Genetic Manipulation of Folate in Rice

Nampeung Anukul, BSc, MSc

A thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy, August 2009.

ABSTRACT

Folate malnutrition is a major problem in many countries around the world especially in Asia and Africa. Stable foods such as rice contain very low amounts of folate. White rice which is the most popular form for human consumption contains less than 80 μg folate per 100 g. Given that it forms a major part of the South East Asian diet, rice represents an important target for enhancing folate levels.

The objectives of this thesis is study the mechanisms that regulate total folate levels in rice grains and attempt to enhance and stabilise folate in rice endosperm. Three main strategies were adopted. First, the natural variation of folate biosynthesis gene expression was probed using RT-qPCR. Second, functional genomic approaches were used to manipulate the activity of rice folylpolyglutamate synthetase (FPGS), the enzyme which adds glutamate residues to folate. Third, genetic engineering was used to express FPGS enzymes and mammalian folate binding proteins in rice endosperm.

RT-qPCR revealed that the variation in folate biosynthesis transcript abundance was closely correlated with total folate levels among rice varieties. High transcript abundance of all folate biosynthesis genes was associated with high total folate levels in Moroberekan rice mature seed. Comparative genomic studies revealed that rice FPGS is encoded by two distinct genes, *FPGS* Os03g02030 and *FPGS* Os10g35940. Transcript abundance of *FPGS* Os03g02030 appeared higher than that of *FPGS* Os10g35940 in seed, whilst, transcript abundance of *FPGS* Os10g35940 was higher in leaf.

To determine the function of the *FPGS* Os03g02030 gene in rice seed, a T-DNA knock out line was characterised. Disrupting Os03g02030 gene expression resulted in delayed seed maturation and decreased mono- and polyglutamylated folate pools in mutant seed. RT-qPCR detected an increase in the transcript abundance of folate biosynthesis genes in seed of the knock out plant, whereas the folate deglutamylating enzyme γ-glutamyl hydrolase (GGH)

mRNA level was reduced. A potential feedback mechanism to maintain folate abundance during rice development was uncovered through the alternative functional *FPGS* Os10g35940 activity and reduction of folate breakdown.

Protein-bound folate forms are better protected from oxidative degradation resulting in greater folate stability (Suh *et al.* 2001). Two rice *FPGS* and mammalian folate binding proteins was successfully introduced into rice endosperm using *Agrobacterium* based transformation in an attempt to retain and stabilise folate pool within rice endosperm. Analysis in terms of folate abundance and bioavailability will form part of future studies.

ACKNOWLEDGEMENTS

First of all, I would like to express my great thanks to the Royal Thai government for the scholarship supporting my PhD study. It is a great opportunity leading me to the UK and achieving the greatest experiences.

Most sincere gratitude and appreciation are expressed to my supervisor, Professor Malcolm Bennett, for his valuable guidance, continuous discussion and encouragement throughout my study. His kindness and understanding will be memorized forever.

This thesis would not have been completed successfully, without the help from many students, post-doc and staffs from Lab A29 and Plant and crop sciences department as well as the help from collaborators in CIRAD (Montpellier, France) and Dominique Van Der Straeten's group (Ghent University). I wish to express my gracious thanks for their valuable lab guidance and all support.

At last but most important, I would like to express my deepest gratitude to my family as well as my dear friends in Nottingham for their love, understanding and encouragement; not only for academic inspiration but also for sharing all happiness and sadness together throughout my PhD life. Thank you so much, indeed.

LIST OF FIGURES AND TABLES

ABBREVIATIONS

°C	Degree celcius
AdoHcy	Adenosylhomocysteine
AdoMet	Adenosylmethionine
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
cFBP	Cow's milk folate binding protein
DHF	Dihydrofolate
DHFR/TS	Dihydrofolate reductase/thymidylate synthase
DHFS	Dihydrofolate synthase
DHNA	Dihydroneopterin aldolase
DHNTP	Dihydroneopterin triphosphate
DHPR	Dihydropteridine reductase
DHPS	Dihydropteroate synthase
FBP	Folate binding protein
FPGS	Folylpolyglutamate synthetase
FST	Flanking sequence tag
GDC	Glycine decarboxylase complex
Glu	Glutamate
GNMT	Glycine N-methyltransferase
GTP	Guanosine 5'-triphosphate
GTPCHI	GTP cyclohydrolase I
GTPCHI	GTP cyclohydrolase I
GUS	β-glucuronidase
$H_4PteGlu_n$	5,6,7,8-tetrahydropteroyl-γ-glutamate
HCOOH	Formate
Hcy	Homocysteine
HPLC	High-performance liquid chromatography
HPPK/DHPS	Hydroxylmethyldihydropterin pyrophosphokinase/Dihyropteroate synthase
Kd	Dissociation constant

LCMS/MS	Liquid chromatography tandem mass spectrometry
Met	Methionine
MTHFR	Methylene tetrahydrofolate reductase
NTDs	Neural tube defects
*p*ABA	*p*-aminobenzoic acid
Pi	Inorganic phosphate
RDAs	Recommended dietary allowances
RNA	Ribonucleic acid
rpm	Rounds per minute
RT-qPCR	Reverse transcriptase-quantitative polymerase chain reaction
RU	Resonance units
s	Second
SAH	S-adenosylhomocysteine hydrolase
SAM	S-adenosylmethionine
SD	Standard deviation
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SHMT	Serine hydroxymethyltransferase
SPR	Surface plasmon resonance
THF	Tetrahydrofolate
tRNA	Transfer RNA
TS	Thymidylate synthase

TABLE OF CONTENTS

CHAPTER 1 GENERAL INTRODUCTION

1.1 GENERAL INFORMATION ABOUT FOLATE

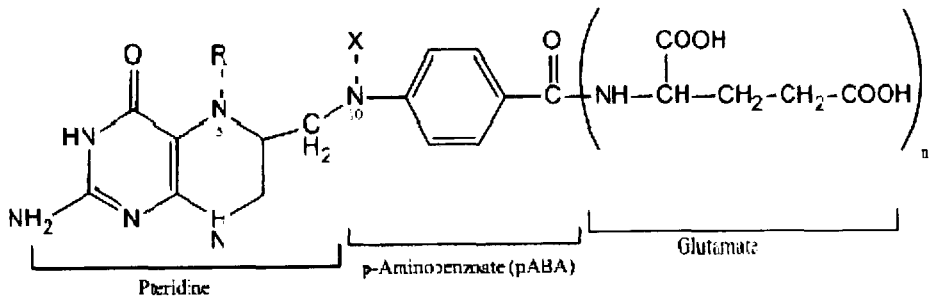
Vitamins are organic molecules required in small amounts for metabolic processes (Stryer, 1995). Vitamins are classified by their physiological and chemical properties into two groups, fat soluble (A, D, E, K) and water soluble (C and B-complex) vitamins. The vitamin B complex mainly act as coenzymes or cofactors of metabolic pathways. For example, vitamins such as biotin are directly involved in the synthesis of essential compounds involving in fatty acid biosynthesis. Vitamins can also function as hormones (vitamin D) or antioxidants (vitamin E).

Animals are unable to synthesize all the vitamins they need, so they have to obtain these from outside sources such as animal products, fruits and vegetables. As most vitamins are not stored in the body, a consistent daily source is required. The requirement of each vitamin depends on age, gender, and other factors such as pregnancy and diseases. The recommended uptake amount of each vitamin has been established by the Food and Nutrition Board (FNB) of the National Academy of Sciences as RDA for people in different ages and condition. Recommended dietary allowances (RDAs) are defined as the levels of essential nutrient intake that practically meet the nutritional needs of all healthy people, average in a particular age and gender group (Bender, 2003). Although the nutritional requirement is different among types of vitamins and ethnic groups, the best way to reach the daily requirement is to have a balanced diet. Folate or vitamin B9 is the focus of this thesis.

1.1.1 Folate structure

The folate molecule consists of a pteridine ring, *p*-aminobenzoic acid (*p*ABA) and glutamate moieties of a variable length with 1-11 residues depending on the source (Figure 1.1) (Hanson and Gregory, 2002). One-carbon (C1) unit

such as formyl (-CHO), methylene (-CH_2-) and methyl (-CH_3) can be attached either at N_5 of the pteridine ring or at N_{10} of the *p*ABA molecule or bridged between these two units. Those folate derivatives, for example, 10-formyl tetrahydrofolate, 5-methyltetrahydrofolate and 5,10-methylenetetrahydrofolate, have ability to carry the one-carbon units and transfer to other molecules. These folate derivatives are able to interconvert within cells (Appling, 1991). A complex mixture of many forms of tetrahydrofolate monoglutamate and polyglutamate could be detected in extracellular and intracellular, respectively, according to the specific requirement of cells.



folate derivatives	R	X
THF	H	H
5-methyl-THF	CH3	H
5-formyl-THF	CHO	H
5,10-methylene-THF	CH2	CH2
5,10-methenyl-THF	CH	CH
10-formyl-THF	H	CHO

Figure 1.1 Chemical structures of tetrahydrofolate and its C1 substituted derivatives

1.1.2 Folate sources and human requirement

The RDA for folate is 400 micrograms per day for adults (except pregnant or lactating women) and children over 4 years of age, whilst pregnant and lactating women require 600 micrograms per day (IOM, 2005, www.iom.edu). Table 1.1 presents RDA for folate in various groups of people.

Table 1.1 Recommended Daily Allowance (RDA) for folate (modified from the complete set of dietary reference intakes table, www.iom.edu, July 2007)

Life stage group	Folate (µg/day)
Children	
1-3 year	150
4-8 year	200
Males	
9-13 year	300
14-18 year	400
19-30 year	400
31-50 year	400
51-70 year	400
>70 year	400
Females	
9-13 year	300
14-18 year	400
19-30 year	400
31-50 year	400
51-70 year	400
>70 year	400
Pregnancy	
14-18 year	600
19-30 year	600
31-50 year	600
Lactation	
14-18 year	500
19-30 year	500
31-50 year	500

Table 1.2 Folate content in foods (modified from Bekaert *et al.* 2007)

Crop	Folate content (µg/100 g)
Rice (white, raw)	6-8
Wheat (hard, white, raw)	38-43
Maize (yellow, seeds, raw)	19
Tomato (fruit, ripe, raw)	9-29
Peas (green, raw)	65
Spinach (leaves, raw)	194
Beans (mature seeds, raw)	463
Lentils (mature seeds, raw)	433

To prevent the diseases caused by folate deficiency, people must obtain an adequate folate intake to reach the minimum daily requirement. The major sources of folate are asparagus, avocados, broccoli, whole grains, spinach, tomatoes and oranges (IOM, 2005). Table 1.2 indicates the folate content in each food. Although folate is found in most plant foods, some staple crops such as rice and wheat are poor in this vitamin. Raw white rice appears to have the lowest level, only 10 microgram per 100 gram. As folate is sensitive to heat, light and oxidation conditions, the folate level in cooked white rice tends to be lower than this level.

1.1.3 Folate-mediated one-carbon metabolism in eukaryotes

Even though folate is required daily in a small amount (in microgram), its requirement in target tissues is high. Polyglutamylated folate forms are the major cellular active coenzymes required by most folate-dependent enzymes acting as the donor and acceptor for one-carbon unit (Shane, 1995; Scott *et al.*, 2000). However, methionine synthase enzyme needs monoglutamate folate for methionine synthesis (Scott *et al.*, 2000). Hence, monoglutamate folate is mostly found extracellular, in the intestinal tract and blood plasma of mammalian cells, and is the form transported across cells (Suh *et al.*, 2001).

In mammalian cells, small intestine is a major path of ingested polyglutamate folates which are first converted to monoglutamate folate forms in either methyl- or methylene- and formyl-derivatives depending on folate sources by brush border conjugase and then absorbed through the intestinal membrane and transported into the liver. Absorbed monoglutamate folate forms are then converted to 5-methyltetrahydrofolate monoglutamate in the liver, then transported in the blood to target tissues (Scott *et al.*, 2000).

Similarly, folic acid which can be absorbed directly into cells will be reduced by dihydrofolate synthetase (DHFS) to dihydrofolate and by dihydrofolate reductase (DHFR) to tetrahydrofolate (van der Put *et al.*, 2001). Monoglutamate folate taken into cells is sequentially glutamylated into the active coenzymes.

Generally, polyglutamylated folates are synthesised through a polyglutamylation process catalysed by the folylpolyglutamate synthetase (FPGS) enzyme. This enzyme is responsible for the intracellular conversion of monoglutamate folate into polyglutamylated folates in most organisms. Polyglutamylated folates cannot be transported across the cell membrane; therefore, they are the essential forms for folate retention in cells (Appling, 1991).

From mammalian studies, most cellular folate is protein-bound. In the liver, 60% of cytosolic folate and 20% of mitochondrial folate are protein-bound forms (Zamierowski and Wagner, 1977). More than half of bound polyglutamylated folates are contained in the liver (Clifford *et al.*, 1990). The major folate binding proteins in cytosol are glycine N-methyltransferase (GNMT) and 10-formyltetrahydrofolate dehydrogenase enzymes. Both enzymes provide high binding affinity to folate polyglutamates with dissociation constants (Kd) in the 100 nM range (Yeo *et al.*, 1999; Min *et al.*, 1988). Activities of both enzymes tend to be regulated by folate derivatives. Hexaglutamate folate is a potent inhibitor of 10-formyltetrahydrofolate dehydrogenase enzyme purified from pig liver (Min *et al.*, 1988), whilst, pentaglutamate folate is an inhibitor for GNMT (Wagner *et al.*, 1985). In mitochondria, sarcosine dehydrogenase and dimethylglycine dehydrogenase enzymes are the major tightly-bound folate binding proteins which bind strongly to pentaglutamate folate (Wittwer and Wagner, 1981). Moreover, enzymes in one-carbon metabolism also provide high binding affinity to folate polyglutamates with Kd in the 100 nM range, for example, serine hydroxymethyltransferase (SHMT) and folylpolyglutamate synthetase (FPGS) (Strong *et al.* 1990; Cichowicz and Shane, 1987). Although bound-cellular folate polyglutamates are studied mostly in mammalian system, it could be applied to the plant system. With folate-dependent enzymes in all plant cell compartments, folate binding capacity of these enzymes can affect intracellular folate accumulation.

Studies of folate polyglutamylation in yeast and mammalian systems have shown that protein binding protects folate molecules from oxidative breakdown (Suh *et al.* 2001; Jones and Nixon 2002). Metabolic enzymes which are favoured by polyglutamylated folate appear to be an efficient folate binding proteins resulting in folate stability in cells. Furthermore, protein-binding aids the protection of deglutamylation of polyglutamyl folates from γ-glutamyl hydrolase (GGH) (Wang *et al.*, 1993). Hence, these mechanisms could retain cellular polyglutamylated folate forms for the demand of cells.

The major pools of folate molecules are in the cytosol and mitochondria of mammalian and plant cells. Only in plant cells, folate coenzymes can be found in the chloroplast (Cossins and Shuh, 1972). Hence, a duplication of folate-dependent enzyme activities in these cell compartments can fulfil different metabolic roles within specific compartment (for summary, see Figure 1.2 and 1.3). Addition or removal of one carbon (C1) unit is required by a number of metabolic pathways; for example, nucleic acid biosynthesis, amino acid biosynthesis, methylation reactions on lipids, proteins and DNA as well as mitochondria and chloroplast molecule biosynthesis.

1.1.3.1 One-carbon metabolism in the cytosol

Three major one-carbon transfer reactions requiring folate coenzymes in the cytosol are purine, thymidylate and methionine biosynthesis. Purine and thymidylate molecules are involved in DNA synthesis, whilst, methionine is an essential molecule in the methylation cycle (Figure 1.2 and 1.3).

DNA synthesis in both mammalian and plant cells require polyglutamylated folates for the production of purine and thymidylate bases. To produce purines, 10-formyl tetrahydrofolate provides a one-carbon unit to a transformylase enzyme to insert a carbon group into the purine structure. For thymidylate synthesis, 5,10-methylene tetrahydrofolate provides a one-carbon unit to the thymidylate synthase (TS) enzyme for converting dUMP into dTMP. This reaction generates dihydrofolate which can be converted into tetrahydrofolate by dihydrofolate reductase (Scott *et al.*, 2000).

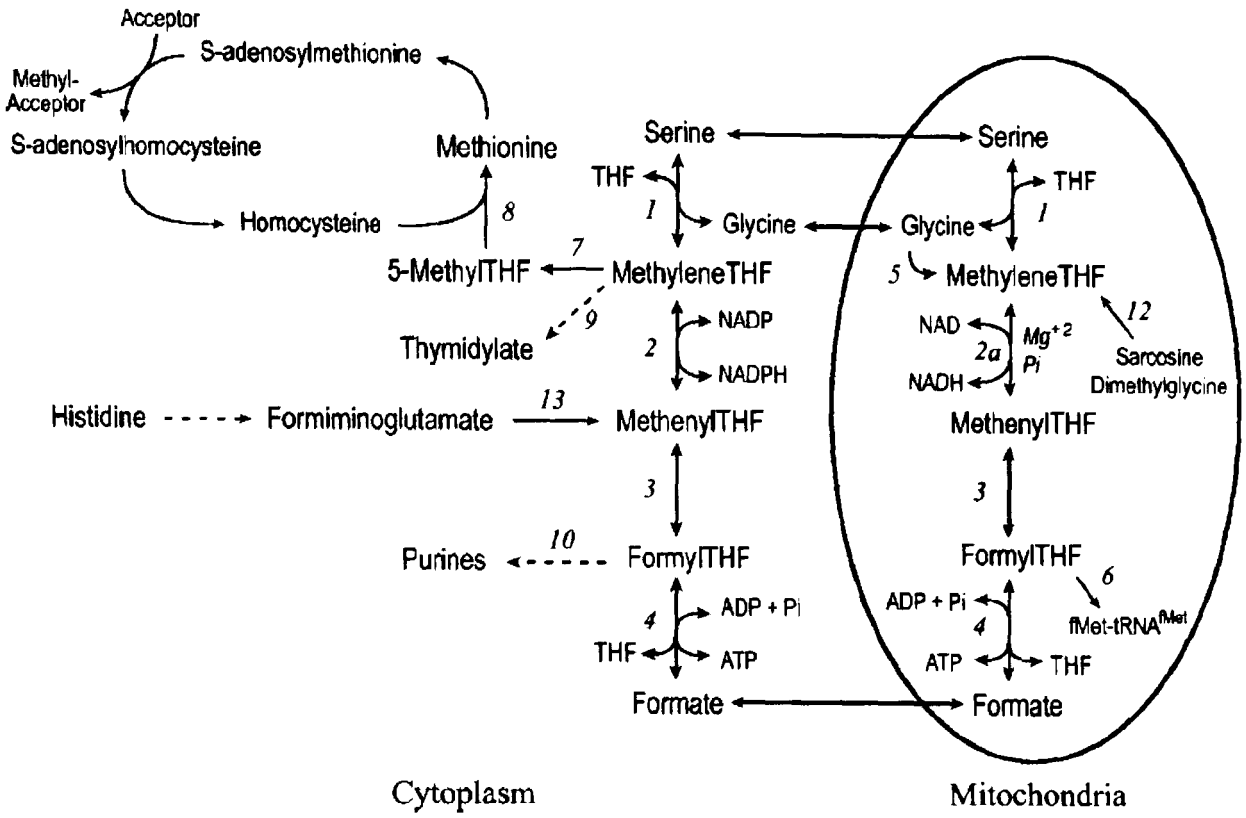


Figure 1.2 One-carbon metabolisms in mammalian cells

Numbers represent enzymes involved in one-carbon transfer reactions. 1, serine hydroxymethyltransferase (SHMT); 2,3,4 trifunctional NADP-dependent methylene-tetrahydrofolate dehydrogenase (MTHFD)/ methenyl-tetrahydrofolate cyclohydrolase (MTHFC)/ 10-formyl-tetrahydrofolate synthetase (FTHFS); 2a, 3, bifunctional NAD-dependent MTHFD/MTHFC; 4, FTHFS; 5, glycine decarboxylase complex (GDC); 6, methionyl-tRNA formyl transferase; 7, NADH-dependent methylene-tetrahydrofolate reductase (MTHFR); 8, methionine synthase (MS); 9, thymidylate synthase (TS); 10, glycinamide ribonucleotide transformylase (GART) and aminoimidazolecarboxamide ribonucleotide transformylase (AICART); 12, sarcosine dehydrogenase and dimethylglycine dehydrogenase; 13, formiminotransferase-cyclodeaminase; THF, tetrahydrofolate. Picture was taken from Christensen and MacKenzie, 2006.

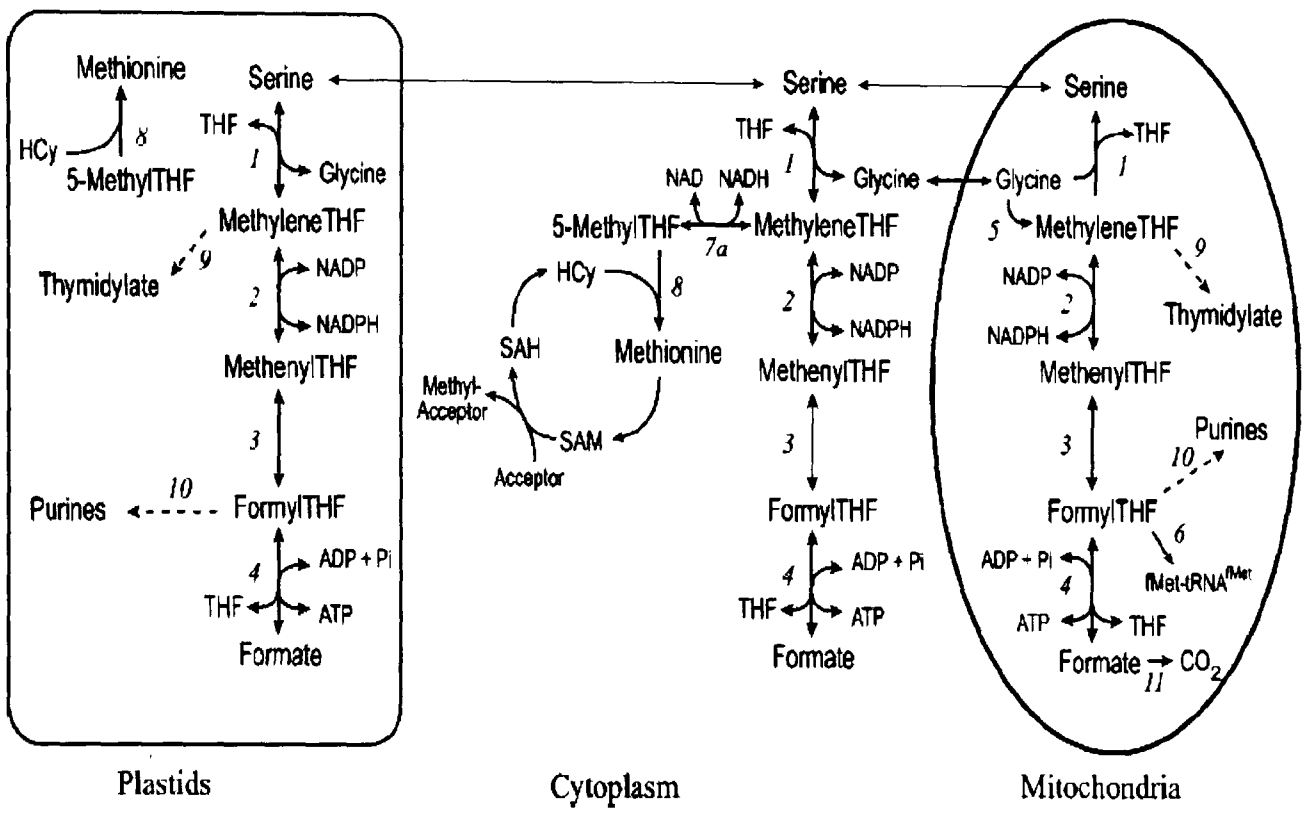


Figure 1.3 One-carbon metabolisms in plant cells

Numbers represent enzymes involved in one-carbon transfer reactions. 1, serine hydroxymethyltransferase (SHMT); 2,3, bifunctional NADP-dependent methylene-tetrahydrofolate dehydrogenase (MTHFD)/ methenyl-tetrahydrofolate cyclohydrolase (MTHFC); 4, 10-formyl-tetrahydrofolate synthetase (FTHFS); 5, glycine decarboxylase complex (GDC); 6, methionyl-tRNA formyl transferase; 7a, NADH-dependent methylene-tetrahydrofolate reductase (MTHFR); 8, methionine synthase (MS); 9, thymidylate synthase (TS); 10, glycinamide ribonucleotide transformylase (GART) and aminoimidazolecarboxamide ribonucleotide transformylase (AICART); 11, formate dehydrogenase; THF, tetrahydrofolate. Picture was taken from Christensen and MacKenzie, 2006.

As folates are involved directly in DNA synthesis, anti-folate drugs are the target for treatment of several diseases such as cancer, malaria and rheumatoid arthritis (Appling, 1991). Moreover, 5,10-methylene tetrahydrofolate can be converted to 5-methyl tetrahydrofolate monoglutamate by the action of 5,10-methylene tetrahydrofolate reductase (MTHFR). This monoglutamate molecule participates in the methylation cycle.

Methylation processes are required in all organisms. The efficient one-carbon donor in methylation processes of lipids, proteins and DNA is S-adenosylmethionine (SAM). It is produced by adenylation of a methionine molecule. To produce methionine, homocysteine is remethylated by the methionine synthase enzyme with vitamin B12 and 5-methyltetrahydrofolate monoglutamate coenzyme. This reaction is also very important in cell detoxification because homocysteine is toxic at low level; hence methionine synthase protects cells from homocysteine toxicity (Hanson and Roje, 2001).

