singh, G Zhang, N Kumaravadivel, J Bennett, GS Khush. 1997. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor Appl Genet 95 : 313}320

Lander ES, D Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121 : 185}199

Lang NT and BC Buu. 2004. Molecular genetic analysis and marker assisted selection for restorer line and bacterial blight resistance in hybrid rice. SABRAO Journal of breeding and Genetics . Vol 36 (2) 83-93.

Lang NT, BC Buu. 2005 . Genetic analysis of grain protein content in rice. Tap chi Nong Nghiep & PTNT 12: 14-20

Lang NT, PKSubudhi, SS Virmani, DS Brar, GS Khush, Z Li, N Huang. 1999. Development of PCR based marker for thermosensitive genetic male sterility gene tms 3 (t) in rice ( Oryza sativa L.) Hereditas 131:121-127.

Lang NT, S Yanagihara, BC Buu. 2001. A microsatellite marker for gene conferring salt tolerance on rice at the vegetative and reproductive stages. SABRAO Journal of breeding and Genetics .Vol 33(1) 1-10 .

Prasad M, RK Varshney, A Kumar, HS Balyan, PC Sharma, KJ Edwards, H

Singh, HS Dhaliwal, JK Roy, PK Gupta. 1999. A microsatellite marker associated with a QTL for grain protein content on chromosome arm 2DL of bread wheat. Theor Appl Genet 99:341–345

Reddy JN, MR Baraoidan, MA Bernardo, MLC George, R Sridhar. 1997. Application of marker assisted selection in rice for bacterial blight resistance. Curr Sci 73 : 873}875

Shenoy VV, DV Seshu, JKS Sachan. 1991. Inheritance of protein per grain in rice. Indian J Genet 51(2): 214–220.

Shi C, J Zhu, X Yang, Y Yu, J Wu..1999. Genetic analysis for protein content in *indica* rice. Euphytica 107: 135-140.

Singh NB, HG Singh. 1982. Gene action for quality components in rice. Indian J Agric Sci 52(8): 485–488.

Tinker NA. 1996. New strategies for QTL mapping. Barley workshop (no. 15) at the Plant Genome IV Conf. (http://www.css.orst.edu/research/barley/qtl)

Vu hieu Dong. 2005. A microsatellite marker associated with a QTL for grain protein content on chromosome 7. *MSc Thesi on Biotechnology Course.* Can Tho University, Can Tho, Vietnam.

Warner MH, J Warner. 1970. A rapid method for nitrogen determination in plant tissue. Commun Soil Sci Plant Anal 1: 106-114

**Đánh dấu gen điều khiển hàm lượng protein trong hạt thóc bằng microsatellite**

Hàm lượng protein trong hạt (GPC) do đa gen điều khiển được nghiên cứu trên tổ hợp lai giữa IR64 (GPC=7,5%) và Nàng Thơm Chợ Đào (GPC=10,62%) để phát triển quần thể $BC_3F_2$ với 149 cá thể chọn lọc. Phân tích protein trong hạt theo phương pháp đo lường N tổng số và suy theo công thức để biết hàm lượng protein. Phân bố chuẩn được ghi nhận trên quần thể hồi giao $BC_3F_2$ cho thấy tính chất đa gen được khẳng định. Phân tích QTL theo phương pháp SMA được tiến hành. Biến di di truyền của GPC được QTL giải thích 18.1% theo kiểu hình. Marker RM234 định vị trên nhiễm sắc thể số 7 được khẳng định liên kết khá chặt với gen mục tiêu.