In mammalian cells, it is clear that SAM acts as a methyl donor to synthesize a wide range of methylated products. For example, methylated lipids represent a major part of the cell membrane, whilst methylated proteins can be the activated or inactivated forms of enzymes and hormones (Scott *et al.*, 2000). The ratio of SAM/SAH is regulated by the glycine N-methyltransferase (GNMT) enzyme, which is one of the major folate binding proteins (FBP) in rat liver (Kerr, 1972). However, it is not clear what the importance of the SAM/SAH ratio is in plant cells. However, some reports indicated that a high value of SAM/SAH ratio can also be found in plants which suggests similar effects as in mammalian cells (Edward, 1995; She *et al.*, 1994).

1.1.3.2 One-carbon metabolism in the mitochondria

The crucial role of one-carbon metabolism in the mitochondria is to generate the amino acids glycine from serine (Figure 1.2 and 1.3). This reaction is catalysed by serine hydroxymethyltransferase (SHMT) enzyme with tetrahydrofolate polyglutamate coenzyme. Formate, the second product produced from this reaction, can be transferred to the cytosol as a major source

of one-carbon units for cytosolic purine synthesis in mammalian cells (Barlowe and Appling, 1988). However, the transportation of formate would not occur in plants (Christensen and Mackenzie, 2006).

5,10-methylene tetrahydrofolate polyglutamates, are mainly produced from the conversion of serine, and are needed in purines and thymidylate biosynthesis in cytosol (Cossins and Chen, 1997). Basically, polyglutamylated folates in mammalian cells are not transported into mitochondria but they can be effluxed from mitochondria into cytoplasm without prior hydrolysis (Kim *et al.*, 1993). SHMT enzyme can also be found in cytoplasm as well as in mitochondria. The cytosolic isozyme is abundant in liver, whilst, the mitochondria isozyme is expressed in Chinese hamster ovary cells and yeast (Schirch, 1980; Chasin *et al.*, 1974; Zelikson and Luzzati, 1977).

The polyglutamate forms of tetrahydrofolate are also required by the glycine decarboxylase complex (GDC). The GDC reaction provides 5,10-methylene tetrahydrofolate polyglutamates from the oxidation of glycine. The eukaryotic GDC complex consists of a number of enzymes which are a dimeric, pyridoxal phosphate binding protein (P-protein), a lipoic acid-containing protein (H-protein), a tetrahydrofolate-binding protein (T-protein) and a lipoamise dehydrogenase protein (L-protein) (Cossins and Chen, 1997).

In plants, polyglutamylated folates are involved in the photorespiratory pathway (Figure 1.4). The photorespiration is initiated in chloroplast by ribulose 1,5-biphosphate carboxylase/oxygenase to produce 3-phosphoglycerate and 2-phosphoglycolate (Lorimer, 1981). This pathway involves the chloroplast, peroxisome and mitochondria. The key step for utilising folates in the photorespiration pathway is in mitochondria with the interconversion between glycine and serine molecules.

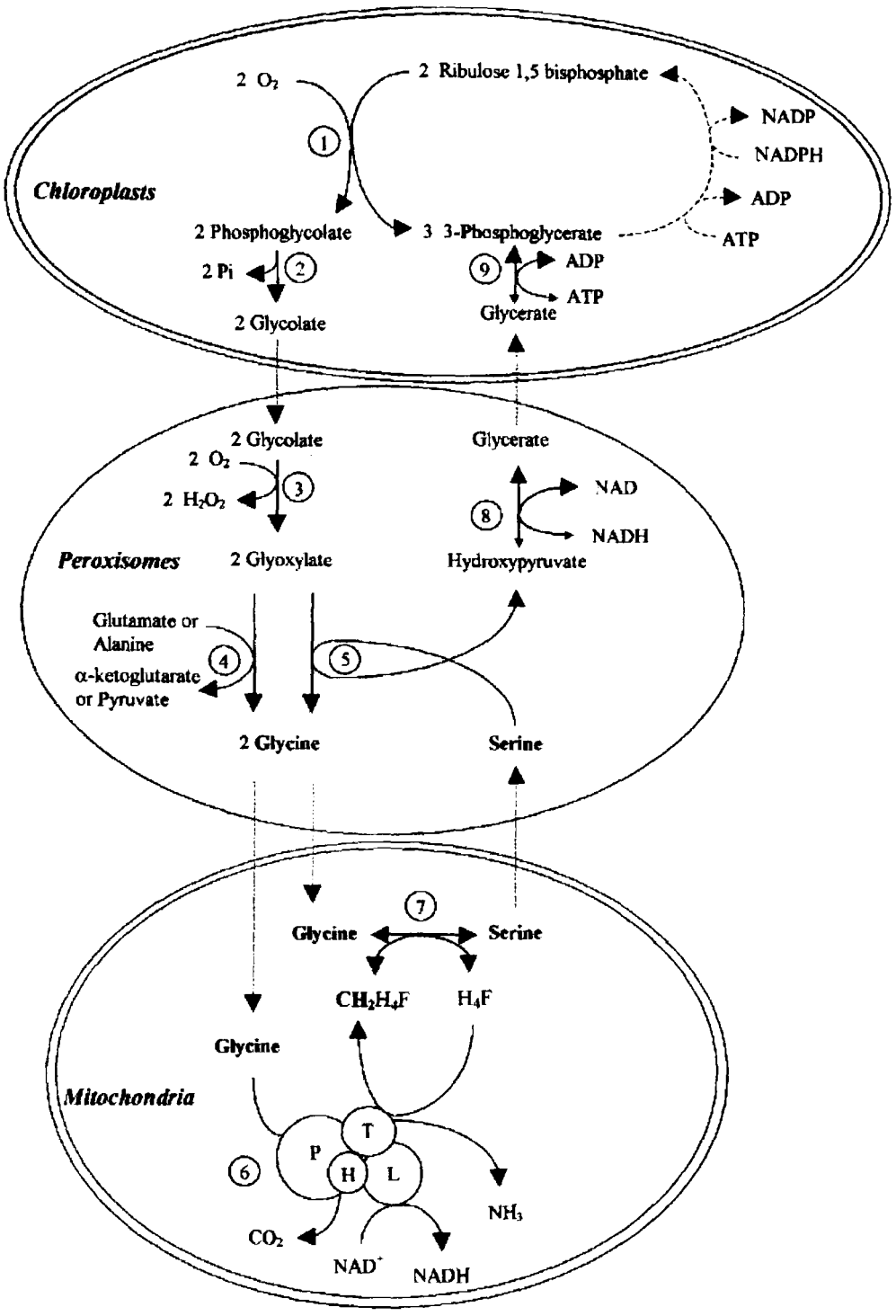


Figure 1.4 Folate-mediated photorespiratory cycle in plant cells

Folate-dependent enzymes are represented in numbers. 1, ribulose 1,5-bisphosphate carboxylase/oxigenase; 2, phosphoglycolate phosphatase; 3, glycolate oxidase; 4, glyoxylate:glutamate (alanine) aminotransferase; 5, glyoxylate:serine aminotransferase; 6, glycine decarboxylase; 7, serine hydroxymethyltransferase (SHMT); 8, NADH-dependent hydroxypyruvate reductase; 9, glycerate kinase. H_4F, tetrahydrofolate, CH_2H_4F, 5,10-methylenetetrahydrofolate. Picture is taken from Scott *et al.*, 2000.

Glycine provided from the photorespiration cycle in peroxisome is used as a substrate by SHMT and GDC enzymes to produce serine (Oliver, 1994; Hanson and Roje, 2001). Hence, GDC and SHMT in plant mitochondria are present in larger amounts than in mammalian cells and are very active in light-grown plants (Oliver and Raman, 1995; Turner *et al.*, 1993). Also a large pool of 5,10-methylene tetrahydrofolate is released in the pea leaf mitochondria indicating photorespiration is an important source of carbon units in plants (Bourguignon *et al.*, 1988). This folate pool may account for 65-80% of total folate in plant cells (Rebeille *et al.*, 1994).

Moreover, formylation of initiator transfer RNA (tRNA) in mitochondria catalysed by methionyl-tRNA formyltransferase enzyme needs 10-formyltetrahydrofolate coenzyme. fMet-tRNA is required for protein translation in both mammalian and plant organelles (Staben and Rabinowitz, 1984).

1.1.3.3 One-carbon metabolism in the chloroplast

Some folate-dependent enzymes seem to duplicate in this compartment (Figure 1.3). Within the plastid, SHMT enzyme is active to generate methylene tetrahydrofolate coenzyme for use in thymidylate synthesis. In contrast to mitochondria, plastids lack GDC enzyme to use glycine as a carbon donor and it is not transported out of the cell (Luo *et al.*, 1997). Plastid appears to have ability to synthesise DNA as purine molecules can be produced using formyl-folate coenzymes (Atkins *et al.*, 1997). Methionine synthase enzyme is also found in plastid for methionine synthesis; but plastids lack MTHFR to produce the folate coenzyme (Hanson *et al.*, 2000). Hence, it seems that plastid is autonomous for methionine synthesis but cannot generate 5-methyltetrahydrofolate coenzyme to use in methionine synthesis. However, methionine appears to be produced in plastid using S-adenosylmethionine transported from the cytosol but it is still unclear how plastids make the 5-methyltetrahydrofolate coenzyme (Ravanel *et al.*, 2004). In plants, the

existence of chloroplasts make folate metabolism more complicate than in other organisms.

1.1.4 The consequences of folate deficiency

The deficiency of folate can cause several problems in the retardation of cell growth and affect folate-mediated physiological functions.

In human, folate deficiency can cause severe birth problems such as neural tube defects (NTDs) in neonatal and increased risk of cardiovascular disease due to the reduction of the DNA synthesis capacity in cells and elevated serum homocysteine (Gerhard and Duell, 1999; Geisel, 2003). Low folate levels in pregnant women are associated with devastating birth defects such as spina bifida, a group of developmental defect of incomplete closure of the embryonic neural tube resulting in an incompletely formed spinal cord or open and unfused spinal cord. This defect is considered the second main cause for infant mortality worldwide (Copp, 1993). The etiology of NTDs worldwide is 1 in 500 in average, whilst, the wide-range of the birth defect prevalence is reported across the western countries (van der Put *et al.*, 2001). In 1991, the etiology number ranged from 1 in 2500 in Finland, 1 in 700 in the Netherlands, 1 in 300 in Mexico, and 1 in 80 in South-Wales (Dolk *et al.*, 1991). However, the prevalence of birth defects has declined over 80% in the last 20 years in the United States, the United Kingdom and Australia when prenatal prevention has been applied (Stone, 1987; Leech, 1991; Chan *et al.*, 1993). Several case-control studies, cohort studies, and randomized and nonrandomized intervention studies demonstrated the evidence of the preventive effect of folate supplementation in pregnant women (Frey and Hauser, 2003). The birth defect could be reduced by half in the patient group receiving 0.8 mg folate supplementation daily for at least a month during the preconception period (Czeizel and Dudás, 1992). With the mouse folate-impaired mutants, the frequency of NTDs can be significantly lower by supplementation of folic acid (Fleming and Copp, 1998).

Genetics is one factor causing inherited NTDs. Mutation in the methylenetetrahydrofolate reductase (MTHFR) gene can cause insufficient re-methylation of homocysteine to methionine and elevate homocysteine levels (van der Put *et al.*, 2001; Eikelboom *et al.*, 1999). Folate intake can normalize the homocysteine concentration and can prevent the etiology of cardiovascular diseases by 25% (Finglas *et al.*, 2006). High levels of homocysteine in blood may also increase the risk of dementia and Alzheimer's disease (Seshadri, 2002).

Megaloblastic anemia results from DNA mismatch of uracil instead of thymidylate in bone marrow cells which causes clumping and fragmentation of nuclear chromatin. The impairment of thymidylate synthesis is the major cause (Blount, 1997). Large, immature red blood cells are released from the bone marrow into blood circulation, causing anemia. Cancer can be caused by defective methylation due to a lack of S-adenosylmethionine (SAM) and/or 5,10-methylene-tetrahydrofolate, co-substrate in cytosine methylation (Storozhenko *et al.*, 2005). DNA methylation is an important tool of cells to silence genes. This defect leads to hypomethylation of CpG sequences in promoter region of several oncogenes causing increased oncogene expression (Storozhenko *et al.*, 2005; Rébeillé *et al.*, 2006). Enhancing folate intake can reduce risk of colorectal (Giovannucci, 2002), pancreatic and breast cancer (Ericson *et al.*, 2007). However, the high folic acid dose intake (1 mg/day for 5 years) could promote colorectal carcinogenesis (Cole *et al.*, 2007). Hence, folate is required to prevent these diseases but not at the high dose that may promote any other risks.

1.1.5 Folate bioavailability

Obtaining a sufficient amount of daily folate not only depends on consuming folate enriched food but also on an ability to absorb folate in the human intestinal tract. Food stuffs contain a high proportion of folate in a polyglutamylated form. Many aspects from the absorption process should be considered for the bioavailability of folate. For example, the amount of folate

released from the food matrix in the intestinal tract, the ability of polyglutamated-conjugases in cleaving polyglutamylated folates and the amount of monoglutamate folate absorbed in the intestinal mucosa.

In pharmacokinetics, folate bioavailability testing is important to estimate *in vivo* metabolism of folate derivatives. Comparisons between folic acid, a stable folate form in fortification process, and natural folate compounds, have been shown that the latter is less bioavailable than folic acid (Brown *et al.*, 1973; Scott *et al.*, 2000). Natural folates have as low as 50% bioavailability compared to folic acid (Cuskelly *et al.*, 1996; Gregory, 1997; Bailey, 1998). From spinach, 52% bioavailable of folates was reported (Perry and Chanarin, 1970; Prinz-Langenhe and Pietrzik, 1998; Scott *et al.*, 2000). However, natural folate tends to have higher retention in plasma (85-90%) than folic acid which might be because folic acid enters the urinary excretion rapidly (Perry and Chanarin, 1970; Tamura and Strokstad, 1973; Scott *et al.*, 2000). In contrast, 76% absorption and 24% faeces excretion were reported for folic acid while 62% absorption and 38% faeces excretion were analysed for polyglutamylated folates (Godwin and Rosenberg, 1975; Scott *et al.*, 2000).

Folic acid is more stable and has higher bioavailability than natural food folate. All fortification programmes have used folic acid to enhance folate status as the supplemented tablets and the fortified white flour, bread and cereal products. However, due to the progress in genomics and metabolomics knowledge in plant foods, biofortification represents long-term solutions.

1.2 FOLATE BIOFORTIFICATION

In theory, plant foods can provide almost all essential micronutrients to human, however, in practice, most plant foods especially staple crops do not contain sufficient amounts of vitamins and minerals to reach the minimum daily requirement. Consequently, fortifying food with micronutrients such as folic acid, the fully oxidized monoglutamyl form of folate, has been applied in

micronutrient enhancement programmes for over 15 years but it has been only partially successful and severe nutritional problems still continue worldwide even in developed countries due to ineffective supplementary programmes and poor eating habits.

In the developed world, eating habits of westerners have changed in the last decade. Less frequent daily eating of fruits and vegetables has been observed in both children and adults along with changing life style and less exercise. Moreover, the consumption of fast food, carbohydrate and fat-rich meals is increasing. Even though medication and hospital facilities are advanced, due to those bad habits people still suffer with having bad nutritional status and the prevalence of folate-related diseases is increasing (McNulty and Scott, 2008).

In Asia where the economic gap is high between people, most of the population cannot afford the variation in food needed to balance their nutrition. Rice is a staple food crop providing more than 70% of the calorie intake per day for people in South East Asia and African counties (Fitzgerald *et al.*, 2009). Polished rice grains which has very low folates; less than 10 microgram per 100 gram, is the common staple food of the middle and lower class. Hence, monotonous food consumption is a factor resulting in malnutrition across much of the population. Obviously, folate supplementation and fortified food need specialized techniques and infrastructure which many poor countries cannot afford. Also the products from these processes are aimed for the good benefit of the consumers, so, the price of those products is not as cheap as the natural food. Hence, these fortification methods are not practical in most developing countries. A survey of the prevalence of folate deficiency among Thai people has shown that folate deficiency is quite common in poor income people, in manual male workers and over 60 year old males; with the prevalence of approximately 38.8% (Assantachai and Lekhakula, 2005). In Venezuela, the prevalence of folate deficiency reached 81% in adolescents and 36% in pregnant women (Garcia-Casal *et al.*, 2005).

Biofortification offers an alternative solution to supplementation in developing countries when combined with breeding programmes (Jeong and Guerinot, 2008). The information obtained from the rice genome provides the opportunity to manipulate vitamin biosynthetic pathways. The ability to manipulate plant nutritional content leading to the creation of biofortified foods would benefit the farmer, consumer and overall health of developing nations.

Several examples of folate biofortification have been reported in the last few years using metabolic engineering approaches. Total folate level was increased firstly in *Arabidopsis* by engineering the pteridine branch of folate biosynthesis. Overexpression of *E.coli GTP cyclohydrolase I* gene, which encodes for the first enzyme in pteridine synthesis, gave a 1250 fold increase of unconjugated pterins but increased total folate by only 2-4 fold (Hossain *et al.*, 2004). A similar strategy was employed in tomatoes where a mammalian *GTP cyclohydrolase I* gene was overexpressed using a fruit specific promoter. The pteridine content in tomato fruits was raised by 3-140 fold. However, total folate was raised by only 2 fold compared to wild type tomatoes (Diaz de la Garza *et al.*, 2004).

The *p*ABA branch of folate biosynthesis is another target to increase folate level. Indeed, supplying *p*ABA to the pteridine engineered fruit could increase folate up to 10 fold (Diaz de la Garza *et al.*, 2004). In addition, crossing tomatoes overexpressing *GTPCHI* and *Arabidopsis ADC synthase* increased folate content by 25 fold in ripened fruits (Diaz de la Garza *et al.*, 2007). Recently, Storozhenko *et al.* (2007) reported folate biofortification in rice seeds by overexpressing *Arabidopsis GTPCHI* and *ADC synthase* genes. It could be enhanced by a maximum of 100 folds above wild type; 89% of the folate produced was 5-methyltetrahydrofolate in this transgenic rice.

Although strategies to enhance total folate level have been successfully developed in tomatoes and rice, additional approaches targeting other traits such as polyglutamylated-mediated folate stability are further targets to enhance folate status in rice species.

1.3 AIMS AND OBJECTIVES OF THE THESIS

Rice is the most important staple food for Asian and African people but contains very low folate concentrations. Due to genetic advances, rice full genome sequence was completed and publicly available (Ware *et al.*, 2002) to identify the biosynthesis and regulatory genes for folate metabolism. Rice would be a model crop for genetic study because of its small genome and great synteny with other cereal genomes has also been shown (Goff *et al.*, 2002). Metabolomics is an important study to improve the quality and nutritional value of rice as the essential metabolic pathways for producing the nutrient based molecules have been identified. Hence, enhancing folate status using genetics and metabolomics knowledge would be focused on rice species. The aim of this thesis is to investigate whether genetic approaches for biofortification can enhance and retain folate levels in rice grains. Some of the following questions and hypotheses have been put forward and will be covered in this thesis.

- Folate synthesis in plants might be regulated at the transcriptional level, so, looking at gene expression might explain the natural variation of synthesized folate among rice varieties.
- Studies of folate polyglutamylation in yeast and mammal systems have shown that polyglutamylation protects folate from oxidative breakdown and enzymatic digestion by protein binding (Suh *et al.*, 2001, Jones and Nixon 2002) resulting in folate stability. Hence, increasing polyglutamylated folate forms might facilitate stability and retention of cellular folate concentration in rice.
- As in iron-rich rice, increasing of ferritin, iron bound molecule, in rice endosperm can enhance by 2-fold of iron level (Lucca *et al*, 2002). Hence, introducing a folate binding protein might increase folate levels in rice endosperm in the same idea.

To examine these questions and hypotheses, the following experiments have been executed below:-

- Comprehensive characterisation of folate biosynthesis gene expression among rice varieties (chapter 3).
- Functional characterisation of the folylpolyglutamate synthase (FPGS) enzyme in the polyglutamylation process in rice grain (chapter 4).
- Manipulate expression of folate related genes which are responsible for synthesizing polyglutamylated folates aiming to retain folate levels in the rice endosperm (chapter 5).
- Develop alternative folate biofortification strategies employing mammalian folate binding proteins (chapter 5).

CHAPTER 2 MATERIALS AND METHODS

2.1 PLANT MATERIALS

2.1.1 Seed materials

Dry grains of rice variety *japonica* cultivar Nipponbare, variety *indica* cultivar IR72 and Moroberekan were obtained from the International Rice Research Institute (IRRI). Dry grains of Shan Yon 63 and Mungur varieties were provided by Dominique Van Der Straeten from Ghent University. T-DNA knock out line, FST number A16772 for FPGS LOC_Os03g02030 gene, FST number D02773 for FPGS LOC_Os10g35940, Dongjin and Hwayoung wild type were obtained from French Rice Functional Genomics Centre, Montpelier (originally from Postech, Korea).

2.1.2 Leave samples for qPCR analysis

Rice seeds were soaked with distilled water on a petri dish and kept in growth chamber at 26-28°C until germination. Germinated seeds were transfer into the compost soil pot containing 1:1 ratio of Levington M3 and John Innes no.3. Pots were then placed in the same growth chamber with a 12 hours light cycle at 28-30°C during the day and 21°C during the night. The third green leaves were collected in 1.5 ml eppendorf tube from at least 5 plants per pool. At least 3 pools of samples were collected to perform RT-qPCR assay. All samples were immediately frozen in liquid nitrogen and stored at -80°C until RNA isolation.

2.1.3 Grain samples for qPCR analysis

The outer layer of approximately 40 dry seeds was removed by hand. All brown seeds were homogenised using Retsch® MM301 ball mill equipment (Retsh, UK) at a frequency of 30 $1s^{-1}$ for 1 minute. Homogenised samples were stored at -80°C until RNA isolation.

2.2 BIOINFORMATICS ANALYSIS

The rice genome databases were explored for folate-related genes based on the sequences of Nipponbare rice. Gramene database (http://www.gramene.org/) and the TIGR rice genome annotation database (http://www.tigr.org/tdb/e2k1/osa1/) were selected as the tools. Amino acid and genomic sequences of rice were blasted across the *Arabidopsis* AtEnsembl genome database (http://atensembl.arabidopsis.info/) and the *Arabidopsis* information resource (TAIR) (http://www.arabidopsis.org/index.jsp) to check the orthologs of folate-related genes. Nucleotide alignment and Boxshade of genomic sequences were analysed using http://searchlauncher.bcm.tmc.edu/multialign/multi-align.html. For the information of folate-related enzyme, Brenda enzyme database was used (http://www.brenda.uni-koeln.de/). Other tools for DNA, RNA and protein analysis were linked via Expasy website (http://www.expasy.org).

2.3 MOLECULAR BIOLOGY AND CLONING TECHNIQUES

All reagents and chemicals used for molecular work were molecular grade. *E.coli* DH5α and Origami B (DE3) strains were purchased from Novagen (UK). Agrobacterium EHA105 strain was a gift from Helen Parker, Nottingham. LB media was used as a bacteriological media for both *E.coli* and Agrobacterium. Antibiotics used in plates and broth were at the following concentration: ampicillin 50 μg/ml; kanamycin 50 μg/ml and rifampicin 25 μg/ml.

2.3.1 Nucleic acid isolation

2.3.1.1 RNA isolation

Total RNA was extracted from leaf and seed rice samples by using TRIzol reagent (Invitrogen, UK) according to the manufacturer's protocol. In brief,

plant samples were homogenised in liquid nitrogen. 1 ml of TRIzol reagent was added to 100 mg of homogenised samples and was mixed thoroughly. After centrifugation at 5790 x g for 10 minutes at 4°C to remove cell debris, 200 µl of chloroform was added and mixed by vigorous shaking for 15 seconds. After centrifugation at the same speed, the aqueous phase was kept. To precipitate RNA, 500 µl of isopropanol was added. For glycogen-rich samples such as grains, 250 µl of isopropanol and 250 µl of 0.8 M sodium citrate containing 1.2 M sodium chloride solution were added instead. After 10 minute incubation at room temperature, the mixture was centrifuged at the same speed as mentioned above and RNA pellet was kept. Then, the pellet was washed by 75% ethanol, resuspended in 180 µl of distilled water and treated with 2 units of DNase enzyme (Sigma) at 37°C for 30 minutes. Then, RNA was cleaned up using phenol and chloroform method. After DNase incubation, equal volume of phenol was added and mixed well by vortex. Centrifugation was carried out at 6800 x g for 5 minutes at 4 °C. Supernatant was collected and equal volume of chloroform was added. The mixture was mixed by vortex and centrifuged at 6800 x g for 5 minutes at 4°C. Supernatant was collected and 0.1 volume of 3 M sodium acetate and 1 volume of isopropanol were then added. The mixture was left at -20°C for at least 5 hours and then centrifuged at 6800 x g for 30 minutes at 4 °C. Supernatant was discarded and 100 µl of 70% ethanol was added to wash a pellet. After centrifugation at 6800 x g for 5 minutes at 4 °C, all supernatant was discarded and allow pellet to air dry. Pellet was then resuspended in 20 µl of RNase/DNase-free water and stored at -80°C until use.

2.3.1.2 DNA isolation

For cloning, genomic DNA was extracted from Nipponbare leaves by using GenElute plant genomic DNA miniprep kit (SIGMA, UK). According to the manufacturer's protocol, in brief, plant samples were homogenised in liquid nitrogen. Lysis solution was added to lyse plant cells. Precipitation solution was then added and all supernatant was transferred to prepared blue filtration

column. Column was then washed twice and DNA was eluted from column with 50 µl of distilled water. Eluted DNA was stored at -20°C until used.

For genotyping of T-DNA knock out lines, crude DNA was extracted from leaves using 0.2 M Tris-EDTA, 0.1 M NaCl and 0.5% SDS extraction buffer. 600 µl of extraction buffer was applied to 100 mg of homogenised leaf samples and then centrifuged at 6800 x g for 5 minutes at room temperature. Only supernatant was transferred to the new tube and 500 µl of isopropanol was added. The mixture was incubated on ice for 5 minutes and centrifuged again. The supernatant was decanted and only the pellet was retained. The pellet was then washed in 75% ethanol and left to air dry. To suspend the pellet, 30 µl of distilled water was applied and crude DNA solution was used straight in PCR reaction.

2.3.2 cDNA synthesis

First-stranded cDNAs were synthesised from 500 ng of treated total RNA with Superscipt™ II reverse transcriptase (Invitrogen, UK) following the manufacturer's protocol. Briefly, 100 pmole of oligo(dT) primer and 10 mM dNTPs were added to 500 ng of total RNA in 13 µl total reaction. After heating at 65°C for 5 minutes, 4 µl of 5x first-strand buffer, 2 µl of 0.1 M DTT and 1 µl of reverse transcriptase enzyme were added and incubated at 42°C for 2 hours. This reaction was then inactivated by heating at 70°C for 15 minutes.

2.3.3 Polymerase Chain Reaction (PCR)

Target genes were amplified from an amount of 25 ng of rice cDNA using 10 pmol of gene specific primers in 25 µl of the reaction mixture containing 10x PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, and 1 unit of Taq DNA polymerase (red hot Taq polymerase, ABgene, UK). The PCR conditions used to amplify the genes of interest were as follow; an initial denaturation step of 2 minutes at 94°C, 35 cycles of 30s for denaturation at 94°C, 30s for annealing at (Ta)°C (depends on primer pairs, see the appendix 1) and 1 minute for extension step at 72°C. The final extension was carried out for 1 cycle at 72°C, 10 minute. The PCR products were analysed on a 1% agarose gel containing

0.4 µg/ml ethidium bromide and were purified using Gen Elute Gel Extraction kit (SIGMA, UK). The purified PCR products of each gene were sequenced.

2.3.4 Quantitative real-time PCR (qPCR)

RNA extraction and DNase treatment was carried out as described in 2.3.1.1. The concentration of total RNA from all samples was measured using nanodrop 1000 spectrophotometer (Thermo Scientific, UK). 500 ng of total RNA was used to performed cDNA synthesis as described in 2.3.2. Approximately 25 ng of cDNA (1 µl from cDNA synthesis reaction) was used in single qPCR reaction. Prior to performing qPCR, tradition PCR was done in order to examine DNA contamination by including a no-RT control. The PCR conditions for examining DNA contamination used to amplify the actin internal control was described in 2.3.3 which annealing temperature for actin primer was at 55°C. Samples which showed negative result (no amplified band) in the no-RT control were used in qPCR experiment. The absolute qPCR assays were performed on a Corbett Rotorgene 6000 QPCR system (Corbett life science, UK) using SYBR Green I detection kit (Stratagene, UK). The PCR conditions for short targets (<300 bp) were as follows; 1 cycle of 10 minutes initial denaturation step at 95°C, 40 cycles of 30s denaturation step at 95°C, 1 minute annealing step at 55°C or 58°C (depends on probes, see appendix 1) and 1 minute extension step at 72°C. The melting curve was run from 55°C or 58°C to 90°C for checking a single band of a specific product. A standard curve was prepared by serial dilution of a known number of actin PCR product molecules using known DNA concentration of 60 ng/µl, 15 ng/µl, 1.5 ng/µl, 150 pg/µl and 15 pg/µl. The gene expression levels were interpreted by comparing the threshold cycles of samples to the linear standard curve. The acceptable standard curve should reach 90-110% reaction efficiency and 0.975 or greater of an amplification efficiency value based on the instruction manual of the machine. The absolute expression value of each folate-related gene was normalised to the expression level of rice action gene. All samples were performed in at least 3 replicates. The expression values of each gene in each rice cultivar were then statically compared using SPSS version 14.0. A

probability value of less than 5% ($P \leq 0.05$) was considered statistically significant.

2.3.5 Gene cloning techniques

The fragments of target genes were amplified by PCR using Phusion High-Fidelity DNA Polymerase (NEB, UK). An amount of 25 ng of template cDNAs was added into 50 µl of total reaction mixture containing 5x Phusion HF buffer, 0.2 mM dNTPs, 1 unit of Phusion DNA polymerase enzyme and 15 pmol of specific primers. The amplification was started with an initial denaturation at 98°C for 30s. This was followed by 35 cycles of denaturation at 98°C for 10s, annealing at (Ta)°C of each primer for 30s and extension at 72°C for 1 minute. The final extension was carried out for 1 cycle of 72°C for 10 minutes.

In this thesis, there were 5 gene fragments amplified, cFBP, GNMT, glutelin-1 promoter, FPGS03g and FPGS10g, with the specific primers (see the appendix 1). The templates used to amplify these 5 genes were different. cFBP was amplified from cow udder cDNA, GNMT was amplified from rat liver tissue cDNA, glutelin-1 promoter was amplified from Nipponbare rice DNA, FPGS03g and FPGS10g were amplified from Nipponbare rice cDNA. The final PCR products contained the complete attB sites, essential for gateway cloning, were gel purified and then cloned into pGEM-T Easy vector (Promega, UK).

2.3.5.1 Purification of PCR products from agarose gel

The PCR products were purified from agarose gel using QIAquick Gel Extraction kit (QIAGEN, UK). From the instruction manual, briefly, after gel electrophoresis, the gel area containing the DNA fragment was excised with a sterile razor blade. Then the gel slice was dissolved in a solubilising solution containing guanidine thiocyanate and indicator dye with the ratio of gel: solution at 1:3 (v/v). Thereafter, 1 gel volume of isopropanol was added and all of the mixture was applied onto a QIA column. The DNA was adsorbed onto membrane of the column while the column was washed out with a buffer

containing ethanol. Finally, the DNA was eluted with distilled water and stored at -20°C until use.

2.3.5.2 Ligation into pGEM-T Easy vector

To check the sequence of the amplified gene, the purified cFBP, GNMT, glutelin-1 promoter, FPGS03g and FPGS10g fragments were ligated to the pGEM-T Easy vector with the standard method described by the company. Because Phusion polymerase generates a blunt end PCR product, generating 3'overhang for PCR product is initially required. In brief, PCR reaction was set up without the specific primers in total amount of 10 µl using purified DNA fragments as the templates and Taq DNA polymerase. This reaction was incubated at 72°C for 15 minutes and 2 µl of reaction was then used for ligation. The ligation mixture (10 µl in total volume) comprised 10 ng of the DNA, 30 ng of pGEM-T Easy vector, and 3 Weiss units of T4 DNA ligase. The reaction was carried out at room temperature (25°C) for 1 hour. Then, half volume of the ligation mixture was transformed into the competent cells of *E.coli* DH5α, and cells were grown overnight at 37°C. Ten single white colonies, expected to contain the plasmid with a DNA insert, were selected to perform colony PCR to check the insertion. A colony of each gene which gave the expected band size was cultured overnight in 2 ml of LB medium containing an appropriate antibiotic. The plasmids were purified from bacteria cells using Wizard Plus SV miniprep DNA purification kit (Promega, UK). Size of the inserted DNA was checked by the restriction analysis, and the sequence of the inserted gene was confirmed by DNA sequencing.

2.3.5.3 Restriction analysis of plasmids

Because there are multiple cloning sites within the vectors, the presence and size of the DNA of interest in the vectors can be determined by single or double digestion of the plasmids with the appropriate restriction enzymes. For double digestion, 5 µl of the plasmid (correspond to 250-500 ng) was mixed with 10 unit of restriction enzymes in the total reaction volume of 10 µl. The

reaction mixture was incubated at 37°C for 2 hours prior to the electrophoresis which was performed to check an appropriate band pattern.

2.3.5.4 Gateway cloning for cFBP and GNMT

Gel purified cFBP and GNMT gene fragments were integrated into Gateway entry vector, pDONR221, by a BP reaction. 10 µl reaction was set up including 1 µl of 300 ng of pDONR221, 2 µl of each purified gene fragments, 2 µl of BP clonase enzyme (Invitrogen, UK), 2 µl of 5x BP reaction buffer and 3 µl of 1x TE buffer. This mixture was incubated at 25°C overnight and then 1 µl of proteinase K was added to inactivate the reaction and incubated at 37°C for 10 minutes. 5 µl of each proteinase K-treated reaction was next transformed into 50 µl of the competent *E.coli* DH5α cells. The positive colonies were selected on LB plates containing 50 µg/ml kanamycin. Colony PCR was done to check the insert and the plasmids which gave the expected band size were selected for plasmid purification.

Genes inserted in pDONR221 vectors were transferred into the destination vector, pDEST17 which was used as the *E.coli* expression vector, by LR reaction. 10 µl reaction was set up including 1 µl of 150 ng of pDEST17, 1 µl of 50 ng of gene-inserted pDONR221, 2 µl of LR clonase enzyme (Invitrogen, UK), 2 µl of 5x LR clonase buffer and 4 µl of 1x TE buffer. This mixture was incubated at 25°C overnight and then 1 µl of proteinase K was added and incubated at 37°C for 10 minutes. 5 µl of each proteinase K-treated reaction was next transformed into 50 µl of the competent *E.coli* DH5α cells. LB plates containing 100 µg/ml ampicillin were also used in colony selection. The colony which gave the inserted band from colony PCR was chosen for plasmid miniprep. Purified gene-inserted pDEST17 was transformed into Origami B (DE3) *E.coli* strain for protein expression.

2.3.5.5 Gateway cloning for FPGS Os03g02030 and FPGS Os10g35940

After checking the sequences of glutelin-1 promoter, FPGS Os03g02030 and FPGS Os10g35940 fragments cloned in pGEM-T Easy vectors, glutelin-1

promoter was firstly ligated into pENTR11 vector. The ligation mixture (10 µl in total) comprised 10 ng of gene fragments, 20 ng of restricted pENTR11 vector, and 3 Weiss units of T4 DNA ligase. The reaction was carried out at 16°C overnight. Then, half volume of the ligation mixture was transformed into the competent *E.coli* DH5α cells using heat-shock method, and cultured overnight at 37°C on LB containing 50 µg/ml kanamycin plates. Twenty single colonies were selected randomly to perform colony PCR to check the insertion. A colony of each gene which gave the expected band size was cultured in 2 ml of LB-kanamycin broth. The plasmid was then purified. DNA insertion and orientation were confirmed by restriction analysis and PCR using gene specific primers. After inserting glutelin-1 promoter in pENTR11, two FPGS genes were also inserted into glutelin-inserted pENTR11 with the same strategies. Finally, each pENTR11 had glutelin promoter followed by FPGS gene.

Due to gateway compatible sites (attL) in pENTR11 vector, glutelin promoter and FPGS gene cassettes could be transferred into binary vector, pGWB7, using gateway reaction. LR reaction was set up consisting of 1 µl of 150 ng linearized pGWB7, 0.5 µl of glutelin: FPGS inserted pENTR11, 2 µl of LR clonase enzyme, 2 µl of 5x LR clonase buffer and 4.5 µl of 1x TE buffer. After incubation at 25°C overnight and proteinase K adding, 5 µl of each treated reaction was transformed into 50 µl of the competent *E.coli* DH5α cells. LB plates containing 50 µg/ml kanamycin were used in cloning selection. The colony which gave the inserted band was chosen for plasmid miniprep. Purified gene-inserted pGWB7 was transformed into EHA105 Agrobacterium for rice transformation.

2.3.6 Bacterial transformation

2.3.6.1 Transformation of *E.coli*

Heat shock method was used to transform cells with foreign DNA, The mixture of gene-inserted vector and competent *E.coli* cells in a polypropylene tube was incubated on ice for 10 minutes. Then cells were subjected to heat-shock by placing the tube containing cells to a heat block of 42°C and incubated for

exactly 90 seconds without shaking and immediately placed on ice for 2 minutes. Thereafter, 450 µl of LB broth was added into the reaction tube and then it was shaken at 37°C for 1 hour. Aliquot 200 µl of the culture was plated on LB containing antibiotic plates and incubated at 37°C overnight.

2.3.6.2 Transformation of Agrobacterium

Electroporation method was used for introducing genes into Agrobacterium and EHA105 strain was chosen for rice transformation. The mixture of gene-inserted binary vector and competent EHA105 cells was firstly placed on ice for 1 minute. Then, this mixture was transferred to ice-cold 0.2 cm electroporation cuvette (Molecular BioProducts, UK) and placed on ice for 10 minutes. The cuvette was then put on the MicroPulser electroporation chamber (BioRad, UK) and pulsed once at 25 µF and 2.5 kV. After that, the cuvette was removed from the chamber, immediately added 1 ml of LB broth and incubated at 28°C for 1 hour without shaking. About 100 µl of cell suspension was placed on LB agar containing 25 µg/ml rifampicin and the vector-specific antibiotic for plasmid selection. Culture plates were incubated at 28°C for 2 days until colonies appeared.

2.4 RICE TRANSFORMATION

Agrobacterium-mediated Nipponbare rice transformation was modified from the protocol of Hiei *et al* (1994) and Maarten Volckaert (University of Gent, personal communication). There were 5 major steps in the transformation procedure.

2.4.1 Callus production

Sterile seeds were put on the Callus Induction Medium (CIM) plates, closed the lids with parafilm, incubated at 28°C in the dark about 4 weeks until the embryogenic callus formed. Then, the callus was moved onto fresh CIM plates for callus multification and continually incubated for 2 weeks to get bigger callus. The composition of CIM includes basal NB medium added 30 g/L sucrose, 2.5 mg/L 2,4-D and 2.5 g/L Phytagel at pH 5.8.

2.4.2 Callus cocultivation

Agrobacteria cells containing gene-inserted binary vector were cultured in LB-kanamycin broth. Overnight culture was spread on the Agrobacterium (AB) plates and incubated at 28°C in the dark for 3 days. After that, Agrobacterium on plates were scraped and dissolved in liquid Cocultivation medium (CCM) was added to make the absorbance at 600 nm equals 1. The composition of CCM includes basal R2 medium added 10 g/L glucose, 2.5 mg/L 2,4-D and 400 μM acetosyringone at pH 5.2. The callus from CIM plates was put into Agrobacterium suspension, shaken for 15 minutes, removed from the suspension and the callus pieces blotted on the sterile filter papers until they dried. Dry callus was then put on solid CCM and incubated at 28°C in the dark for 3 days. The callus on CCM plates was washed in liquid CCM containing vancomycin and cefotaxine to kill Agrobacteria. Washed callus was then put on selection medium and incubated at 28°C in the dark for 4 weeks until the resistant callus started to grow.

2.4.3 Improvement of embryogenicity

Hygromycin resistant callus was moved onto Embryogenicity-improvement medium (EIM) and incubated at 28°C in the dark for 4 weeks. The composition of EIM includes basal LS medium added 30 g/L sucrose, 2.5 mg/L 2,4-D, 100 ml/L coconut water (Gibco BRL), 100 mg/L cefotaxime, 100 mg/L vancomycin, 100 mg/L hygromycin and 7 g/L agarose type I (Sigma) at pH 5.8.

2.4.4 Regeneration of plantlet

Regeneration (REG) medium was used to induce shoot formation. Embryonic callus tissues were placed on REG medium and kept at 28°C in 12/12 dark/light environment. The composition of REG medium includes basal LS medium added 40 g/L sucrose, 0.3 mg/L BAP, 0.5 mg/L IAA, 100 mg/L cefotaxime, 100 mg/L vancomycin and 7 g/L agarose type I (Sigma) at pH 5.8.

2.4.5 Plant development

When the plantlets were about 3 cm, they were moved onto Plantlet-development medium (PDM) with hygromycin for 1 week. PDM includes basal ½ MS- medium added 10 g/L sucrose, 0.05 mg/L 1-NAA and 2.5 g/L phytagel at pH 5.8. Then, the fresh root growth plantlets were moved to PDM without hygromycin and grown at the same condition as regeneration step. When the plants grew big enough, they were transferred to soil and kept in the room with high humidity, 28°C and light condition of 30 Lux. Seeds were harvested after 3-4 months.

2.4.6 Composition of basal medium solutions

Basal NB medium : 100 ml/L N6 Macro I, 100 ml/L N6 Macro II ($CaCl_2.2H_2O$ 1.66 g/L), 10 ml/L B5 vitamins, 10 ml/L MS FeNaEDTA, 1 ml/L B5 Micro, 500 mg/L Proline, 500 mg/L Glutamine, 300 mg/L Casein enzymatic hydrolysate or CEH.

Basal R2 medium : 100 ml/L R2 Macro I, 100 ml/L R2 Macro II ($CaCl_2.2H_2O$ 1.46 g/L), 25 ml/L LS vitamins (10 g/L Myo-inositol and 0.04 g/L Thiamine-HCl), 10 ml/L R2 FeNaEDTA, 1 ml/L R2 Micro.

Basal LS medium : 100 ml/L MS Macro I, 100 ml/L MS Macro II ($CaCl_2.2H_2O$ 4.4 g/L), 10 ml/L LS vitamins, 10 ml/L MS FeNaEDTA, 1 ml/L MS Micro

Basal ½ MS- medium : 50 ml/L MS Macro I, 50 ml/L MS Macro II, 5 ml/L B5 vitamins, 5 ml/L MS FeNaEDTA, 0.5 ml/L MS Micro

N6 Macro I : KNO_3 28.3 g/L, $(NH_4)_2SO_4$ 4.63 g/L, KH_2PO_4 4.00 g/L, $MgSO_4.7H_2O$ 1.85 g/L

B5 vitamins : Myo-inositol 10 g/L, Thiamine-HCl 1 g/L, Nicotinic acid 0.1 g/L, Pyridoxine-HCl 0.1 g/L

B5 Micro : $MnSO_4.H_2O$ 1000 mg/100 ml, Kl 75 mg/100 ml, H_3BO_3 300 mg/100 ml, $ZnSO_4.7H_2O$ 200 mg/100 ml, $CuSO_4$ 2.5 mg/100 ml, $Na_2MoO_4.2H_2O$ 25 mg/100 ml, $CoCl_2.6H_2O$ 2.5 mg/100 ml

R2 Macro I : KNO_3 40 g/L, $(NH_4)_2SO_4$ 3.3 g/L, $NaH_2PO_4.H_2O$ 3.12 g/L, $MgSO_4.7H_2O$ 2.46 g/L

R2 Micro : H_3BO_3 283 mg/100ml, $ZnSO_4.7H_2O$ 220 mg/100ml, $MnSO_4.H_2O$ 160 mg/100ml, $CuSO_4.5H_2O$ 19.5 mg/100ml, $Na_2MoO_4.2H_2O$ 12.5 mg/100ml

MS Macro I : KNO_3 19 g/L, NH_4NO_3 16.5 g/L, $MgSO_4.7H_2O$ 3.7 g/L, KH_2PO_4 1.7 g/L

MS Micro : $MnSO_4.H_2O$ 1690 mg/100ml, $ZnSO_4.7H_2O$ 860 mg/100ml, H_3BO_3 620 mg/100ml, Kl 83 mg/100ml, $Na_2MoO_4.2H_2O$ 25 mg/100ml, $CuSO_4.5H_2O$ 2.5 mg/100ml, $CoCl_2.6H_2O$ 2.5 mg/100ml

2.4.7 GUS staining

T-DNA activation tagging lines for FPGS Os03g02030 and FPGS Os10g35940 received from CIRAD, France, were checked for GUS expression in rice developmental tissues. Rice tissues were incubated in GUS solution at 37°C for 1 hour or overnight. GUS solution contain 1 mM 5-bromo-4-chloro-3-indolyl ß-D-galactopyranoside sodium salt (X-Glc), 0.5% (v/v) Dimethylformamide (DMF), 0.5% (v/v) Triton X-100, 1 mM Ethylenediaminotetraacetic acid (EDTA pH 8), 0.5 mM Potassium ferricyanide ($K_3Fe(CN)_6$), 0.5 mM Potassium ferrocyanide ($K_4Fe(CN)_6$) and 500 mM Phosphate buffer (Na_2PO_4 pH7). This was checked every hour to see the appearance of blue colour on rice tissues. Up to 1 ml of 70% ethanol was added to stop reaction and stained tissues were kept in ethanol for over a month.

2.5 PROTEIN ANALYSIS TECHNIQUES

2.5.1 Recombinant protein expression

The Origami B (DE3) clones containing cFBP and GNMT were induced by IPTG to synthesize the soluble proteins in *E.coli*.

2.5.1.1 Small scale optimization

A single colony of the clones on LB-ampicillin plate was inoculated into 5 ml of LB-ampicillin broth and shaken at 37°C overnight in an orbital shaking incubator (200 rpm). The overnight culture was then subcultured into 30 ml LB-ampicillin broth and cultured until the absorbance at 600 nm reached 0.5 to 1. The culture was split into 6 tubes (5 ml per tube) for temperature and IPTG concentration optimization. The trial conditions were as follows: 20°C/0 mM, 20°C/0.1 mM, 20°C/0.5 mM, 37°C/0 mM, 37°C/0.1 mM and 37°C/0.5 mM. An appropriate amount of 1 M IPTG was added into the culture and 500 µl of the culture supernatant was collected. For the induction at 20°C, the culture pellets were collected at 3 and 24 hours after induction as the bacteria growth rate is slow at the low temperature. For the induction at 37°C, the culture pellets were collected at 1 and 6 hours after induction. All pellets were stored at -20°C until analysis. The recombinant intracellular proteins were analysed by SDS-PAGE and the specific protein band was detected by Coomassie blue staining and western blot. The condition that gives the highest amount of produced proteins was chosen.

2.5.1.2 Large scale protein production

To produce a huge amount of recombinant proteins, the culture volume was increased up to 200 ml using the best optimized condition from the small scale. All culture was collected by centrifugation at 1000 x g for 10 minutes at 4°C. The pellet was stored at -20°C until used.

2.5.2 Sample preparation for SDS-PAGE

Cell pellets collected at each time point were thawed and resuspended in 500 µl of lysis buffer. The cell suspension was froze and thawed for 3 times and then

sonicated for 5 minutes. The lysis cells were centrifuged at 6800 x g for 1 minute at 4°C to pellet insoluble proteins. The supernatant was then transferred to the fresh tube and stored on ice or kept at -20°C until used. SDS-PAGE sample buffer was added into both supernatant and pellet prior to loading gel.

2.5.3 Polyacrylamide gel electrophoresis

The soluble and insoluble forms of synthesized recombinant FBPs were determined by Polyacrylamide gel electrophoresis under denaturing condition with sodium dodecyl sulfate (SDS-PAGE) according to Laemmli and Favre, 1973. The proteins were separated by electrophoresis in 10% acrylamide resolving gel, pH 8.9, and 4% acrylamide stacking gel, pH 6.8. The protein samples were incubated at 100°C for 5 minutes prior to electrophoresis at 120 volts. The protein bands were visualized by staining with Coomassie brilliant blue R-250.

2.5.4 Coomassie blue staining

After electrophoresis, gel was placed in the staining solution containing 0.2% Coomassie brilliant blue R-250, 50% methanol and 10% acetic acid for 1-2 hour. Then, an excess dye was removed by placing the gel in a solution containing 50% methanol and 10% acetic acid. Then gel was stored in a solution containing 7% methanol and 5% acetic acid.

2.5.5 Western blot

To detect the specific protein bands of recombinant FBPs, antibody against cFBP (Abcam, UK) and GNMT (Abgent, USA) was used. After running SDS-PAGE, protein bands were transferred onto a nitrocellulose membrane and detected using SuperSignal West Pico Chemiluminescent substrate (PIERCE, UK) following the instruction from the company. Briefly, after removing the blot from the transfer apparatus, 5% skimmed milk was added as the blocking reagent and incubated at room temperature for 1 hour with shaking. Blocking reagent was then removed and the appropriate primary antibody dilution was added (1:10,000 for cFBP antibody and 1:500 for GNMT antibody) and incubated for 1 hour with shaking. The membrane was washed at least 4-6

times in wash buffer. The blot for GNMT was further incubated with 1:5,000 anti-rabbit IgG-HRP-conjugated antibody for 1 hour at room temperature with shaking. Non-bound antibody was washed by TBS-tween (20 mM Tris-HCl pH 7.4, 500 mM NaCl and 0.1% Tween 20) several times. Working solution of chemiluminescent substrate was prepared by mixing 1 ml of the stable peroxide solution and the Luminol/Enhancer solution. Blot was incubate with this working solution for 5 minutes and removed from the solution. The membrane was placed in the plastic wrap prior to placing in a film cassette for film exposure. Exposure time for cFBP was approximately 1 minutes, whilst the exposure time for GNMT was approximately 5 minutes.

2.5.6 Protein purification by affinity chromatography

To characterise the properties of recombinant FBPs, the purified proteins were needed. An affinity chromatography on Ni-NTA column (Invitrogen, UK) was performed under native conditions according to the manufacturer's instructions. Briefly, the bacterial cells were harvested from 50 ml of culture by centrifugation at 1000 x g for 10 minutes. Harvested cells were resuspended in 8 ml of native binding buffer and 8 mg of lysozyme was added. The suspension was incubated on ice for 30 minutes and then sonicated for 1 minute. The lysate was centrifuged at 1000 x g for 15 minutes to pellet the cellular debris. The supernatant was then transferred to a fresh tube and added to a prepared Ni-NTA column under native conditions. The column was inverted slowly for 60 minutes to allow proteins to bind to the resin. Then, the resin was allowed to settle under gravity and the supernatant was aspirated and kept at 4°C for further analysis. The column was washed with 30 ml of native wash buffer containing 50 mM sodium phosphate buffer pH 8.0 and 20 mM imidazole. Protein was eluted from the column with 8 ml of native elution buffer containing 50 mM sodium phosphate buffer pH 8.0 and 250 mM imidazole. The eluting fractions were collected, and protein concentration was monitored for UV absorbance at 280 nm. FBPs in the fractions were detected by SDS-PAGE followed by Coomassie blue or silver staining. The fractions containing FBPs were pooled, concentrated and stored at -20°C in TBS.

2.5.7 Determination of protein concentration

The concentration of proteins was assayed using the Bradford reagent (Sigma, UK). Bovine serum albumin (BSA) was used as the standard. The reaction mixture comprised of 0.1 ml of protein samples and 3 ml of the Bradford reagent. The complex formation between proteins and Brilliant Blue G in the reagent was allowed to occur at room temperature for 5 minutes and the optical density of 595 nm of the mixture was measured. The standard curve was performed using 0.25, 0.5, 1.0 and 1.4 mg/ml BSA and used to calculate the concentration of unknown samples.

CHAPTER 3 NATURAL VARIATION OF FOLATE BIOSYNTHESIS GENE EXPRESSION

3.1 INTRODUCTION

3.1.1 Folate biochemical pathway

Folate is a tripartite molecule consisting of pteridine, *p*-aminobenzoate or *p*ABA and one or more glutamate residues. Each of these molecules is produced in different subcellular compartments (Figure 3.1).

Enzymes localised in the cytosolic compartment are responsible for hydroxymethyldihydropterin (pteridine) synthesis. GTP is a substrate catalysed by three catalytic enzymes; the GTP-cyclohydrolase I (GTPCHI) enzyme to form dihydroneopterin; a nudix hydrolase enzyme to remove pyrophosphate group from dihydroneopterin and the dihydroneopterin aldolase (DHNA) enzyme to release hydroxymethyldihydropterin (H_2Pterin). The GTPCHI reaction is considered as the rate-determining step of pteridine synthesis in *E.coli* and mammals due to feedback inhibition by the end product, tetrahydrobiopterin (Schoedon *et al.*, 1992; Yoneyama *et al.*, 2001) unlike in plants; this end product is not produced (Basset *et al.*, 2002). In animals, this enzyme is a decamer protein containing N- and C-terminal domains with distinct substrate binding and catalytic sites (Basset *et al.* 2002; Auerbach *et al.*, 2000). In contrast, plant GTPCHI is a dimer molecule with these N- and C-termini (Basset *et al.*, 2002).

Recombinant tomato GTPCHI proteins having only N- or C-terminal domains do not show GTPCHI activity in yeast cell extracts (Basset *et al.*, 2002). Hence, both domains are required to have a complete set of GTPCHI activity. The alignment of GTPCHI cDNA sequences of *E.coli*, human, *Arabidopsis* and tomato shows similarly conserved regions in both binding and catalytic sites

and proteins encoded from those cDNAs have full enzymatic activity (Basset *et al.*, 2002).

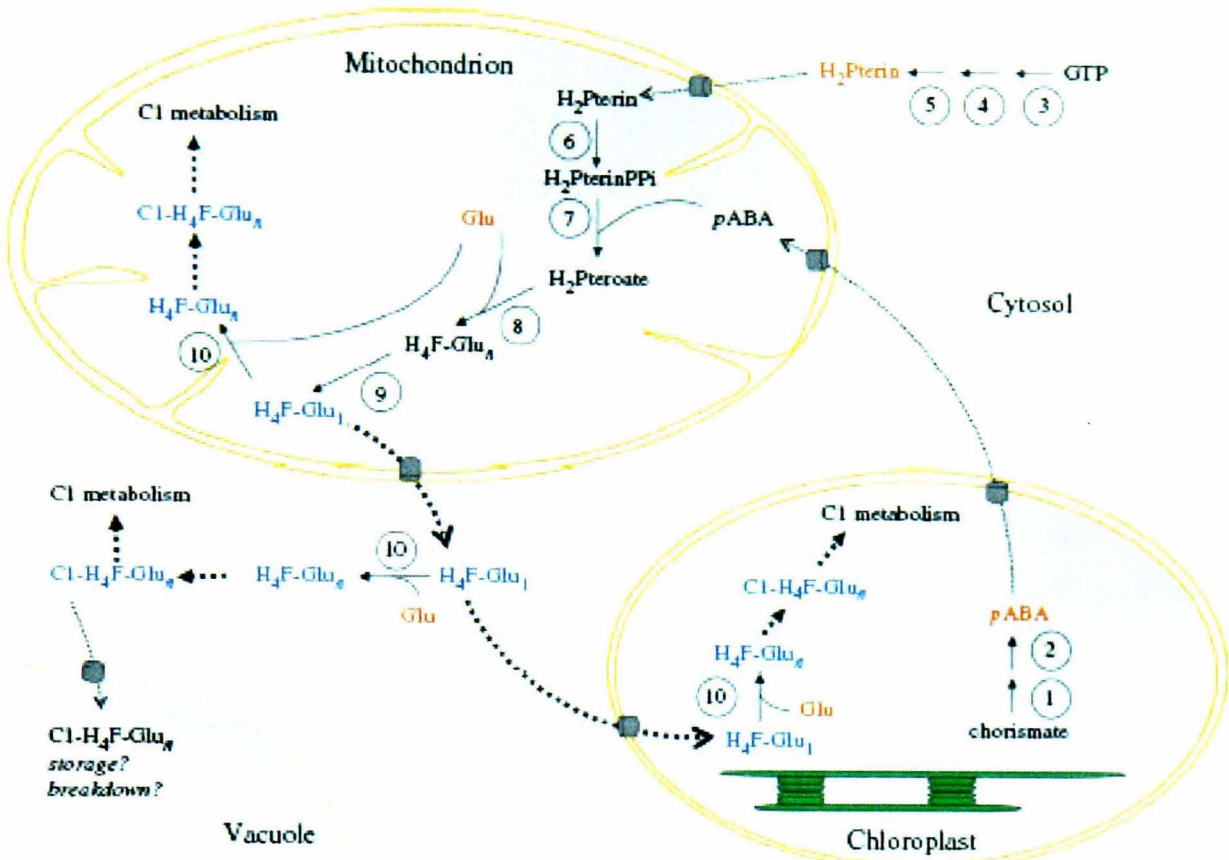


Figure 3.1 Folate biosynthesis pathway in plant cells

The enzymes involving in this pathway are (1) aminodeoxychorismate (ADC) synthase, (2) ADC lyase, (3) GTP cyclohydrolase I, (4) nudix hydrolase, (5) dihydroneopterin aldolase (DHNA), (6) hydroxymethyldihydropterin pyrophosphokinase (HPPK), (7) dihydropteroate synthase (DHPS), (8) dihydrofolate synthetase (DHFS), (9) dihydrofolate reductase (DHFR) and (10) folypolyglutamate synthetase (FPGS). H_2Pterin, hydroxymethyldihydropterin; H_2PterinPPi, hydroxymethyldihydropterin pyrophosphate; H_2Pteroate, dihydropteroate; H_4F-Glu$_n$, tetrahydrofolate. The picture was taken from Rébeillé *et al.*, 2006.

In cereals, the *GTPCHI* gene was first described in wheat by McIntosh *et al.* (2008). Wheat and rice *GTPCHI* nucleotide sequences exhibit 85% identity which shows very high conservation in the Poaceae family as they retain only 53% identity to *Arabidopsis GTPCHI* gene. Comparing protein structure in plants versus other eukaryotes, plants lack one important functional part in both

domains conserved in other eukaryotes, which is an EF-hand-like-motif required for calcium binding (Steinmetz *et al.*, 1998). However, plant GTPCHI is still active.

In chloroplasts, *p*-aminobenzoic acid (*p*ABA) is synthesized from chorismate which requires two main enzymes; aminodeoxychorismate (ADC) synthase catalyzes the alteration of chorismate into ADC and ADC lyase catalyzes the alteration of ADC into *p*ABA. During evolution, the mechanism of *p*ABA synthesis has been changed. *E.coli* synthesises *p*ABA in two steps with three separate peptides (Viswanathan *et al.*, 1995). The first reaction is catalysed by the heterodimeric complex consisting of the PabA and PabB proteins. It is named ADC synthase later on which transfer the amino group from glutamine to chorismate producing ADC molecule (Viswanathan *et al.*, 1995). The second step requires separately PabC peptides named ADC lyase to eliminate pyruvate and aromatize ADC ring into *p*ABA (Green *et al.*, 1992). In contrast, *p*ABA in plants is synthesised in plastids by two distinct enzymes; ADC synthase and ADC lyase. ADC synthase in plants is a bipartite polypeptide consisting of PabA- and PabB-like domains but lacks a lyase activity necessitating a separate ADC lyase enzyme to generate *p*ABA (Basset *et al.*, 2004a, b).

Hydroxymethyldihydropterin from the cytosol and *p*ABA from the plastid are coupled together in the mitochondria by a bifunctional enzyme, hydroxymethyldihydropterin pyrophosphokinase (HPPK) and dihydropterate synthase (DHPS), to form dihydropterate (H_2Pteroate). HPPK/DHPS was firstly identified in bacteria as two separated monofunctional proteins (Slock *et al.*, 1990) while plant and *Plasmodium* carried the bifunctional protein for these two activities (Triglia *et al.*, 1994). Pea, wheat and rice genomes contain a single gene coding for the putative mitochondria HPPK/DHPS enzyme while two genes are presented in the *Arabidopsis* genome, one encoded for mitochondria isoforms and another gene encoded for cytosolic form (Rébeillé *et al.*, 1997, Jabrin *et al.*, 2003, Storozhenko *et al.*, 2005, McIntoch *et al.*, 2008). From kinetic studies of these two enzymes, only DHPS is strongly

feedback-inhibited by dihydropteroate and also dihydrofolate and tetrahydrofolate monoglutamate. Thus, the DHPS domain may be a key regulatory part of the entire folate biosynthetic pathway (Mouillon *et al.*, 2002).

The later step is an attachment of glutamate residues to dihydropterate molecule to form dihydrofolate, tetrahydrofolate and tetrahydrofolate polyglutamates which are catalyzed by dihydrofolate synthetase (DHFS), dihydrofolate reductase (DHFR) and folylpolyglutamate synthetase (FPGS), respectively. Loss of HPPK, DHPS and DHFS enzymes in animal cells blocks folate biosynthesis (Scott *et al.*, 2000). DHFS and FPGS activities in some bacteria exist as a bifunctional protein encoded by the *folC* gene (Cossins and Chen, 1997), whereas in fungi and plants, it is a monomeric protein (Cossins and Chen, 1997). In *Arabidopsis*, there is a single gene encoding DHFS whereas there are three genes encoding DHFR and FPGS enzymes (Ravanel *et al.*, 2001). DHFS is located in the mitochondria making this compartment the sole site of dihydrofolate synthesis in plant cells. On the contrary, distinct FPGS isoforms exist in the mitochondria, cytosol, and chloroplast. The details about FPGS enzymes and genes will be described more in chapter 4.

The mitochondria, the main site of folate synthesis, contains about 40% of the total folate, whilst, the cytosol, the vacuoles and the plastids contain 30%, 20% and 10% of synthesized folates, respectively (Rébeillé *et al.*, 2006). Hence, folate transporters must exist in the mitochondria to carry folates to other compartments. From animal studies, polyglutamylated folate forms cannot cross membranes, whereas folic acid or dihydrofolate can rapidly move. As glutamate is an acidic amino acid, polyglutamate forms of folate would have a highly negative charge and thus be retained in the place they are synthesized. Hence, the glutamylation process will facilitate the retention of folate in organelles (Cybulski and Fisher, 1981). It seems to correlate with having three isoforms of *FPGS* presented in three different cell compartments (Scott *et al.*, 2000) in *Arabidopsis* in order to retain folate and specifically use folate in

those compartments. Folates need to be depolyglutamylated into its monoglutamate form prior to being transported across cell membranes.

The folate transport system and folate transporters in plants have not been characterised to date. Transporters must be involved in several synthesis steps (Figure 3.1) including 1) transporting *p*ABA out of the plastids into the mitochondria; 2) transporting pterin from the cytosol to the mitochondria; 3) transporting the monoglutamate forms out of the mitochondria into the cytosol and the chloroplasts; 4) transporting polyglutamylated forms into the vacuole for degradation (Rébeillé *et al.*, 2006); and 5) folate uptake from outside of cells (Basset *et al.*, 2005). According to these transport systems, folate can be distributed to several cellular compartments at different levels.

Folate catabolism and turnover is poorly understood in plants. In animals, folate is spontaneously oxidized via a non-enzymatic process producing two main metabolites, the pteridine ring obtained from the cleavage between C_9 and N_{10} units and hydrolyzed glutamate moieties attached to *p*ABA residues, which are consequently excreted and normally found in urine (Scott *et al.*, 2000). In plants, the breakdown products might not be excreted but turned over in specific compartments to recycle molecules in the new synthesis pathway. Orsomando *et al.* (2006) reported that folate salvage pathways also exists in plants and are important to the folate pool size by comparing the amount of by-products from detached and attached fruits. The breakdown rate of folate pool is estimated to be very high, about 10% per day, and the cleaved products normally enter the salvage process so that the total folate pool of the attached fruits does not change much. This conclusion is supported by evidence obtained from tomatoes and *Arabidopsis*. Total folate levels of harvested tomatoes declined by 62% in 7 days compared to a little change in attached tomatoes (Orsomando *et al.*, 2006). Hence, the high degradation rate of folate occurs naturally and does not affect the total folate pool much. *p*ABAGlu and pterin products were also quantified to prove whether they were accumulated after degradation. Both of them were not accumulated in *Arabidopsis* leaves

which implied the existing of salvage activity in plant cells (Orsomando *et al.*, 2006).

According to the high level of polyglutamylated folate in plants, catabolic studies have focused on the hydrolysis of the polyglutamate chain by γ-glutamyl hydrolase (GGH) enzyme in the vacuoles where 5-methyltetrahydrofolate was predominantly found (Orsomando *et al.*, 2005). In *Arabidopsis*, the GGH enzyme appears to play roles mainly in the vacuoles in either cleaving polyglutamate folates or folate storage (Orsomando *et al.*, 2005). There are at least two specificities described for GGH enzymes. One gene cleaves pentaglutamates mainly to di- and triglutamates, while another provides mostly monoglutamates. This enzyme affects the folate pools in plant cells therefore its reaction is suggested as an important step in the regulation of folate transport between cell compartments (Orsomando *et al.*, 2005). Recently, Noiriel *et al.* (2007) and Bozzo *et al.*, (2008) reported additional ways of folate salvage in plants through pteridine aldehyde reduction system (PTAR) and *p*-aminobenzoylglutamate hydrolase enzyme (PGH), respectively, as shown in Figure 3.2.

After tetrahydrofolate polyglutamates ($THFGlu_n$) are cleaved during the plant salvage process. GGH and PGH generate free *p*ABA. Even PGH activity has been found in *Pseudomonas* carboxypeptidase G-like protein (CPG), there is only one enzyme known having this activity (McCullough *et al.*, 1971). Similar proteins in plants have not been identified and no homologs of CPG have been found in plants (Bozzo *et al.*, 2008). There is only PGH activity could be detected by 89% in cytosol plus vacuoles parts without recognising the specific protein (Bozzo *et al.*, 2008). Thus, the protein with PGH activity might be either a rare novel protein or a known enzyme exhibiting PGH activity.

For the cleaved pterin molecules, dihydropterin-6-aldehyde (DHPA) is a stable end-product form of pterin found *in vivo* (Orsomando *et al.*, 2006). It can be altered to the intermediate hydroxymethyldihydropterin (HMDHP) by NADPH-dependent reductase (PTAR) activity in the cytosol and recycled into

folate synthesis (Noiriel *et al.*, 2007). Unlike PGH, the *Arabidopsis* PTAR protein encoded by At1g10310 gene is present at high abundance in dry seeds (Noiriel *et al.*, 2007). To complete the salvage process, GGH, PGH and PTAR enzymes are needed in the plant system.

Greater understanding of folate metabolism in *Arabidopsis* will aid development of strategies for crop improvement.

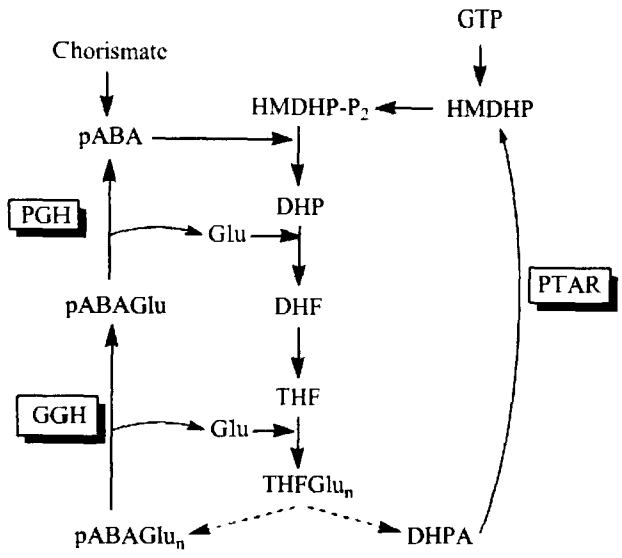


Figure 3.2 Folate salvage pathway in plants

The $pABAGlu_n$ molecules may be cleaved by γ-glutamyl hydrolase (GGH) enzyme prior to pABAGlu hydrolase activity (PGH). Free *p*ABA can be reused in folate synthesis. The pterin part (dihydropterin-6-aldehyde or DHPA) can be recycled by changing to the intermediate molecule, HMDHP, through the action of pterin aldehyde reductase (PTAR). This picture is modified from Bozzo *et al.* (2008).

3.1.2 Rice development stages

Rice is in the genus *Oryza* which contains 21 species but only two, *Oryza sativa* originating from Southeast Asia and *Oryza glaberrima* originating from West Africa, have been cultivated (Juliano, 1985). To date, many thousands of rice varieties grown and distributed around the world are *Oryza sativa* and these can be divided into 3 main sub-species, *indica*, *japonica* and *javanica*. *Indica* and *japonica* are the most common varieties in all markets, while, *javanica* is rarely found (Kennedy and Burlingame, 2003).

There are 3 main rice developmental stages which are seedling, vegetative and reproductive stages. Classification of stages in this thesis will rely on the reference guide of the International Rice Research Institute (IRRI) which is publicly available at www.knowledgebank.irri.org. The growth stages of rice plant are explained in a numerical scale of 0-9 (see Figure 3.3).

Stage 0 is germination stage of which the primary leaf will emerge on the third day after germination with elongation of radicles. Stage 1 is seedling. The leaf and root are developed at this stage, the young seedling at around the 18^{th} day is ready for transplanting. Stage 2 is tillering where stem and tiller are developed to have a maximum tiller number. This stage is started by the 30^{th} day after transplanting. Stage 3 is stem elongation which can overlap with stage 2 and begins before panicle initiation. Tiller and stem will be increased by number and height, respectively. Stage 4 is panicle initiation and booting. Stage 5 is heading of panicles protruding from the flag of leave sheets. Stage 6 is flowering, where anthers protrude from the spikelet and fertilization happens. Stage 7 is the milk grain stage which the fertilized grain begins to fill with a white and milky liquid. Stage 8 is the dough grain stage. The milky portion of grain will be turned to soft dough and then hard dough. After that, the filled grains become yellow indicating the senescence stage of rice plants. Finally, stage 9 is the mature grain stage. Mature grains are filled, hard and turned to yellow and brown. Usually, the rice developmental cycle takes about 120 days in tropical weather.

3.1.3 Rice grain composition

Grains have unique characteristics among rice varieties; some have a special fragrance such as Thai jasmine *indica* rice. *Indica* grains are medium to long, narrow and flat compared to *Japonica* grains which are shorter and wider. Also, *japonica* varieties tend to be more tolerant to cold temperature but less tolerant to drought, insects and disease than *indica* varieties which are mostly found in tropical countries (Kennedy and Burlingame, 2003).

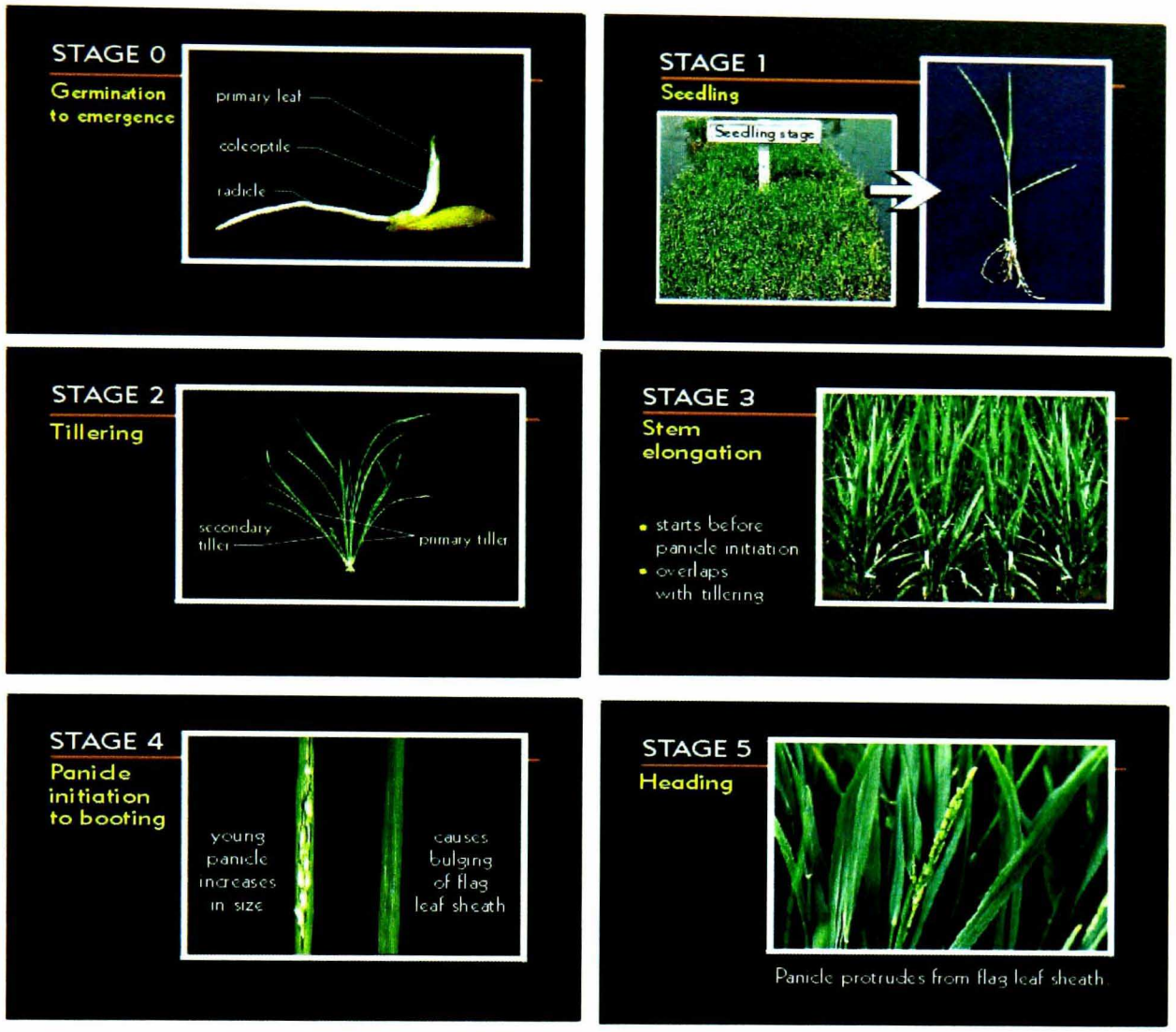


Figure 3.3 The nine developmental stages of rice plants

(www.knowledgebank.irri.org)

All rice grains share the same feature composing of the brown hull and the bran in the outer layers which can be removed by milling. The endosperm and embryo are the inner layers which are consumed as part of the human diet (see Figure 3.4).

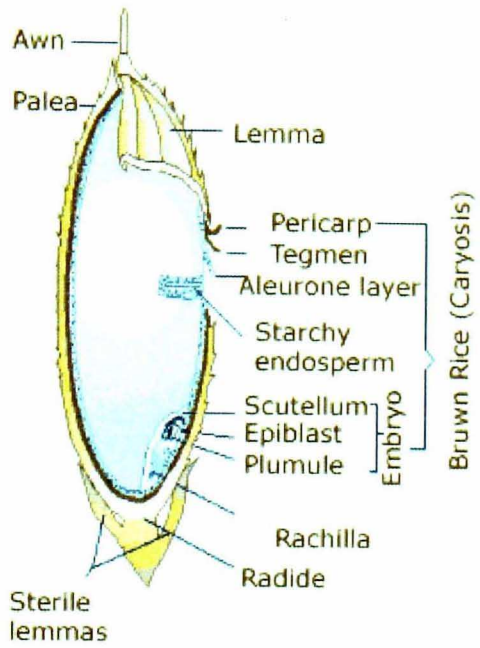


Figure 3.4 The structure of rice grain (www.foodmarketexchange.com)

The bran layer contains most of the nutrients including fibres, vitamin B, protein and fat. The review of Kennedy and Burlingame (2003) reported the huge variation of nutrient composition in unpolished rice grains for over 2000 varieties. The average protein assayed by IRRI was about 8.55 g/100g with the highest protein variety, Chinese fragrant long grain rice, containing up to 15g/100g. Additionally, endosperm and embryo parts contain a range of amylose content which was about 0.5-33 g/100g. The stickiness of cooked rice is determined by amylose whilst lower amylose cultivars, *Japonica*, are stickier and softer. They also reported the levels of other micronutrients which are iron, zinc, thiamine (B1), riboflavin (B2), niacin and calcium. It was quite a wide range in these micronutrients, 0.70-6.35 mg/100g for iron, 0.79-5.89 mg/100g for zinc, 0.117-1.74 mg/100g for thiamine, 0.011-0.403 mg/100g for riboflavin, 1.972-9.218 mg/100g for niacin and 1-65 mg/100g for calcium. However, the highest vitamin content was reported in red and purple rice varieties whilst

high iron and zinc has been linked to aromatic varieties such as jasmine and basmati rice (Graham *et al.*, 2001).

Considering the folate content in rice grains, very few studies have reported folate distribution in traditional rice. Quantification of folates using tri-enzymes treatment and microbiological assay suggested 25 µg/100g in unfortified raw white rice and 9 µg/100g in 30 min boiled rice (Han *et al.*, 2005). Unfortunately, no data exists showing the amount of folate between rice varieties and also rice tissues during development.

3.2 AIMS AND OBJECTIVES

In order to enhance the total folate level, rice can be genetically manipulated using several different approaches (chapter 4 and 5) such as overexpression of genes in the folate biosynthesis pathway (for review see Bekaert *et al.*, 2007). Before that, this chapter aims to explore the rice folate biosynthesis pathway and at genetic regulation. The transcript abundance of folate-synthesis related genes was analysed using Reverse Transcription quantitative Polymerase Chain Reaction (RT-qPCR) among rice varieties. To do this, a number of genes encoding folate related enzymes have been identified in rice based on homology to *Arabidopsis* folate related genes. Differences in mRNA levels are then related to variation in folate abundance in rice varieties. The discussion will identify the target genes required to manipulating folate in rice using either marker assisted breeding or GM-based approaches.

3.3 RESULTS

3.3.1 Bioinformatic identification of genes encoding folate related biosynthetic enzymes in rice

There are 9 enzymes involved in the folate pathway (Figure 3.5) including GTP cyclohydrolase I (GTPCHI); dihydroneopterin aldolase (DHNA); aminodeoxychorismate synthase (ADC synthase); aminodeoxychorismate lyase (ADC lyase); hydroxymethyldihydropterin pyrophosphate kinase (HPPK); dihydropteroate synthase (DHPS); dihydrofolate synthase (DHFS); dihydrofolate reductase (DHFR); and folypolyglutamate synthase (FPGS); plus gamma-glutamyl hydrolase (GGH), involved in the folate salvage pathway.

BLASTP was firstly used as a search tool to identify rice folate related enzymes using peptide sequences of folate related enzymes in *Arabidopsis*. The protein identity of *Arabidopsis* and rice enzymes was compared and shown in Table 3.1. Based on the percentage of protein identity between *Arabidopsis* and rice, homologous enzymes in these plant species exhibit 45-75% similarity at the peptide sequence level.

Since folate related enzymes are located in three different subcellular compartments of plants (Figure 3.1), enzyme putative localization was searched through the experiment reports reviewed by Storozhenko *et al.* (2005) to confirm the place of reaction for each enzyme and are summarized in Table 3.2. Compared to the reports of *Arabidopsis* proteins, many rice enzymes are predicted to be similarly localized. However, localization experiments for rice enzymes have not been performed, hence it relies on using targeting prediction software such as TargetP 1.1.

Firstly, in the pteridine synthesis branch, one gene encoding GTPCHI and three genes encoding DHNA were identified in both *Arabidopsis* and rice genomes. Their proteins are predicted to lack targeting signals. Hence, it is likely that these two enzyme activities are located in the cytoplasm (Storozhenko *et al.*, 2005).

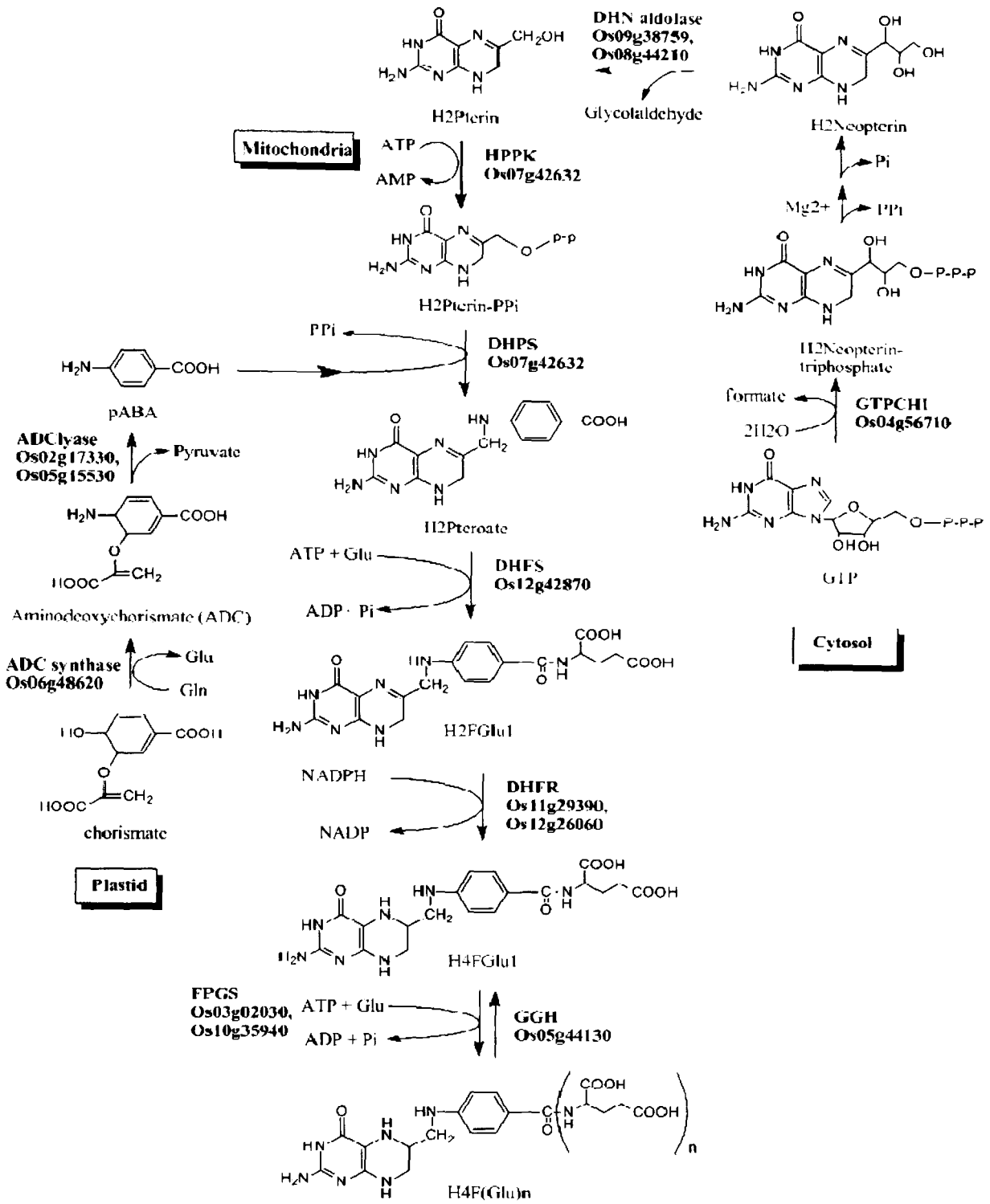


Figure 3.5 Folate biosynthesis pathway

Enzymes responsible in metabolism are in bold; GTP cyclohydrolase I (GTPCHI), dihydroneopterin aldolase (DHN aldolase), aminodeoxychorismate synthase (ADC synthase), aminodeoxychorismate lyase (ADC lyase), hydroxymethyldihydropterin pyrophosphate kinase (HPPK), dihydropteroate synthase (DHPS), dihydrofolate synthase (DHFS), dihydrofolate reductase (DHFR) and folypolyglutamate synthase (FPGS) and gamma-glutamyl hydrolase (GGH). The Gramene gene codes encoding these enzymes in rice are presented in bold along with the enzyme names.

Table 3.1 Protein identity of folate-related enzymes among *Arabidopsis* and rice

Name of enzymes	gene code in *Arabidopsis*	gene code in Rice	%protein identity
GTP cyclohydrolase I	AT3G07270	LOC_Os04g56710	44
DHN aldolase	AT3G11750	LOC_Os09g38759	56
		LOC_Os06g06100	57
		LOC_Os08g44210	58
	AT3G21730	LOC_Os09g38759	52
		LOC_Os06g06100	50
		LOC_Os08g44210	49
	AT5G62980	LOC_Os09g38759	64
		LOC_Os06g06100	58
		LOC_Os08g44210	60
ADC synthase	AT2G28880	LOC_Os06g48620	52
ADC lyase	AT5G57850	LOC_Os05g15530	53
		LOC_Os02g17330	64
hydroxymethyldihydropterin pyrophosphokinase (HPPK)/ dihydropteroate synthase (DHPS); bifunctional protein	AT4G30000	LOC_Os07g42632	64
	AT1G69190	LOC_Os07g42632	61
dihydrofolate synthetase (DHFS)	AT5G41480	LOC_Os12g42870	63
dihydrofolate reductase (DHFR) ; DHFR/TS	AT2G16370	LOC_Os11g29390	75
		LOC_Os12g26060	75
	AT2G21550	LOC_Os11g29390	52
		LOC_Os12g26060	51
	AT4G34570	LOC_Os11g29390	71
		LOC_Os12g26060	73
folylpolyglutamate synthetase (FPGS)	AT5G05980	LOC_Os03g02030	57
		LOC_Os10g35940	56
	AT3G55630	LOC_Os03g02030	52
		LOC_Os10g35940	51
	AT3G10160	LOC_Os03g02030	52
		LOC_Os10g35940	50
gamma-glutamyl hydrolase (GGH)	AT1G78660	LOC_Os05g44130	59
	AT1G78670	LOC_Os05g44130	55
	AT1G78680	LOC_Os05g44130	57

Gene code in *Arabidopsis* was searched by enzyme names in the *Arabidopsis* AtEnsembl genome database (http://atensembl.arabidopsis.info/) and the *Arabidopsis* information resource (TAIR) (http://www.arabidopsis.org/index.jsp). *Arabidopsis* protein was blasted for the orthologs in rice in the Gramene database (http://www.gramene.org/) and the TIGR rice genome annotation database (http://www.tigr.org/tdb/e2k1/osa1/). Protein identity and gene code in rice were based on the Gramene database.

Table 3.2 Enzyme localization from each folate-related gene in *Arabidopsis* and rice

Name of enzymes	gene code in *Arabidopsis*	localization	gene code in Rice	localization
GTP cyclohydrolase I	AT3G07270	cytoplasm	LOC_Os04g56710	cytoplasm
DHN aldolase	AT3G11750	cytoplasm	LOC_Os09g38759	cytoplasm
	AT3G21730	cytoplasm	LOC_Os06g06100	cytoplasm
	AT5G62980	cytoplasm	LOC_Os08g44210	cytoplasm
ADC synthase	AT2G28880	plastids	LOC_Os06g48620	plastids/mitochondria?
ADC lyase	AT5G57850	plastids	LOC_Os05g15530	cytoplasm
			LOC_Os02g17330	plastid
hydroxymethyldihydropterin pyrophosphokinase (HPPK)/ dihydropteroate synthase (DHPS); bifunctional protein	AT4G30000	mitochondria	LOC_Os07g42632	mitochondria
	AT1G69190	cytoplasm		
dihydrofolate synthetase (DHFS)	AT5G41480	mitochondria	LOC_Os12g42870	mitochondria
dihydrofolate reductase (DHFR) ; DHFR/TS	AT2G16370	plastids	LOC_Os11g29390	mitochondria?
	AT2G21550	cytoplasm	LOC_Os12g26060	plastids?
	AT4G34570	mitochondria		
folylpolyglutamate synthetase (FPGS)	AT5G05980	plastids	LOC_Os03g02030	mitochondria?
	AT3G55630	cytoplasm	LOC_Os10g35940	plastids?
	AT3G10160	mitochondria		
gamma-glutamyl hydrolase (GGH)	AT1G78660	vacuole	LOC_Os05g44130	vacuole
	AT1G78670	vacuole		
	AT1G78680	vacuole		

Gene code in *Arabidopsis* was obtained from the *Arabidopsis* information resource (TAIR) (http://www.arabidopsis.org/index.jsp). Gene code in rice was obtained from the Gramene database (http://www.gramene.org/). The localization of each enzyme was reviewed by Storozhenko *et al.* (2005). ? in the localization column indicates no data existing to confirm the location of enzymes.

For the enzymes ADC synthase and ADC lyase in the *p*ABA biosynthesis branch, the number of genes and their localization appears not to be conserved between rice and *Arabidopsis*. Although one gene encoding for ADC synthase was identified in *Arabidopsis* and cereals, their localization appears to be different. *Arabidopsis* ADC synthase employing a GFP fusion in *Arabidopsis* protoplasts was localized in plastid (Basset *et al*, 2004a). In contrast, the rice ADC synthase enzyme was predicted to have both chloroplast and mitochondria signals from TargetP software, but without experimental evidence. However, wheat ADC synthase was predicted to have a chloroplast signal peptide even though its overall protein sequence showed little similarity to *Arabidopsis* enzyme (McIntoch *et al.*, 2008).

In the case of ADC lyase, the *Arabidopsis* enzyme is encoded by a single gene with a chloroplast transit peptide. The rice ADC lyase enzyme is encoded by two distinct genes; the LOC_Os05g15530 gene does not include any transit peptide whilst the second gene LOC_Os02g17330 is predicted to contain a chloroplast signal peptide.

In terms of THF synthesis, the HPPK/DHPS and DHFS enzymes in rice are encoded by single genes and their deduced amino acid sequences are predicted to contain mitochondria targeting peptides. In contrast, HPPK/DHPS in *Arabidopsis* is encoded by two genes which contain mitochondria and cytoplasm transit signals (Neuburger *et al.*, 1996; Storozhenko *et al.*, 2007). Even though HPPK/DHPS in rice is located only in the mitochondria, this compartment is proposed as the main site of folate synthesis in plant cells. Indeed, Neuburger *et al.* (1996) reported that folate content in pea leaf mitochondria was 150 times higher than in the chloroplasts. About 400 μM in the mitochondria and 1.7 μM in the chloroplasts were the estimated folate concentrations with mitochondria representing about 50% of the total folate pool. Not only in pea, potato tuber mitochondria also contained high folate levels at 200 μM (Neuburger *et al.*, 1996). Thus, mitochondria appear to represent the major source of synthesized folate in plant cells.

For DHFR and FPGS enzymes, three genes were found in *Arabidopsis* and located in plastid, mitochondria and cytoplasmic compartments (Ravanel *et al.*, 2001). Rice, however, contained two genes for these two enzymes and no signal sequence was identified in their deduced amino acid sequences (Storozhenko *et al.*, 2005). Although Storozhenko *et al.* (2005) concluded that DHFR and FPGS enzymes from both genes in rice were located only in the cytoplasm, TargetP software predicted that both mitochondria and chloroplast signal peptides could be found in their sequences. For DHFR, a mitochondria signal sequence could be found in LOC_Os11g29390 and a plastid signal sequence in LOC_Os12g26060. Also, mitochondria and plastid targeting signals for FPGS could be predicted from LOC_Os03g02030 and LOC_Os10g35940, respectively, with very high probabilities.

Although no experimental reports exist about localization of folate biosynthesis enzymes in rice, Neuburger *et al.* (1996) detected most of these tetrahydrofolate synthesis enzymes in pea leaf mitochondria. Very high HPPK/DHPS, DHFS, DHFR and FPGS enzyme activity was detected in the mitochondria, while, they could not detect the activity in the chloroplasts, the cytosol or the nuclei. These tetrahydrofolate synthesis enzymes therefore appear to be primarily located in the mitochondria although other compartments contain very low levels. Hence, those enzymes should be targeted into mitochondria even though no mitochondria signal sequence can be detected.

In the folate salvage pathway, only one gene encoding GGH enzyme is found in rice. In contrast, *Arabidopsis* encodes 3 distinct genes. In *Arabidopsis*, this enzyme encoded a predicted secretory pathway signal peptide to the vacuoles. Enzymes from two isoforms, At1g78660 and At1g78680 were very similar to mammalian GGHs in the conserved glycolation and cysteine-histidine regions which are predicted to be enzyme binding sites. The enzyme encoded by At1g78660 cleaved pentaglutamates mainly to di- and triglutamates, while the At1g78680 generated mostly monoglutamates (Orsomando *et al.*, 2005). The third GGH enzyme (At1g78670) was quite distinct from other GGH members.

Compared to *Arabidopsis*, the single rice GGH gene was much more similar to *Arabidopsis* GGHs At1g78660 and At1g78680 genes (59% and 57% identity, respectively) than At1g78670 (55% identity) (see the percentage of protein identity in Table 3.1).

In summary, folate synthesis mechanism is generally conserved in three distinct cellular compartments in plants. In terms of the specific enzymes for pteridine, *p*ABA and THF synthesis, the localization of each enzyme has been clarified mostly in *Arabidopsis*. Using signal targeting prediction software, most of the rice folate-related enzymes share similarity to *Arabidopsis* enzymes with the exception of ADC synthase, DHFR and FPGS enzymes. However, these results were only predictions. The precise localization of these enzymes should be further investigated experimentally to confirm the common feature of folate biosynthesis pathway in the plant system.

After protein identification, genomic sequences encoded for each enzyme in rice were identified based on the TAIR rice database. The lists of rice gene coding for all 10 enzymes are presented in Table 3.3. TAIR blast results were also double checked using the Gramene database (www.gramene.org). General information such as number of exons and the size of genes, transcripts and proteins obtained from Gramene database is summarized in Table 3.3.

Table 3.3 General information of rice folate-related genes

Name of enzymes	gene code in Rice	Accession no.	No. of exons	Genomic sequence length (bp)	transcript length (bp)	peptide length (residue)
GTP cyclohydrolase I	LOC_Os04g56710	NP_001054158	3	4,585	2,031	367
DHN aldolase	LOC_Os09g38759	NP_001063927	4	2,025	1,064	208
	LOC_Os06g06100	NP_001056847	2	1,114	918	138
	LOC_Os08g44210	NP_001062473	2	1,142	713	135
ADC synthase	LOC_Os06g48620	NP_001058474	2	1,688	1,591	324
ADC lyase	LOC_Os05g15530	NP_001055013	5	2,109	1,291	340
	LOC_Os02g17330	NP_001046531	5	4,307	1,777	389
Hydroxymethyldihydropterin pyrophosphokinase (HPPK)/dihydropteroate synthase (DHPS)	LOC_Os07g42632	NP_001060287	3	2,430	2,219	532
dihydrofolate synthetase (DHFS)	LOC_Os12g42870	NM_001073844	10	4,845	1,797	510
dihydrofolate reductase (DHFR) ; DHFR/TS	LOC_Os11g29390	NP_001067911	12	7,424	2,542	521
	LOC_Os12g26060	DRTS_ORYSJ	10	5,936	1,632	494
folylpolyglutamate synthetase (FPGS)	LOC_Os03g02030	NP_001048721	16	5,020	2,118	554
	LOC_Os10g35940	NP_001064996	15	5,201	2,070	577
gamma-glutamyl hydrolase (GGH)	LOC_Os05g44130	NP_001056052	10	4,368	1,327	337

Gene code in rice and all gene and protein information was obtained from the Gramene database (http://www.gramene.org/).

3.3.2 Folate-related gene expression profiling in Nipponbare wild type

Expression profiling of genes encoding enzymes in the rice folate pathway was performed. The expression data can be used as the basis for selecting suitable genes for engineering rice and also to distinguish the differences of expression level among rice tissues. Fortunately, microarray data of rice folate-related gene expression was provided in RiceGE (Rice Functional Genomic Express Database; http://signal.salk.edu/cgi-bin/RiceGE2) to use for cross-experiment comparison.

Due to the complete genome sequences of rice being publicly available, 13 folate genes were identified and their transcript abundance was measured by RT-qPCR. RT-qPCR probes were designed for all 13 folate-related genes as shown in appendix 1 based on the sequences of wild type Nipponbare *japonica* cultivar. Some enzymes in rice, DHNA, ADC lyase, DHFR and FPGS were encoded by more than 1 gene. Nucleotide alignments were generated to analyse the conserved and divergent regions of their cDNA sequences in order to select specific regions within each gene to design RT-qPCR primers (see appendix 3). This is necessary to distinguish genes with very high similarity.

Considering firstly the quality of cDNA synthesized from total RNA (see method 2.3.2), examination of genomic contamination was performed by tradition PCR with rice actin primers. No positive actin band in the no-RT control was the criteria indicating no genomic contamination in the samples (Figure 3.6). These samples were further used in qPCR reactions.

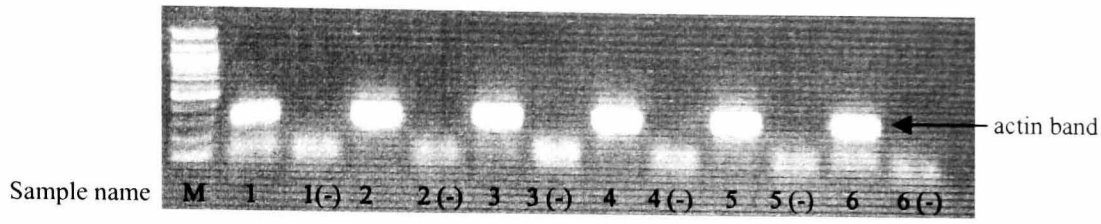


Figure 3.6 Examination of genomic contamination in cDNA samples.

Tradition PCR was performed after cDNA synthesis of RNA samples. The samples with no positive actin band in the no-RT control [1(-), 2(-), 3(-), 4(-), 5(-) and 6(-)] indicate no DNA contamination in all six samples. M represents 1 kb DNA marker.

Data analysis of all qPCR assays were calculated for the absolute transcript levels obtained from the standard curve (Figure 3.7C). A serial dilution of control template of known concentrations was used to prepare the standard curve and the log of the initial template concentration (in nanogram) was plotted against the Ct from each dilution. The standard curve was generated for every single qPCR experiment and all samples were compared within the same run. After completion PCR amplification, data was analysed with the Rotor-gene 6000 series software 1.7 (Corbett life science). To maintain consistency, the threshold baseline was set automatically at fluorescence value of 0.00935 as shown in Figure 3.7B to obtain the threshold cycle (Ct) of samples. Quantification of the samples was calculated from the Ct value by interpolation from the standard curve to yield the initial template quantities of unknown samples reported in nanogram. To compare the amount of transcripts of each folate-related gene within and between samples, normalization was performed in all samples by dividing the amount of transcripts of each gene obtained from the standard curve with the amount of actin transcripts obtained from the standard curve within the same sample.

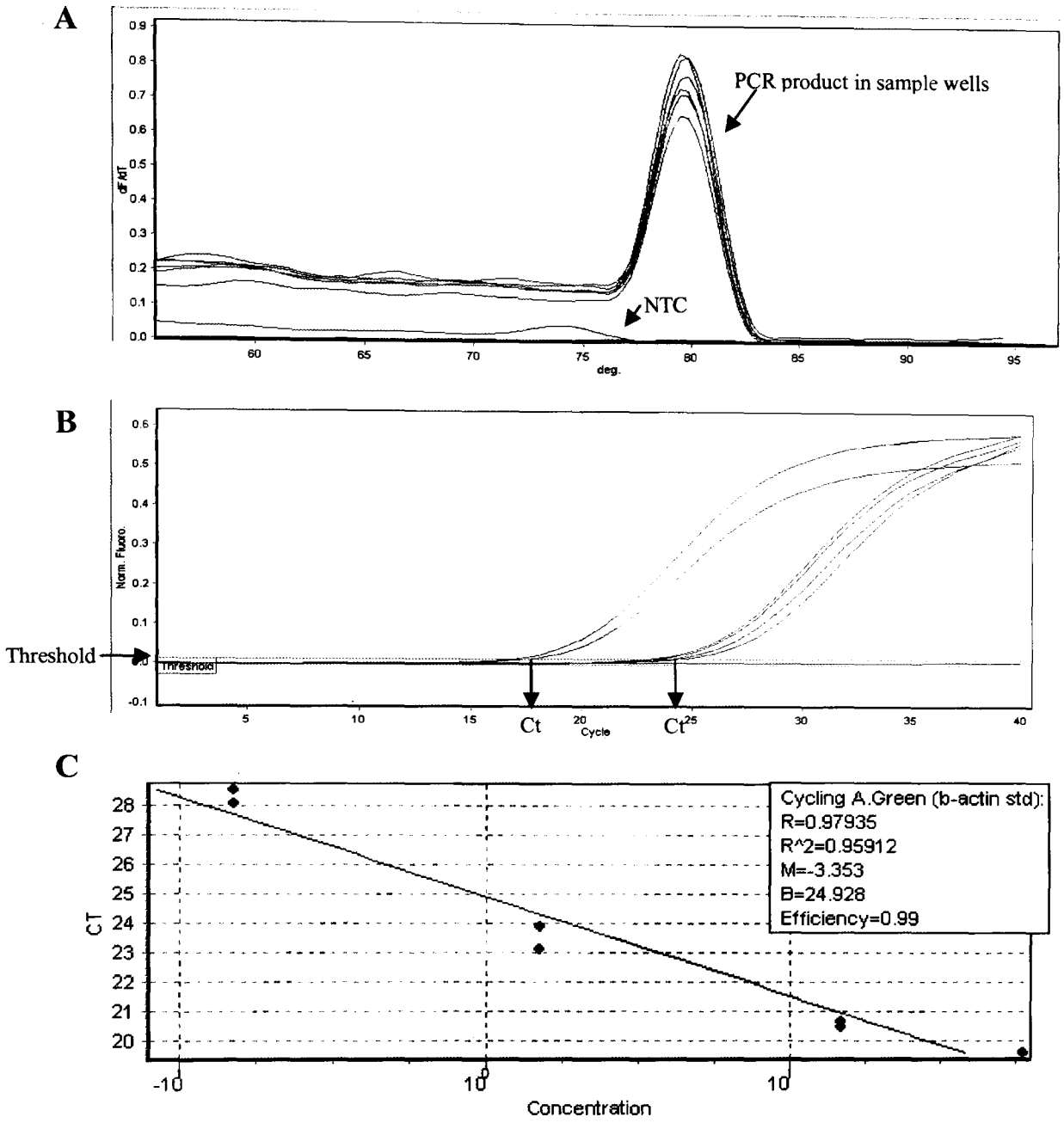


Figure 3.7 Melting point analysis, standard curve and threshold determination of RT-qPCR assays.

A: An example of the melting curve comparing the products obtained from the sample wells and no-template control (NTC) showing the specific amplification with a single peak. B: Threshold base line was set in the region where all plots were in the exponential phase and above the background fluorescence. A threshold Ct value was recorded for each sample; C: Standard curve as generated from known DNA concentration using beta-actin primers for analysis of all rice samples. The series of DNA concentration were at 60 ng/µl, 15 ng/µl, 1.5 ng/µl and 150 pg/µl DNA. Each point was run in duplicate. The linear curve was plotted against initial DNA concentration and threshold cycle (Ct). In this assay, the efficiency of the reaction was 99% with the amplification efficiency value of 0.97935 (R). Comparing Ct values of samples to this standard curve would provide the initial template quantities of unknown samples.

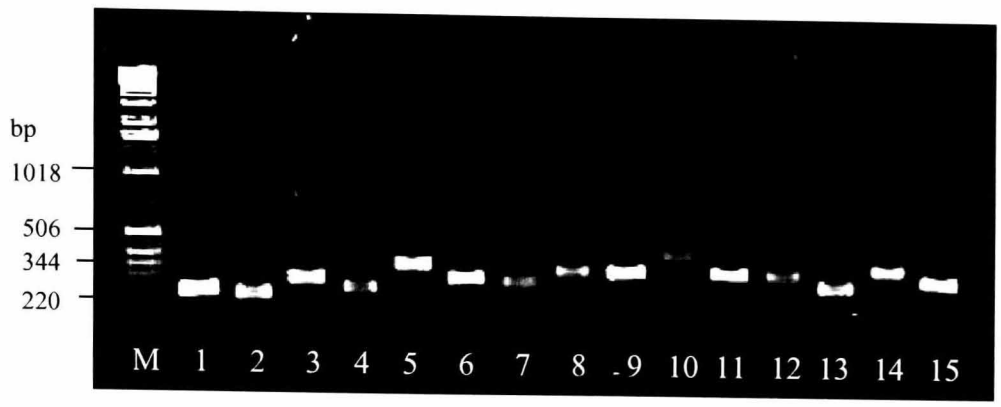


Figure 3.8 PCR products from RT-qPCR.

After the first RT-qPCR, the PCR products from each primer were run on 1% agarose gel electrophoresis. Each lane represents the product from a particular gene. Lane 1 *GTPCHI* LOC_Os04g56710; lane 2 *DHNA* LOC_Os09g38759; lane 3 *DHNA* LOC_Os08g44210; lane 4 *ADC synthase* LOC_Os06g48620; lane 5 *ADC lyase* LOC_Os02g17330; lane 6 *ADC lyase* LOC_Os05g15530; lane 7 *HPPK/DHPS* LOC_Os07g42632; lane 8 *DHFS* LOC_Os12g42870, lane 9 *DHFR* LOC_Os11g29390; lane 10 *DHFR* LOC_Os12g26060; lane 11 *FPGS* LOC_Os03g02030; lane 12 *FPGS* LOC_Os10g35940; lane 13 *GGH* LOC_Os05g44130; lane 14 actin Ta 55°C and lane 15 actin Ta 58°C (for details of primers and product sizes, see appendix 1). M: 1 kb ladder.

Transcript abundance of all 13 genes was initially profiled in Nipponbare rice leaves and dry grains. The average absolute expression value and standard deviation (SD) of each gene in Nipponbare rice are shown in Figure 3.9. Transcript levels in leaves and grains showed significant differences especially in the key genes, *GTPCHI*, *ADC synthase*, *HPPK/DHPS* and *FPGS*. Considering the pteridine synthesis branch, the gene encoding GTPCHI was more highly expressed than both genes encoding the DHNA. In the *p*ABA synthesis branch, the *ADC synthase* gene was more highly expressed than both *ADC lyase* genes. Also, in the middle branch of tetrahydrofolate synthesis, *HPPK/DHPS* was the highest expressed gene. Hence, the first enzyme of each branch appeared to be the most highly expressed gene. Interestingly, the mRNA for the folate polyglutamylation enzymes, *FPGS* LOC_Os03g02030 was more abundant than *FPGS* LOC_Os10g35940 in dry grains, whilst the mRNA of *FPGS* LOC_Os10g35940 was higher in leaves. It was suggested that

two FPGS isoforms encoded by these two *FPGS* genes may function in the different stages of rice development. In addition, the folate salvage pathway gene, *GGH*, was also expressed at a very high level in both leaves and grains indicating the high level of turnover in both tissues.

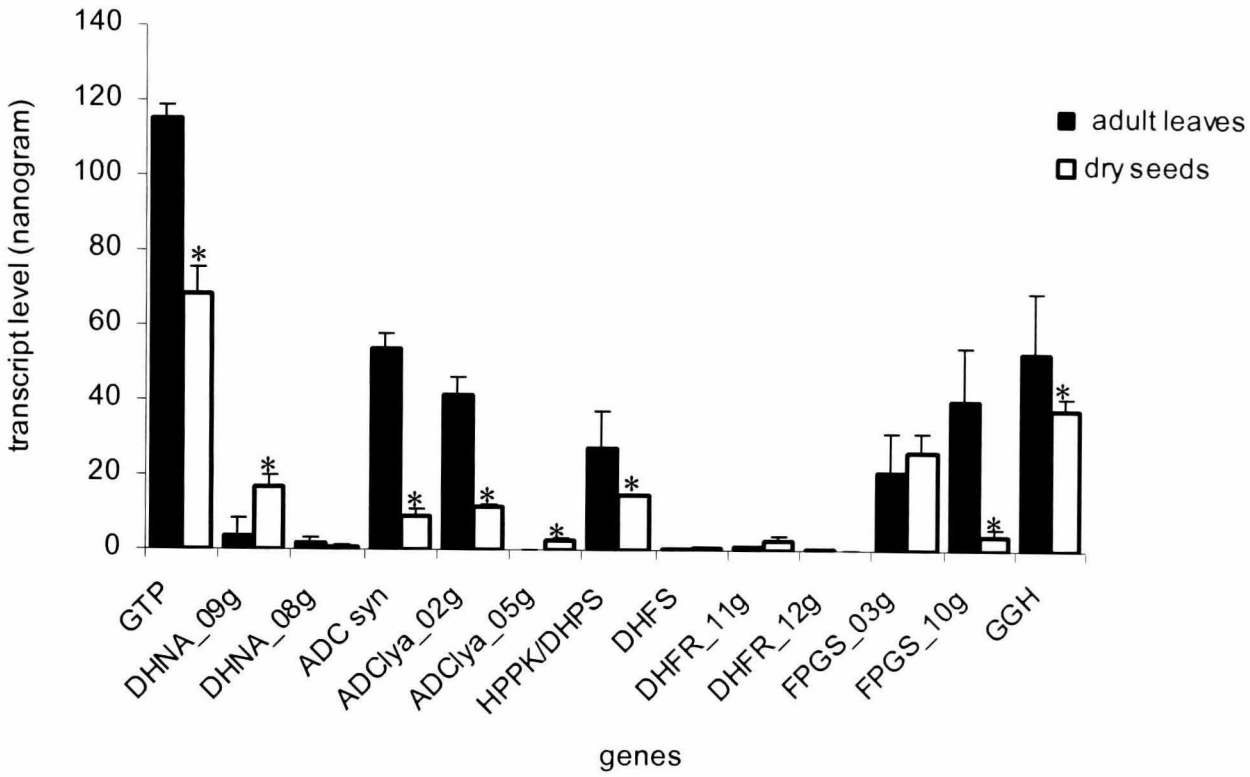


Figure 3.9 Expression profiles of folate-related genes in Nipponbare

Nipponbare rice leaves and whole dry seeds were RT-qPCR profiled. For seed samples, 40 dry mature seeds were pooled per analysis. For leaf samples, the third green leaf from 5 plants was pooled per analysis. The transcript values are provided in nanograms representing the amount of mRNA for a particular gene relative to the actin transcript. All folate-related genes including *GTPCHI* LOC_Os04g56710 (GTP), *DHNA* LOC_Os09g38759 (DHNA_09g), *DHNA* LOC_Os08g44210 (DHNA_08g), *ADC synthase* LOC_Os06g48620 (ADC syn), *ADC lyase* LOC_Os02g17330 (ADClya_02g), *ADC lyase* LOC_Os05g15530 (ADClya_05g), *HPPK/DHPS* LOC_Os07g42632 (HPPK/DHPS), *DHFS* LOC_Os12g42870 (DHFS), *DHFR* LOC_Os11g29390 (DHFR_11g), *DHFR* LOC_Os12g26060 (DHFR_12g), *FPGS* LOC_Os03g02030 (FPGS_03g), *FPGS* LOC_Os10g35940 (FPGS_10g) and *GGH* LOC_Os05g44130 (GGH) were shown in vertical bars. Each bar represents the mean of experiments performed in triplicate and the vertical line indicates their standard deviation. Asterisks represent the significant difference ($P<0.05$) of transcript abundance compared between seed and leaf.

In addition to leaves and grains, other rice tissues collected during rice development stages were also quantified. Figure 3.10 shows mRNA expression profiling from coleoptiles, radicles, adult leaves, adult roots, flowering stage, milk stage, dough stage and dry mature grains. This revealed that during rice development, the key genes of each synthesis branch were expressed significant differently. Almost all mRNA in leaves were more highly abundant than in other tissues. In addition, during seed filling stages mRNAs decreased and then increased again in mature grains.

qPCR results were also compared to microarray data obtained from the RiceGE database (Figure 3.11). Whilst the key genes *GTPCHI*, *ADC synthase* and *FPGS* genes were present on the array, undesirably, the *HPPK/DHPS* and *GGH* genes were not. To evaluate the differences between qPCR and microarray results, the data from rice tissues at similar time points were chosen. Root, 7-day-old seedling in microarray data was equivalent to radicles in qPCR. Mature leaf in microarray was expected to be the adult leaf in qPCR. Young leaf in microarray was assumed as coleoptiles in qPCR. Also young inflorescence in microarray was flowering in qPCR. Seed 5-10 dap in microarray was expected to be milk stage in qPCR, seed 11-20 dap was dough stage and seed 21-29 dap was the mature seed. An obvious point observed in the microarray experiment was that it did not report which rice cultivar have been used to obtain this data, stating only that microarray results were performed on *Oryza sativa* ssp. *Japonica*. Therefore, microarray data will be discussed as if it was obtained from Nipponbare as it is the most common used in rice experiments.

Analysing first the *GTPCHI* gene (Os04g56710), microarray data appeared to correlate with RT-qPCR profiling in leaf samples. The mature leaf contained the highest gene expression level, whereas expression in seeds from these two expression data sets was different. Microarray data showed *GTPCHI* expressed in mature seeds less than in milk and dough stages while RT-qPCR result showed the expression of this gene increased at the mature stage.

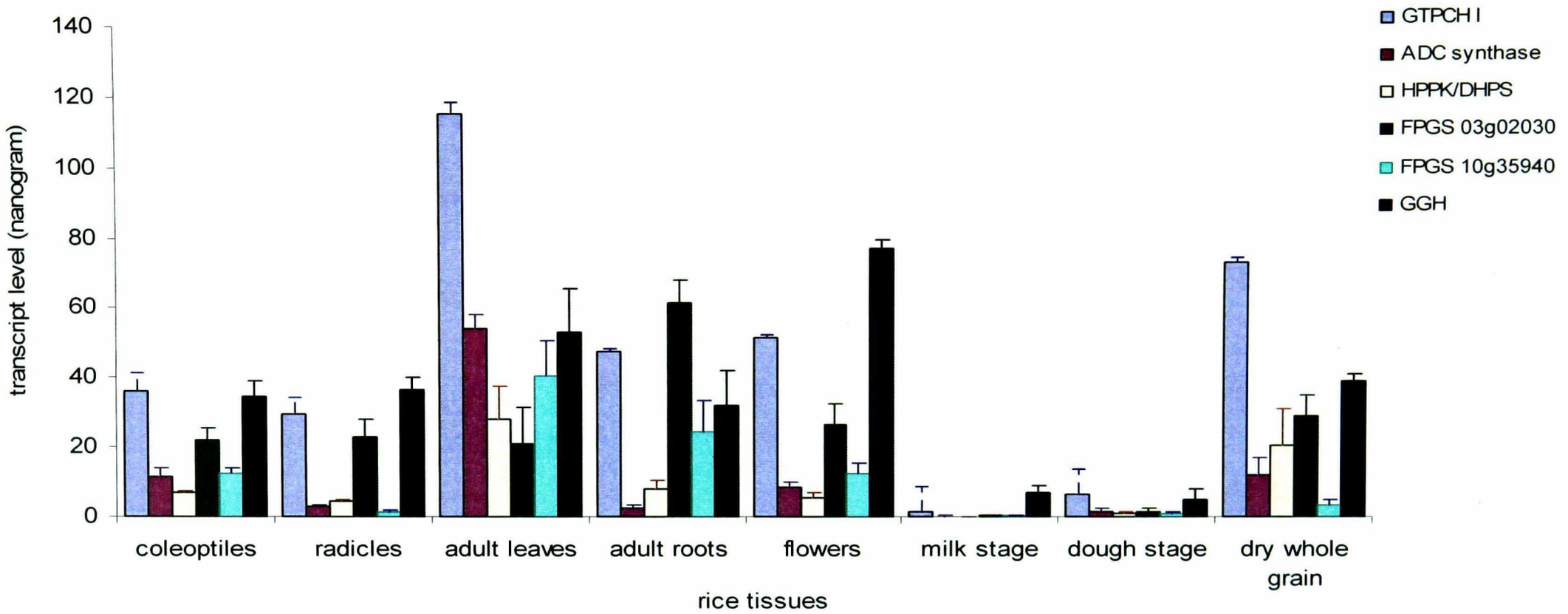


Figure 3.10 Expression profiles of folate-related genes of Nipponbare during rice development stages

Several tissues of Nipponbare rice samples were RT-qPCR profiled including coleoptiles, radicles, adult leaves, adult roots, flowers stage, milk stage, dough stage and dry grains. At least 100 mg of each tissue was collected from 10 different plants per pool per analysis. The expression values were provided in nanogram of gene expression level relying on the standard curve and were normalized by actin expression level. Only the key genes in the pathway were selected including *GTPCHI* LOC_Os04g56710 (■), *ADC synthase* LOC_Os06g48620 (■), *HPPK/DHPS* LOC_Os07g42632 (■), *FPGS* LOC_Os03g02030 (■), *FPGS* LOC_Os10g35940 (■) and *GGH* LOC_Os05g44130 (■) showed in different colour bars. Each bar represents the mean of triplicate determination and the vertical line indicates the standard deviation.

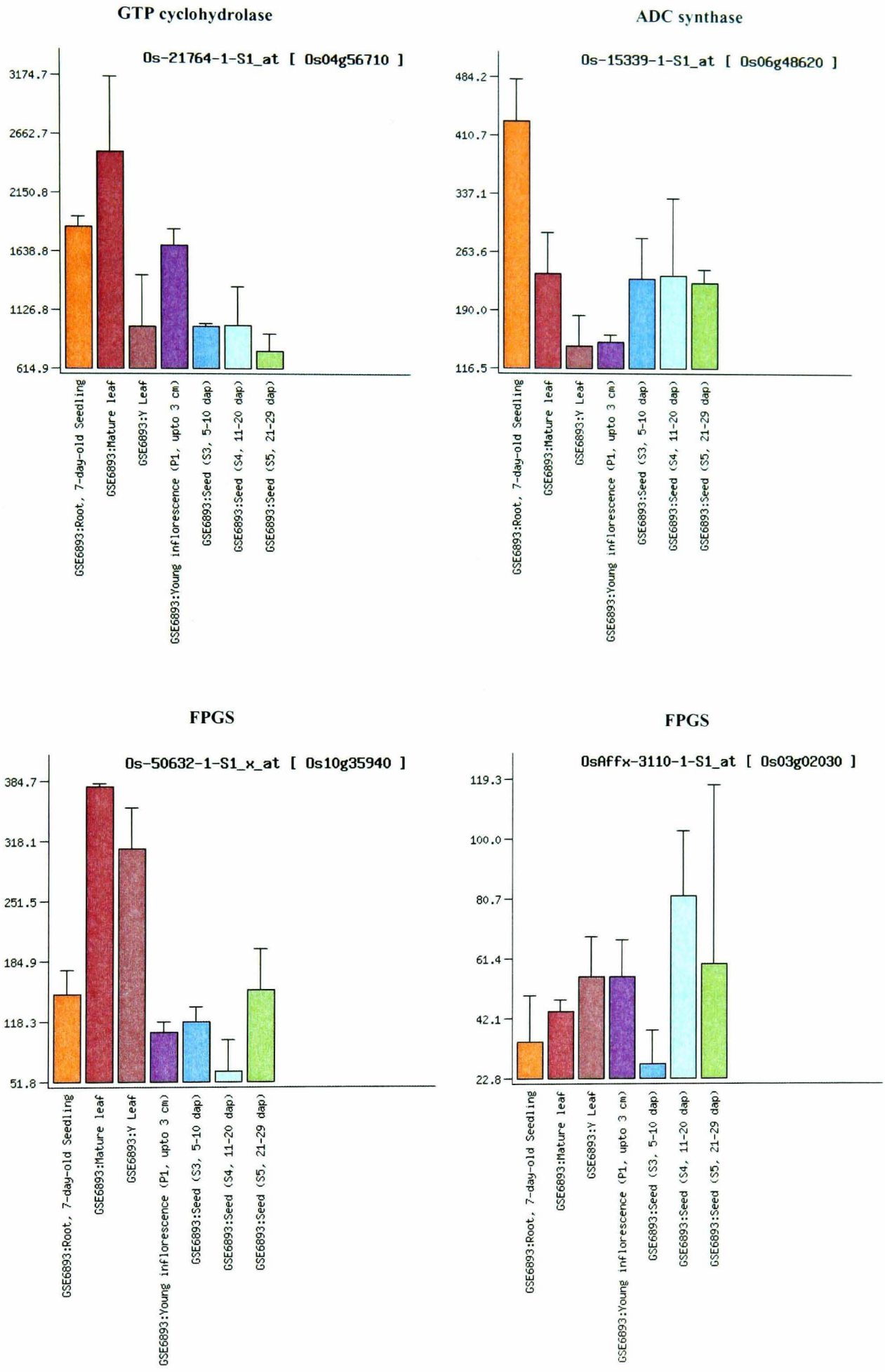


Figure 3.11 Microarray data of *GTPCHI*, *ADC synthase* and *FPGS* genes

(obtained from http://signal.salk.edu/cgi-bin/RiceGE2; 27 January 2009)

Many differences were observed between microarray and RT-qPCR data for the *ADC synthase* gene (Os06g48620). In the beginning of developing stages, microarray showed root seedling shared higher expression than young leaf whilst RT-qPCR presented radicles was less expressed than coleoptiles. Also the same trend happened in seeds. The expression levels between three seed developing stages were not obviously different from the microarray results but in RT-qPCR, it showed increasing mRNA abundance from milk stage to mature stage.

For *FPGS* genes, the expression levels detected by microarray and RT-qPCR for Os03g02030 and Os10g35940 were in agreement. *FPGS* Os03g02030 showed very high expression in seed development stages compared to leaves, whereas, *FPGS* Os10g35940 mRNA level were higher in leaves than in seeds. These results can support the earlier hypothesis that *FPGS* Os03g02030 might play the main role in seeds while *FPGS* Os10g35940 is more responsible in leaves.

3.3.3 Comparison of mRNA abundance for genes encoding folate biosynthesis enzymes in rice varieties

Next, the RT-qPCR assay was used to detect differences in transcript abundance amongst rice varieties. In parallel, 51 rice varieties were screened for total folate level using microbiological assay (MA) (see Figure 3.12) (by Riza Abilgos-Ramos, personal communication). Moroberekan, a West African rice variety, had 3-fold higher folate levels than IR72. MA also detected significant differences in total folate level between wild type varieties, Moroberekan, Nipponbare and IR72 of 63.62, 24.5 and 15.59 μg/100g grain fresh weight, respectively. Moreover, in collaboration with Dominique Van Der Straeten (Ghent University), Shan Yon 63 and Mungur varieties were reported to contain high levels of folate. Employing an LC-MS/MS assay (by Dominique Van Der Straeten, personal communication), they detected 30 and 32 μg/100g grain fresh weight in both varieties. Hence, a large variation of folate content between rice varieties was observed.

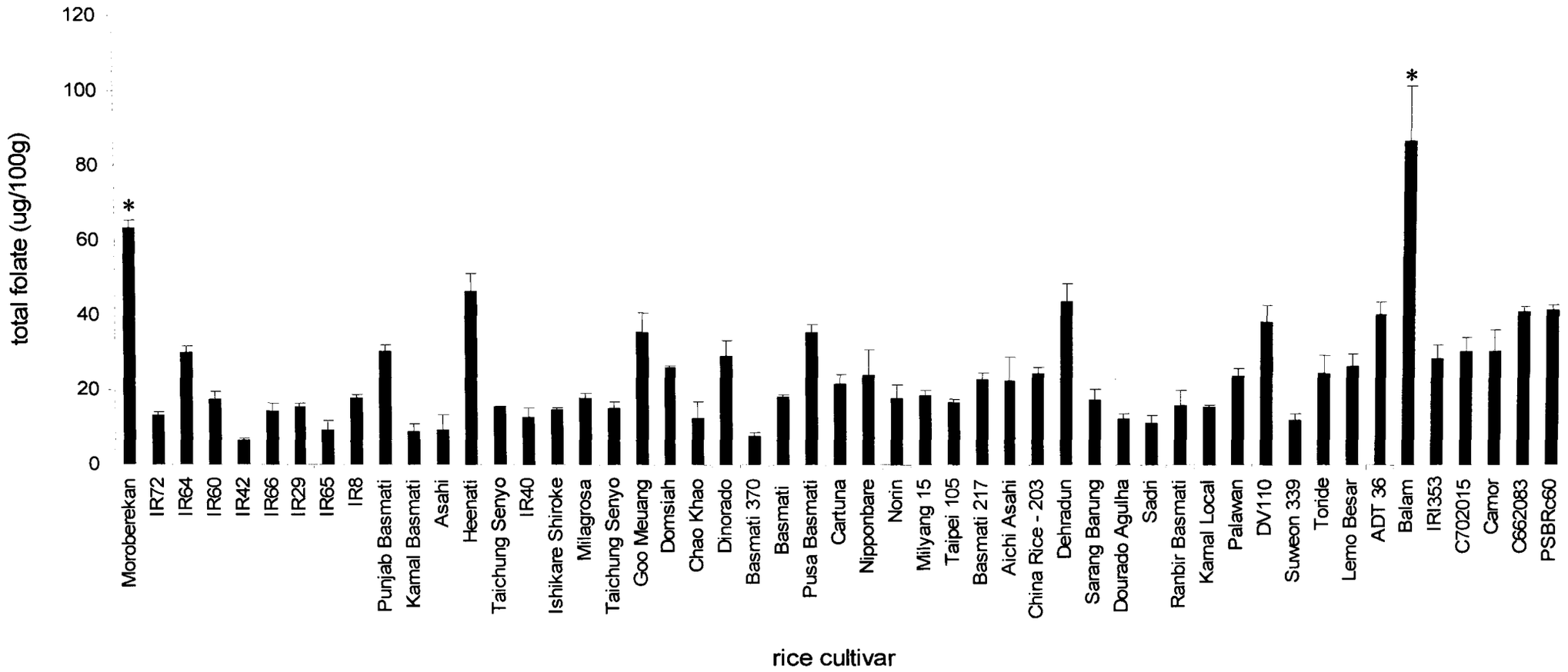


Figure 3.12 Total folate concentrations of unpolished grains of 51 rice cultivars

Mature rice grains of all varieties were obtained from IRRI. They were dehasked and all brown rice seeds were profiled using microbiological assay (MA). Total folate concentration was shown in microgram per hundred gram fresh weight. Asterisk symbols represent the highest folate level-rice varieties. This result was generated by Riza Albigos-Ramos.

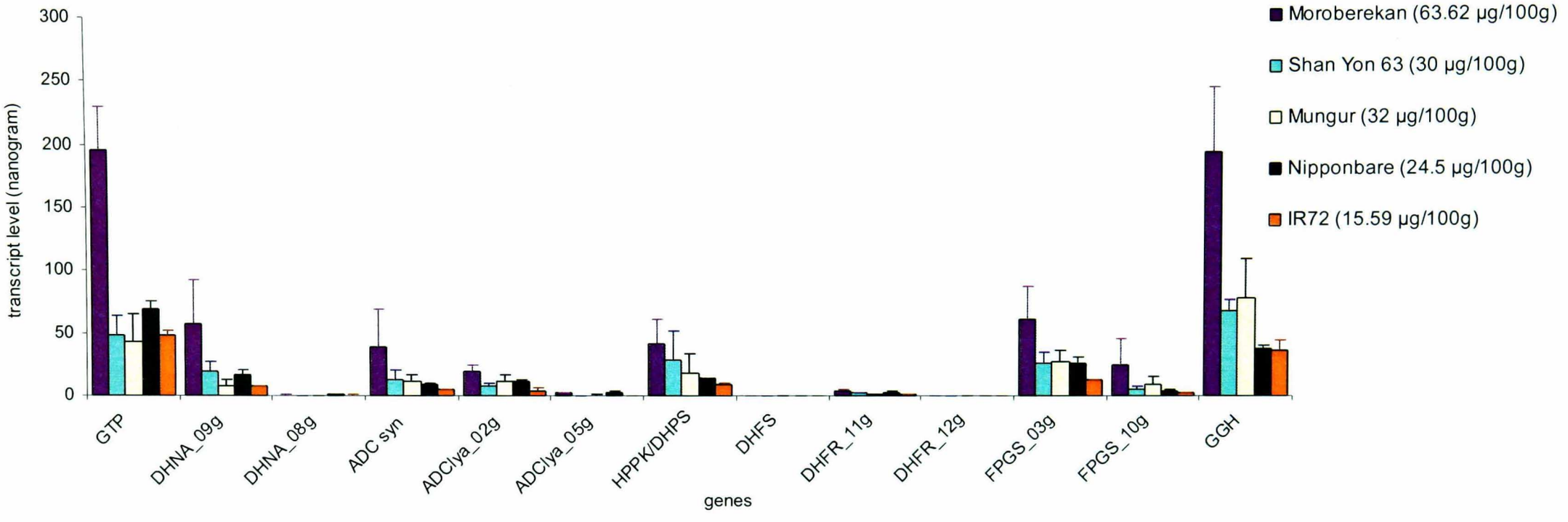


Figure 3.13 Comparison of the expression profile among wild type rice varieties

Mature dry grains of Moroberekan, ShanYon 63, Mungur, Nipponbare and IR72 rice varieties were profiled using RT-qPCR. At least 40 seeds of each variety were pool per analysis. The expression values were provided in nanogram of gene expression level relying on the standard curve and were normalized by actin expression level. All of genes involving in the pathway were presented. *GTPCHI* LOC_Os04g56710 (GTP), *DHNA* LOC_Os09g38759 (DHNA_09g), *DHNA* LOC_Os08g44210 (DHNA_08g), *ADC synthase* LOC_Os06g48620 (ADC syn), *ADC lyase* LOC_Os02g17330 (ADClya_02g), *ADC lyase* LOC_Os05g15530 (ADClya_05g), *HPPK/DHPS* LOC_Os07g42632 (HPPK/DHPS), *DHFS* LOC_Os12g42870 (DHFS), *DHFR* LOC_Os11g29390 (DHFR_11g), *DHFR* LOC_Os12g26060 (DHFR_12g), *FPGS* LOC_Os03g02030 (FPGS_03g), *FPGS* LOC_Os10g35940 (FPGS_10g) and *GGH* LOC_Os05g44130 (GGH) were shown in different colour bars between varieties. Each bar represents the mean of triplicate determination and the vertical line indicates the standard deviation. Total folate concentration was shown after a name of each rice variety in the graph.

To determine the molecular basis for the mRNA abundances, folate related transcripts were examined using RT-qPCR. Rice varieties were divided into 3 groups which were Moroberakan as the high-folate rice; Shan Yon 63, Mungur and Nipponbare as the medium-high folate rice; and IR72 as the low-folate rice. The transcript abundance values and patterns are shown in Figure 3.13.

In comparison between the 5 varieties, gene expression levels seemed to correlate with folate abundance. This is clearly seen for the mRNA levels of the first enzymes in the pteridine, *p*ABA and tetrahydrofolate biosynthesis pathways. Every mRNA was up-regulated several fold in the Moroberekan variety. The medium high folate varieties, Shan Yon 63 and Mungur also had higher mRNA abundance than the low-folate rice, IR72.

Apart from inspecting mRNA and folate levels, rice grain partition was also explored. Basically, there are three main parts in cereal grain which are the embryo (germ), maternal layer (bran and aleurone layers) and endosperm (white starchy part). The quantification of folate in three different sections of wheat seeds was first stated in McIntosh *et al.* (2008). They found folate 2.8 μg/g in embryo, 0.26 μg/g in endosperm and 5 μg/g in aleurone layer (Fenech *et al.*, 1999). This result showed the very low amount of folate in endosperm which is an interesting target for folate biofortification. The low folate level in endosperm correlates to very low amount of mRNA expression because mRNA from only white endosperm could not be extracted from the current study. Hence, transcript level presented in this study represents the sum of the mRNA abundance from embryo and aleurone layer. The size of embryo or aleurone layer may influence the different mRNA and folate levels. The bigger an embryo or thicker an aleurone layer, the higher mRNA and folate level that maybe detected. Figure 3.14 shows the comparison of the embryo between Moroberekan and IR72 (white arrows). Whilst the endosperm size differs, the embryo size appears similar. However, the folate level in endosperm is normally very low compared to embryo (in equal fresh weight). Hence, the differences of folate level among rice grains are dependent on the differences in gene regulation rather than physical size.

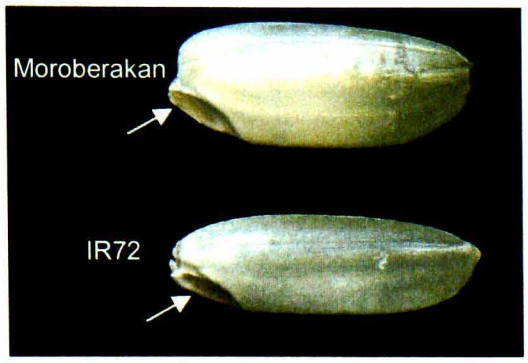


Figure 3.14 Grain phenotype of Moroberekan and IR72

In addition, RT-qPCR profiling between the aleurone layer and embryo did not show any significant difference for genes in the folate pathway (Figure 3.15). Hence, whole brown grain was used for all qPCR experiments.

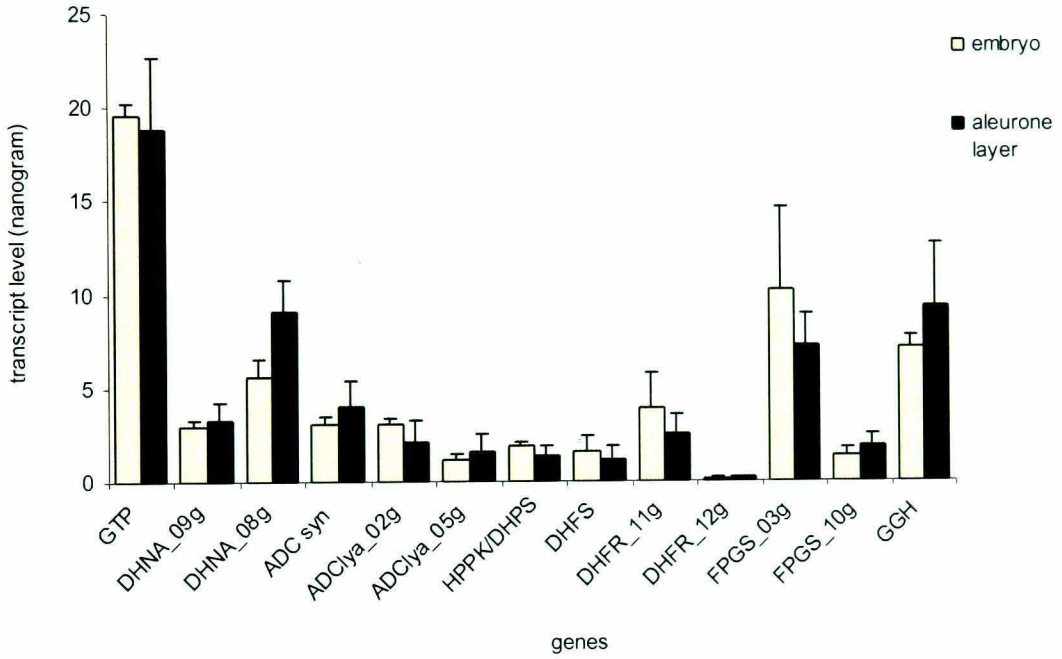


Figure 3.15 Comparison of RT-qPCR profiling of folate-related genes expressed in embryo and aleurone

IR72 embryo and endosperm were RT-qPCR profiled. The expression values were provided in nanogram of gene expression level relying on the standard curve and were normalized by actin expression level. All folate-related genes including *GTPCHI* LOC_Os04g56710 (GTP), *DHNA* LOC_Os09g38759 (DHNA_09g), *DHNA* LOC_Os08g44210 (DHNA_08g), *ADC synthase* LOC_Os06g48620 (ADC syn), *ADC lyase* LOC_Os02g17330 (ADClya_02g), *ADC lyase* LOC_Os05g15530 (ADClya_05g), *HPPK/DHPS* LOC_Os07g42632 (HPPK/DHPS), *DHFS* LOC_Os12g42870 (DHFS), *DHFR* LOC_Os11g29390 (DHFR_11g), *DHFR* LOC_Os12g26060 (DHFR_12g), *FPGS* LOC_Os03g02030 (FPGS_03g), *FPGS* LOC_Os10g35940 (FPGS_10g) and *GGH* LOC_Os05g44130 (GGH) were showed in vertical bars. Each bar represents the mean of triplicate determination and the vertical line indicates the standard deviation.

3.4 DISCUSSION

3.4.1 mRNA abundance represent *de novo* folate synthesis detected in all rice tissues

The first question raised was about the location of folate synthesis in plant tissues. Even though folate can be extracted from several plant tissues including seeds, it is not known either if *de novo* synthesis occurs in all these tissues. The presence of mRNA transcripts in coleoptiles, radicles, mature leaves, mature roots, flowers, milk seeds, dough seeds and dry mature grains indicates that folate synthesis machinery is active in all tissues profiled. The present study initially investigated the difference of transcript abundance of folate related genes in rice tissues. cDNA coding for 13 rice folate related enzymes was identified using genome databases based on their similarity to *Arabidopsis*. The relative amounts of transcripts from the 13 genes were determined in several rice tissues using quantitative PCR. The Japanese variety, Nipponbare, was selected as a model.

Even though folate synthesis gene expression appears active in most tissues, the degree of mRNA abundance varies between tissues. Since folate plays a crucial role in nucleic acid biosynthesis and DNA methylation during the cell cycle; the demand for folate increases during germination and then folate starts to be turned over and more produced at later growth phases (Gambonnet *et al.*, 2001). The demand for folate during rice development correlates with the transcript abundance of folate biosynthesis genes. For example, in Nipponbare folate related gene mRNA levels varies dramatically between leaves and grains (Figure 3.9). Transcripts of almost all genes in leaves are more abundant than in other tissues. This may be because the major portion of synthesized folate accumulates in leaves as folate is required for the photorespiratory cycle and chlorophyll and lignin biosynthesis (Hanson and Roje, 2001; Gambonnet *et al.*, 2001).

Although no data currently exists for folate content in rice tissues during developmental stages, total folate content was reported in other plants such as

pea seedling (*Pisum sativum*) which can be used as a reference to determine the association between transcript abundance and folate levels. Interestingly, folate content assayed by the microbiological method in pea leaves is 2-3 times higher than in other tissues (Gambonnet *et al.*, 2001). Furthermore, folate accumulation is higher in green leaf mitochondria than etiolated leaf mitochondria (Jabrin *et al.*, 2003) indicating that folate accumulation occurs mainly in leaves and reflects the light regime. Moreover, Jabrin *et al.* (2003) reported that high folate content in photosynthetic leaves of pea was associated with high levels of *HPPK-DHPS* mRNA and protein. Interestingly, RT-qPCR results from Nipponbare wild type rice revealed that the transcript level of *HPPK/DHPS* gene in leaves was higher than other genes in the middle branch of the monoglutamyl-tetrahydrofolate biosynthesis pathway. As there is a good correlation between the abundance of *HPPK/DHPS* mRNA and protein as well as folate concentration in pea leaves, the transcript level of folate-related genes in leaves could imply that the high mRNA abundance of folate genes is associated with a high requirement for folate compounds in rice leaf.

RT-qPCR profiling also helps pinpoint the key genes in the folate biosynthesis pathway. Genes encoding GTPCHI, ADC synthase, HPPK/DHPS, FPGS and GGH enzymes exhibit high mRNA abundances. Considering the pteridine synthesis enzymes, the gene encoding the GTPCHI enzyme is more abundant than both genes encoding the DHNA enzyme. This result can be seen in all rice tissues. GTPCHI converts GTP into dihydroneopterin triphosphate (DHNTP). This product is used for other synthesis pathways such as chromopores (Rébeillé *et al.*, 2006). Moreover, several pterin-containing molecules have been isolated from plant tissues. For example, 6-methylpterin, 6-carboxypterin and 6-hydroxymethyldihydropterin were found in pea chloroplasts corresponded as cofactors of the photophosphorylation pathway (Makarov and Stakhov, 1972). Erythro-neopterin, threo-neopterin and isoxanthopterin have also been extracted from soybean seeds (Kohashi, 1980). Due to this evidence, pterin synthesized in cells may not only flow into the folate pathway but are also involved in other pterin pathways. Hence, due to its importance, it is not surprisingly that *GTPCHI* trancript would be highly expressed in plant tissues.

GTPCHI is also a very inefficient enzyme indicated by its turnover number (k_{cat}) compared to other enzymes on the folate pathway (Table 3.4). However, kinetic studies have not yet been performed in rice enzymes. Hence, kinetic values taken from microorganism and animal systems might be different in plant systems such as rice.

Table 3.4 Turnover number of folate related enzymes

Enzymes	**EC code***	**k_{cat} (1/s)***	**substrate**	**organism**
GTPCHI	3.5.4.16	0.0035	GTP	*Thermus thermophilus*
DHNA	4.1.2.25	0.082	7,8-dihydroneopterin	*Escherichia coli*
ADC synthase	2.6.1.85	0.53	Chorismate with L-glutamate cosubstrate	*Escherichia coli*
ADC lyase	4.1.3.38	n/a		
HPPK/DHPS	2.5.1.15	n/a		
DHFS	6.3.2.12	n/a		
DHFR/TS	1.5.1.3	24.6	7,8-dihydrofolate	*Escherichia coli*
FPGS	6.3.2.17	0.65 0.82 0.55 0.34 0.01	pteroyl-Glu pteroyl-Glu2 pteroyl-Glu3 pteroyl-Glu4 pteroyl-Glu5	*Homo sapiens*
GGH	3.4.19.9	8.29 1.1 5.51 1.97	p-aminobenzoylpentaglutamate Pteroylpentaglutamate p-aminobenzoylpentaglutamate Pteroylpentaglutamate	*Arabidopsis thaliana* (GGH1) *Arabidopsis thaliana* (GGH2)

* EC number of all enzymes and the turnover number (k_{cat}) were obtained from Brenda database (www.brenda-enzymes.info/, March 2009); n/a represents no data available in the database.

In the *p*ABA synthesis branch, the transcript of *ADC synthase* gene is much higher than all of the *ADC lyase* genes. The transcript of *ADC synthase* gene was less in grains than in leaves. This might be one factor for having low folate content in grains. Hence, the transcript for first enzymes in pteridine and *p*ABA branches seem to be the most highly expressed genes implying that this could provide a way to engineer increased folate synthesis by metabolic engineering rice tissues.

In the middle branch of the folate synthesis pathway located in the mitochondria, the transcript of *HPPK/DHPS* gene was higher than other genes in the same branch suggesting that the HPPK/DHPS enzyme could be a regulatory point of this branch (Mouillon *et al.*, 2002). It is in agreement of RT-qPCR result between leaf and grain (Figure 3.9). Significant difference of transcript abundance of HPPK/DHPS between leaf and grain implies that this enzyme might determine the difference in tetrahydrofolate monoglutamate levels.

Interestingly, *FPGS* transcript abundance in leaves and grains were significantly different. In leaves, the *FPGS* Os03g02030 mRNA was lower than *FPGS* Os10g35940, whilst the reverse was observed in grains. *FPGS* genes are also differentially expressed in other tissues. Only in leaves, the *FPGS* Os10g35940 mRNA abundance is higher than the other FPGS. This data suggests that *FPGS* Os03g02030 gene plays a major role in polyglutamylation process in grains, whereas *FPGS* Os10g35940 is more important in leaves.

GGH mRNA abundance was at very high levels in almost all stages, most probably because this enzyme is needed to recycle photoinactivated folate. Breaking down and recycling folate appears to be an important process in plant cells in order to maintain folate homeostasis (Orsomando *et al.*, 2006). The high level of its mRNA reflects the essential function of GGH to plant cells.

RT-qPCR studies from flowering to dough stages of grain filling suggests that folate production occurs independently in grain and may not require folate transport from other tissues. Also, mRNA levels of all genes declined sharply after the flowering stage (Figure 3.10). This implied that during rice grain filling, folate levels may decrease. This is similar to previous reports which observed that *GTPCHI* mRNA level and 5-methylTHF declined sharply in tomato fruits during the ripening stage (Basset *et al.*, 2002). From the study of wheat seed development, McIntosh *et al.* (2008) reported total folate levels decreased by 30% post-anthesis until reaching the maturity stage. This result correlated with decreasing mRNA levels of THF producing genes during rice

grain development. Hence, the decrease in transcript levels in anthesis, milk and dough stages may be indicative of a drop in the amount of folate in seeds. During grain filling, it may lack of ability to maintain folate synthesis. However, in mature rice seed, there is an increase of mRNA transcripts which can either reflect the increasing folate synthesis during seed maturation or the accumulation of transcripts to prepare for the next germination. To investigate if mRNA abundance of folate synthesis genes is proportionate with folate levels in mature seeds, samples from a number of rice varieties have been explored.

3.4.2 Variation of folate biosynthesis gene expression correlates with variation of total folate level among rice varieties

Selected rice varieties, Moroberekan, Shan Yon 63, Mungur, IR72 and Nipponbare, were profiled using RT-qPCR in order to examine whether mRNA abundance of folate-related genes differs among rice varieties. From Figure 3.11, the expression profile of the high folate variety, Moroberekan, was compared to other lower folate containing varieties. It was revealed that every mRNA was up-regulated several fold in Moroberekan, consistent with high folate levels being due to high gene expression level of all genes in the pathway.

With reference to the polyglutamylation mechanism, the expression pattern of both *FPGS* genes in Moroberekan seeds was still the same as in other varieties (Figure 3.13). *FPGS* Os03g02030 gene has a higher expression level than *FPGS* Os10g35940 gene. This is a common feature of *FPGS* expression in rice grains which implied that *FPGS* Os03g02030 rather than *FPGS* Os10g35940 might have an important role in enzyme activity required in polyglutamylation in seeds. The different level of *FPGS* transcript among rice varieties was also observed. Higher *FPGS* transcript in Moroberekan might cause higher production of polyglutamylated folate forms than in other varieties. Additionally, *GGH* expression level was also high in Moroberekan seeds which may suggest higher folate turnover in this variety. Moreover, the big

differences of *FPGS* and *GGH* transcript level in IR72 and Moroberekan rice might affect the ratio of monoglutamate folate and polyglutamate folate in seeds, thus, the measurement of the amount of polyglutamylated folate is required to determine this point of view.

LC-MS/MS profiling (by Riza Albigos-Ramos, personal communication) revealed that the most common monoglutamate folate form was 5-methyltetrahydrofolate detected in Moroberakan, IR72 and Nipponbare (see Table 3.5). Total folate levels of all three varieties showed the same trend as that obtained from MA assay which Moroberekan presenting very high amounts of total folates. In terms of polyglutamylated folate forms, different degrees of glutamylation were found. Moroberekan and IR72 contained mostly tetra- and pentaglutamylated folates while only Moroberekan contained the formyl form of pentaglutamylated folate. Nipponbare contained only tetraglutamylated folate.

Table 3.5 Folate contents in three rice varieties obtained by LC-MS/MS

Rice Variety (n=3)	**Folate Form (µg/100g)**					
	$5CH_3$ $H_4PteGlu$	5CHO $H_4PteGlu$	$5CH_3$ $H_4PteGlu_4$	$5CH_3$ $H_4PteGlu_5$	5CHO $H_4PteGlu_5$	Total
Moroberekan	43.2 ± 1.5*	1.99+0.03*	2.87± 1.3	1.64± 0.51	0.32± 0.55	50.02*
IR72	22.3 ± 2.0*	1.3 ± 0.3	3.51 ± 1.2*	1.22 ± 0.02	nd	32.22*
Nipponbare	14.56 ± 0.13*	1.55±0.01*	0.5 ± 0.03*	nd	nd	16.61*

$5CH_3H_4PteGlu$ = 5-methyltetrahydropteroylmonoglutamate;
$5CHOH_4PteGlu$ = 5-formyltetrahydropteroylmonoglutamate;
$5CH_3H_4PteGlu_4$ = 5-methyltetrahydropteroyltetraglutamate;
$5CH_3H_4PteGlu_5$ = 5-methyltetrahydropteroylpentaglutamate;
$5CHOH_4PteGlu_5$ = 5-formyltetrahydropteroylpentaglutamate
The mean±SD was presented with n = 3
* - samples between varieties which are significantly different ($P<0.05$)
nd = not detected

RT-qPCR was compared to polyglutamylated folate concentration. The amount of *FPGS* and *GGH* transcripts may affect the ratio of monoglutamate and polyglutamate folate forms. Due to very high transcript abundance in *GGH* in Moroberekan, folate turnover may occur at a high level resulting in the high

concentration of monoglutamate folates. The *GGH* transcript differences were also observed between Moroberekan and IR72. *GGH* transcript in IR72 appeared 4 fold lower than that in Moroberekan (Figure 3.13). Hence, it can be the result of having lower monoglutamate folates in IR72 than in Moroberekan. However, IR72 contained almost an equal amount of polyglutamylated folate forms as Moroberekan.

In case of Nipponbare, only tetraglutamylated folate form was detected and at the low level. The amount of polyglutamylated folates was not correlated with the mRNA level of *FPGS* and *GGH* as these two mRNA levels presented in Nipponbare was not much higher than in IR72. From our hypothesis if *FPGS* and *GGH* transcripts affecting the polyglutamylation level directly, the amount of polyglutamates in Nipponbare should not have been much different from that in IR72. Hence, the variation of polyglutamylated folate production among rice varieties may be co-regulated by other factors, not only the transcript abundance of FPGS and GGH enzymes.

In summary, by comparing the mRNA and folate levels, the relative levels of mRNA of rice varieties roughly match their total folate contents especially the amount of 5-methyltetrahydrofolate monoglutamate. Hence, the differences of gene expression between rice cultivars could explain the variation of total folate abundance in rice grains. The amount of monoglutamate folates relies on the transcript levels of folate biosynthesis genes and *GGH* in the turnover pathway.

3.4.3 The genes encoding the first enzyme in each pathway are the key genes for metabolic engineering

In the context of metabolic engineering, folate-related gene expression profile in rice was compared to previous transgene studies. From Hanson and his colleagues, folate enhancement by crossing tomato plants overexpressing synthetic *GTPCHI* gene based on mammalian gene (Díaz de la Garza *et al.*, 2004) and *ADC synthase* gene from *Arabidopsis* (Díaz de la Garza *et al.*, 2007) were studied in fruits. The transgenic tomatoes showed higher total folate

levels than in wild type tomatoes with about 2-fold obtained from *GTPCHI* gene overexpression and 25 fold obtained from combined *GTPCHI* and *ADC synthase* gene overexpression (Díaz de la Garza *et al.*, 2007). Their results showed that the ADC synthase enzyme is a key target for enhancing folate in tomatoes.

RT-qPCR profiling has revealed that in contrast to *GTPCHI,* the *ADC synthase* mRNA was significantly lower in rice grains compared to rice leaves (Figure 3.9). Consequently, *ADC synthase* expression in grains, whose abundance is about 2-fold less than in leaves, could be one reason for having very low folate content in rice grains. ADC synthase enzyme therefore represents a target with which to improve rice quality. Recently, folate enhancement was reported in rice by overexpression of *Arabidopsis GTPCHI* and *ADC synthase* genes in rice endosperm (Storozhenko *et al.*, 2007). The maximal enhancement was as high as 100 times more than wild type rice. This success demonstrates that the seed has the ability to synthesize folate and is in good correlation with RT-qPCR results of *ADC synthase* gene expression level as discussed above. Though enhancing folate synthesis in rice endosperm has been successful, there are still several other possible strategies to further improve rice quality. For example, enhancing expression of other folate related genes or folate binding proteins which will be discussed in chapter 5.

CHAPTER 4 CHARACTERISATION OF FOLYLPOLYGLUTAMATE SYNTHETHASE GENE FUNCTION IN RICE

4.1 INTRODUCTION

4.1.1 Biochemical function of folylpolyglutamate synthethase enzyme

Folylpolyglutamate synthethase (FPGS, EC 6.3.2.17), an enzyme identified in different organisms including various plant species (Chan *et al.* 1986; Ravanel *et al.* 2001), is responsible for glutamylation of folate monoglutamate to create polyglutamylated folate forms. Glutamate residues are sequentially conjugated with monoglutamate folate via γ-carboxyl peptide linkages (Suh *et al.*, 2001) as the following reaction:

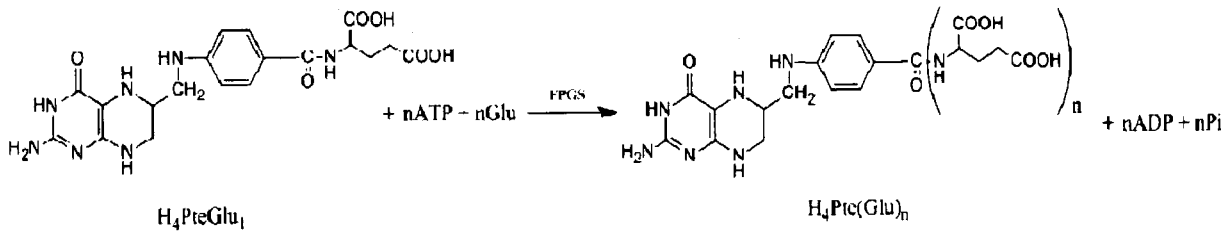


Figure 4.1 The biochemical reaction catalysed by the FPGS enzyme

4.1.2 Functional motifs of eukaryotic FPGS enzyme

In bacteria such as *E.coli* and *Plasmodium falciparum*, FPGS enzyme is a bifunctional protein with DHFS activity which has only one catalytic site for both activities. In contrast, FPGS in fungi, animals and plants is a monomeric protein (Bognar *et al.*, 1985; Cossins and Chen, 1997). In fungi and animals, FPGS contains two isoforms which are the cytoplasmic and mitochondrial

forms derived from the same gene (Chen *et al.*, 1996). The first AUG codon produces the mitochondria isoforms while the second initiation codon provides the cytosolic form lacking a transit peptide (residue 1-42) (DeSouza *et al.*, 2000; Freemantle *et al.*, 1995). In plants, the study of FPGS in *Arabidopsis* reported the presence of three FPGS genes encoding distinct mitochondrial, cytoplasmic and plastidial isoforms of this enzyme (Ravanel *et al.*, 2001).

In terms of enzyme properties, animal and plant FPGS enzymes share common biochemical features such as MgATP-dependence (Sun *et al.*, 2001), an alkaline pH optimum for their activities and a broad specificity for folate substrates (Cossins and Chen, 1997). The major functional domains have been explored.

The structural analysis of eukaryotic FPGS was analysed in *Lactobacillus casei* and used as a model to reveal several conserved domains (Sun *et al.*, 1998). In the glutamylation reaction, FPGS enzyme needs three substrates which are ATP, THF and glutamate molecules. Hence, eukaryotic FPGS consists of three major domains which are the N-terminal ATPase domain, the Ω–loop and the C-terminal glutamate binding site domain (Sun *et al.*, 2001). Figure 4.2 shows the amino acid sequence alignment of FPGS from yeast, human and *Arabidopsis* and indicates their major binding sites.

The ATP-binding site is in a deep cleft between the N-domain (residues 1-294 in *L. casei*) and the C-domain (residue 301-428 in *L. casei*). This site is surrounded by the P-loop (glycine rich, phosphate-binding loop; residues 45-49 in *L. casei*), the Ω–loop (residues 72-82 in *L. casei*) and the interdomain linker (residues 295-300 in *L. casei*).

The P-loop is quite conserved in other nucleotide binding enzyme like FPGS as the putative sequence of P-loop is similar in the family of nucleotide-binding enzymes. xxGxxGKGS sequence was reported in *L. casei* (Saraste *et al.*, 1990; Sun *et al.*, 1998).

```
Arabidopsis_cyto        ---------------------------------------------------------------------------MATEDDGELSARYQNTLDALSSLITKRGRLASNNQSHRFRLLFHY 45
Arabidopsis_plastid     ------MFAVSIVP--------RTTSCRLSSAFLCQLSIPLTLRLHHH----YQHHQPHLPSPLSFQIHSLRKQIDMAAQGGDSYEEALAALSSLITKRSRADKS---NKGDRFELVFDY 99
Arabidopsis_mito        MLVCGKGFLKCRAPGVPFFCDKRKSFFTKTKRGFHSLPLGTGVRVYFNNNLRYSSNSIEVVEKAAINMGSKEDKADNPALS--SYDDAMEALSTLISRRNRGDRTPTKGNRDKLEQVVTY 118
human_mito              ----------------------------------MSRARSHLRAALFLAAASARGITTQVAARRGLSAWPVPQEPSMEYQDAVRMLNTLQTNAGYLEQVKRQRGD---PQTQLEAMELYL 83
L.casei_cyto            ----------------------------------------------------------------------------MNYTETVAYIHSFPRLAKTGD--------------HRRILTLLH 26

                                       P-loop                    Ω–loop                                           implicated in enzyme activity
Arabidopsis_cyto        LKVLELEDAVSQMKIIHVAGTKGKGSTCTFAESILRCYGLRTGLFTSPHLIDVRERFRLNGIEISQEKFVNYFWCCFHKLKEKT-----------------------SNEVDVVILEVGL 142
Arabidopsis_plastid     LKLLDLEEDILKMNVIHVAGTKGKGSTCTFTESIIRNYGFRTGLFTSPHLIDVRERFRLDGVDISEEKFLGYFWWCYNRLKERTN-EEIPMPTYFRFLALLAFKIFAAEEVDAAILEVGL 218
Arabidopsis_mito        LKILDLEDKIKELKVIHVAGTKGKGSTCVFSEAILRNCGFRTGMFTSPHLIDVRERFRIDGLDISEEKFLQYFWECWKLLKEKAV-DGLTMPPLFQFLTVLAFKIFVCEKVDVAVIEVGL 237
human_mito              ARSGLQVEDLDRLNIIHVTGTKGKGSTCAFTECILRSYGLKTGFFSSPHLVQVRERIRINGQPISPELFTKYFWRLYHRLEETKDGSCVSMPPYFRFLTLMAFHVFLQEKVDLAVVEVGI 203
L.casei_cyto            ALGNPQQQG----RYIHVTGTNGKGSAANAIAHVLEASGLTVGLYTSPFIMRFNERIMIDHEPIPDAALVNAVAFVRAALERLQQQQADFNVTEFEFITALGYWYFRQRQVDVAVIEVGI 146

Arabidopsis_cyto        GGRFDATNVIQKPVVCGISSLGYDHMEILGYTLAEIAAEKAGIFKSGVPAFTVAQPDEAMRVLNEKASKLEVNLQVVEPLDS------SQRLGLQGEHQ-YLNAGLAVALCSTFLKEIG- 254
Arabidopsis_plastid     GGKFDATNAVQKPVVCGISSLGYDHMEILGDTLGKIAGEKAGIFKLGVPAFTVPQPDEAMRVLEEKASETEVNLEVVQPLTARLLS--GQKLGLDGEHQ-YVNAGLAVSLASIWLQQIG- 334
Arabidopsis_mito        GGKLDSTNVIQKPVVCGIASLGMDHMDILGNTLADIAFHKAGIFKPQIPAFTVPQLSEAMDVLQKTANNLEVPLEVVAPLEPKKLD--GVTLGLSGDHQ-LVNAGLAVSLSRCWLQRTGN 354
human_mito              GGAYDCTNIIRKPVVCGVSSLGIDHTSLLGDTVEKIAWQKGGIFKQGVPAFTVLQPEGPLAVLRDRAQQISCPLYLCPMLEALEEGGPPLTLGLEGEHQ-RSNAALALQLAHCWLQRQDR 322
L.casei_cyto            GGDTDSTNVIT-PVVSVLTEVALDHQKLLGHTITAIAKHKAGIIKRGIPVVTGNLVPDAAAVVAAKVATTGSQWLRFDRDFS------VPKAKLHGWGQRFTYEDQDGRISDLEVPLVGD 259
                                                                                  glutamate binding
Arabidopsis_cyto        -IEDKNGLDQTNGLPEKFISGLSNAYLMGRAMIVPD----SELPE--------EIVYYLDGAHSPESMEACAIWFSKQIK-QNQ-----------------ERNQKRSEQILLFNCMSVR 343
Arabidopsis_plastid     -KLEVPSRTQMSILPEKFIKGLATASLQGRAQVVPDQYTESRTSG--------DLVFYLDGAHSPESMEACAKWFSVAVKGDNQSGSSGHLVNGSAGSSHDKWSNETCEQILLFNCMSVR 445
Arabidopsis_mito        WKKIFPNESKETEIPVAFCRGLATARLHGRAQVVHDVVSDPQDSSDSMETPCGDLIFYLDGAHSPESMEACGRWFSSAVRGDKSLSTAVN--GYMRHGEYGTDLNRVSKQILLFNCMEVR 472
human_mito              HGAGEPKASRPGLLWQLPLAPVFQPTSHMRLGLRNTEWPGRTQVLRR-----GPLTWYLDGAHTASSAQACVRWFRQALQGRER------------------PSGGPEVRVLLFNATGDR 419
L.casei_cyto            YQQRNMAIAIQTAKVYAKQTEWPLTPQNIRQGLAASHWPARLEKISD------TPLIVIDGAHNPDGINGLITALKQLFSQPIT---------------------------VIAGILADK 346

Arabidopsis_cyto        DPSLLLPRLRSKCIDQGVDFKRAVFVPNVSVYNQVGS-STNVGTR--VESMSWQFGLQRIWESLARGE-----------------------AKSNSKSDSKGKEEEKSFVFSSLPVAVDW 437
Arabidopsis_plastid     DPNLLLPHLKNMCAKYGVNFKKALFVPNMSVYHKVGT-AADLPENDPQVDLSWQFTLQKVWESLVQSE-----------------------RDGEKDGESDGNSE----VFTSLPMAIKC 537
Arabidopsis_mito        DPQVLLPKLVTTCASSGTHFSRALFVPSMSTYNKVISGASAIPSDTRRKDLTWQFRLQRLWEKSIQGTDAGLDHTLKPDGITALPPHDFLCGDAPQCGGPAGTPVTSSAVMPSLPLTINW 592
human_mito              DPAALLKLLQP------CQFDYAVFCPNLTEVSSTGNADQQNFTVTLDQVLLRCLEHQQHWNHLDEEQ-------ASPDLWSAPSPEPGGSASLLLAPHPPHTCSASSLVFSCISHALQW 526
L.casei_cyto            DYAAMADRLTA-------AFSTVYLVPVPGTPRALPEAGYEALHEG---------RLKDSWQEALAAS------------------------LNDVPDQP-------------------- 406

Arabidopsis_cyto        LR---DNARQSK---------------------QVRFQVLVTGSLHLVGDLLRFIKK---- 470
Arabidopsis_plastid     LR---DTVHESSS--------------------ATRFQVLVTGSLHLVGDVLRLIRK---- 571
Arabidopsis_mito        LR---DCVRRNP---------------------SLKLEVLVTGSLHLVGDVLRLLKR---- 625
human_mito              ISQGRDPIFQPPSPPKGLLTHPVAHSGASILREAAAIHVLVTGSLHLVGGVLKLLEPALSQ 587
L.casei_cyto            --------------------------------------IVITGSLYLASAVRQTLLGGKS- 428
```

Figure 4.2 FPGS amino acid sequence alignments

FPGS protein sequences obtained from UniProtKB database for human (accession no Q05932), *L. casei* (accession no P15925) and *Arabidopsis* (accession no UPI00005DC204 for mitochondrial form, UPI000016371B for cytosolic form and UPI0000196BE1 for plastidial form). Alignment was done under CLUSTALW 2.0.10 version in www.ebi.ac.uk. The conserved regions are shown in shaded gray. P-loop involves much in ATP-binding site and Ω–loop is a major part of THF binding cleft. Other two conserved regions are implicated in enzyme activity and glutamate binding site.

The folate binding site is located in the interdomain cleft where residues from both N- and C-domains are surrounded. The pteridine ring is sandwiched by the Ω–loop and the C-domain while the glutamate binding site is in the C-domain (residues His316 and Ser412 in *L. casei*).

From the analysis of a crystal structure of two substrate-bound forms of FPGS, the conformation of the FPGS protein is changed following the attachment of different substrates (Sun *et al.*, 2001). FPGS binds first with THF and this induces a conformational change which allows binding to ATP and to form the binding site for glutamate residues. THF binding is provided by the Ω–loop which is conserved among species (Figure 4.3). This Ω–loop shows high similarity in most organisms and is suggested to be involved in the same important role in THF binding in all organisms (Sun *et al.*, 2001).

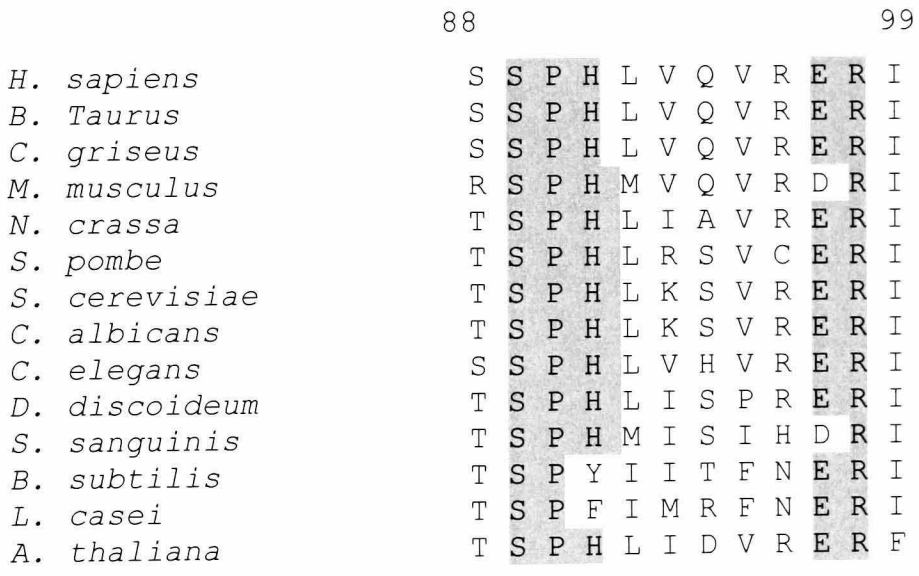


Figure 4.3 The Ω–loop alignment from 14 species

The conserved regions are highlighted in gray (modified from Sun *et al.*, 2001).

4.1.3 Polyglutamylated folate in eukaryotes

The di- and tri-compartmentation of FPGS enzymes in mammalian and *Arabidopsis* systems, respectively, were reported as the key to the homeostasis of folate and folate-dependent metabolites (Suh *et al.*, 2001; Ravanel *et al.* 2001). Since a single glutamate folate molecule is an inefficient coenzyme and

can efflux from cells, cellular polyglutamylated folate forms are required and is indeed the predominant form in yeast, mammalian and plant cells (Cossins and Chen, 1997). As glutamate residues in tetrahydrofolate (THF) make molecules more anionic, this increases solubility and reduces diffusion through hydrophobic membranes (Appling *et al.*, 1991). Polyglutamylated folate is compartmentalized with half in the mitochondria and the remainder in the cytoplasm and the chloroplast (Lin *et al.*, 1993).

The degree of glutamylation varies between developmental organs and with a preference of the target enzymes (Lowe *et al.* 1993). Penta- and hexaglutamates are major forms in mammalian cells; whereas tetra- and pentaglutamates are the main forms in pea leaves (Imeson *et al.*, 1990) and diglutamate is predominant in carrot root (Cossins and Chen, 1997). A wide range of polyglutamated forms was reported among plant species and during different developmental stages. Table 4.1 summarized the variation of glutamate chain length between higher plant species. Tetra- and pentapolyglutamylated folate forms are the major forms and account for the total folate pool in lettuce, soybean and pea.

Table 4.1 The distribution of glutamate chains in crop plants

Tissue	Distribution of glutamyl chains (%)				
	1	**2**	**3**	**4**	**5**
romaine lettuce leaves[2]	33	nd	nd	9	44
soybeans[1]	52	16	nd	nd	32
pea[3] (4 day cotyledons)	12.2	6.8	20.3	6.7	54
pea[3] (14 day leaves)	nd	46.5	<0.1	15.2	38.2

nd, non-specific identified chromatogram;
[1]Shin *et al.* (1975), [2]Batra *et al.* (1977) and [3]Imeson *et al.* (1990)

In eukaryotes, folate-dependant enzymes exhibit different preference for polyglutamylated folate affecting velocity and specificity to their reactions. For example, SHMT and GDC enzymes are two enzymes presented predominantly in plant leaf mitochondria involved in the photorespiratory cycle (Bourguignon

et al., 1988). The kinetic studies of SHMT revealed a higher affinity coefficient (Km) for hexaglutamate than for monoglutamate; being approximately 37 μM for monoglutamate and less than 3.7 μM for hexaglutamate (Besson *et al.*, 1993). Similar in GDC, the affinity coefficient (Km) value for tetrahydrofolate monoglutamate was 17 μM while for tetra- and pentaglutamates were about 0.5 μM. In addition, the minimum number of glutamate residues required in purine and thymidylate synthesis in Chinese hamster ovary (CHO) cell mutant is three (Osborne *et al.*, 1993).

In organisms which synthesize polyglutamylated folate *de novo*, the impairment of FPGS activity results in a defect in the one-carbon metabolism affecting cell growth and development. The study of cell lines bearing a mutation on either cytoplasmic or mitochondrial or plastidial FPGS activities shows the importance of the polyglutamylation process in yeast, mammalian and plant systems.

In yeast, *Saccharomyces cerevisiae*, the disruption on *MET7*, a FPGS encoding gene on chromosome XV, displayed a methionine auxotroph phenotype. Increase rate of loss of mitochondrial DNA and promotion of cell lethality was observed (DeSouza *et al.*, 2000). Lack of mitochondrial FPGS impaired the mitochondrial protein production (Cherest *et al.*, 2000). However, loss of FPGS function does not defect the synthesis of formate, serine or glycine in mitochondria (DeSouza *et al.*, 2000).

Chinese hamster ovary (CHO) cells lacking mitochondria FPGS activity exhibited glycine auxotrophy (Lin and Shane, 1994) indicating mammalian mitochondrial FPGS is essential for glycine synthesis and for maintaining folate pools in the mitochondria.

The physiological defect due to the impairment of folate-dependent enzymes and metabolites involving in one-carbon metabolism in plant systems have been reported (Gallardo *et al.*, 2002; Heineke *et al.*, 2001; Wingler *et al.*, 1997; Collakova *et al.*, 2008). However, very few studies on the physiological effect of impairment of FPGS enzyme activity have been reported. Recently, the

functional importance of polyglutamylated folate to plant development was examined in *Arabidopsis* mutants (Payam Mehrshahi, PhD thesis 2008, University of Nottingham). The knock out mutants in all three FPGS genes exhibited defective proliferation rate, dwarfism and a significant reduction of methionine level. A double knock out targeting the *Arabidopsis* mitochondria and plastidic *FPGS* genes caused embryo lethality, whilst lack of mitochondrial and cytoplasmic *FPGS* resulted in seedling lethality. These defects in *Arabidopsis* mutants involved the reduction of polyglutamylated folate pools in all three compartments.

4.2 AIMS AND OBJECTIVES

Although the physiological importance of polyglutamylated folates in all studied organisms have been shown, the role of FPGS in rice has received little attention. The comparative genomics study revealed two rice *FPGS* genes namely *FPGS* Os03g02030 and *FPGS* Os10g35940; unlike *Arabidopsis* FPGS genes encoding three distinct isoforms in three subcellular compartments. In this chapter, *Arabidopsis*, wheat and rice FPGS deduced amino acid sequences were initially aligned to compare the conserved functional motifs among plant species.

In chapter 3, expression analysis revealed that the mRNA level of *FPGS Os10g35940* was more abundant than *FPGS Os03g02030* in leaf tissue, whilst *FPGS Os03g02030* transcripts were higher in all other tissues including grains. RT-qPCR results suggested that these two *FPGS* genes in rice might be responsible for polyglutamylation during different stages of development.

To identify the biological functions of genes, the most direct way is to disrupt gene function and observe the changes in physiological and morphological traits. Postech (Plant Functional Genomics Laboratory, Korea) have generated a large pool of rice T-DNA inactivation mutants. Thus, the aim of this chapter

was to elucidate the function of rice *FPGS* genes by phenotypic analysis on the impact of knocking out *FPGS* genes by T-DNA.

T-DNA insertion lines FST number A16772 which disrupts the *FPGS Os03g02030* gene in rice *Oryza sativa* cv. Dongjin background and FST number D02773 which disrupts *FPGS* Os10g35940 gene in rice *Oryza sativa* cv. Hwayoung background were identified in collaboration with Emmanuel Guiderdoni (French Rice Functional Genomics Centre, Montpellier). Southern and PCR analysis were performed in order to characterise the mutant genotypes. Phenotypic changes of the mutant line were observed and mRNA abundance was measured by RT-qPCR in order to determine the effect in the mRNA abundance of other folate-related genes.

4.3 RESULTS

4.3.1 Identification of functional domains within rice FPGS enzymes

In *Arabidopsis*, three distinct genes encoding three distinct isoforms were reported to be compartmentalised in the mitochondria, the chloroplast and cytosol compartments (Ravanel *et al.*, 2001). In cereals, a wheat FPGS amino acid sequence has been reported to target in the mitochondria (McIntosh and Henry, 2008). *In silico* analysis of the Gramene rice genome database revealed two genes encoding two rice FPGS sequences. The alignment of amino acid sequences (Figure 4.4) revealed high homology between *Arabidopsis*, wheat and rice FPGS enzymes. FPGS sequences in these three species shared protein similarity up to 60%. The P-loop (VAGTKGKGS), Ω–loop (SPHLxDVRER), glutamate binding site (DGAH) and C-terminal domain (VLVTGSLHLVGD) were conserved across monocot and dicot plant species. Only the motif implicated in enzyme activity (EVGLGG) was not conserved. The rice *FPGS* gene, LOC_Os10g35940, lacked this region which may imply that the enzyme activity of this rice protein may differ from other FPGS enzymes.

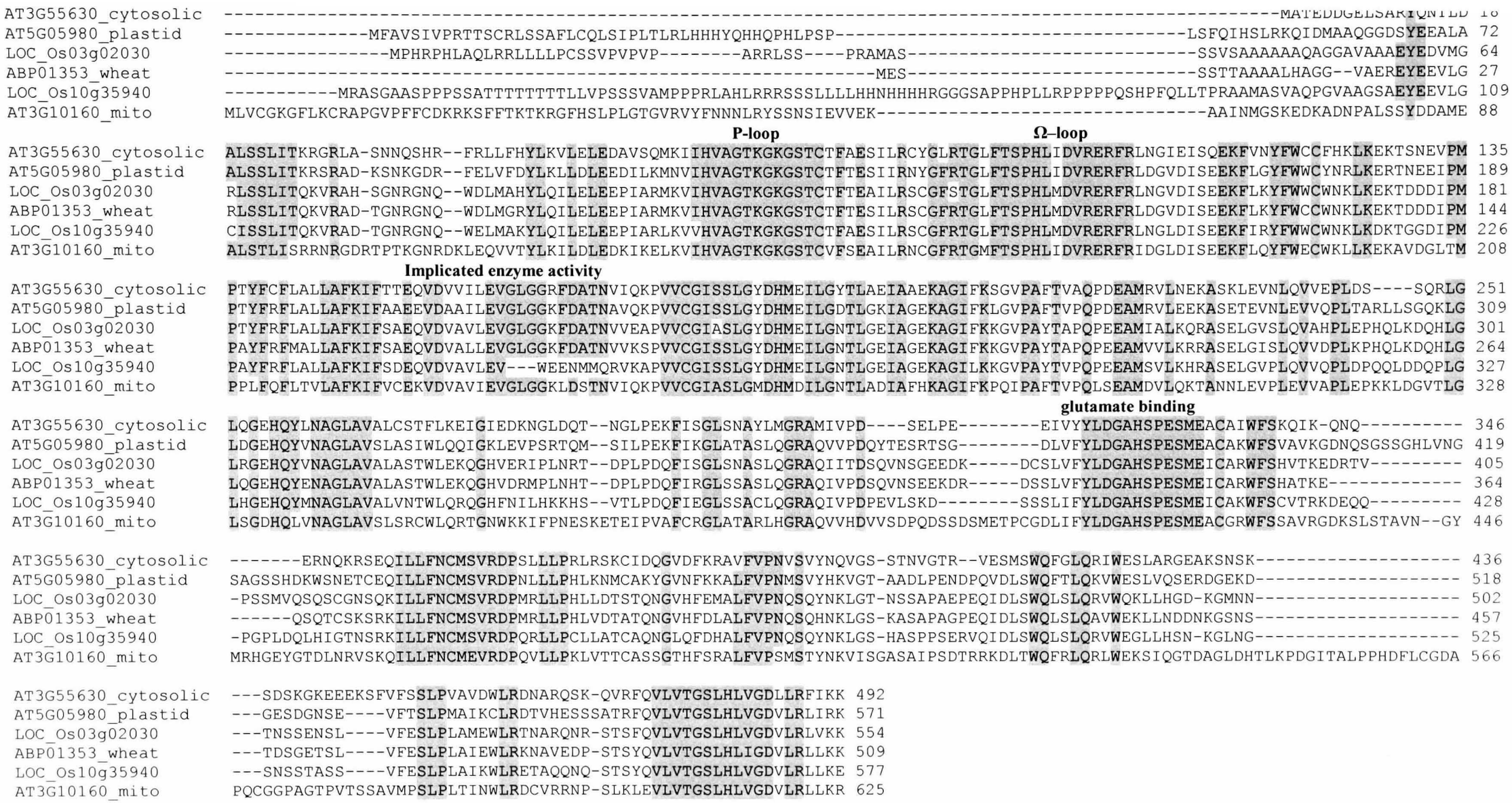


Figure 4.4 Plant FPGS amino acid sequence alignment

FPGS amino acid sequences from *Arabidopsis* (At3g55630, At5g05980 and At3g10160), rice (LOC_Os03g02030 and LOC_Os10g35940) and wheat (ABPO1353) were aligned using CLUSTALW 2.0.10 version provided in www.ebi.ac.uk. The conserved regions are shown in shaded grey. P-loop involves much in ATP-binding site and Ω–loop is a major part of THF binding cleft. The implicated regions of enzyme activity and glutamate binding site are also shown.

Focusing on the targeting sequence, although it was suggested that wheat and rice FPGS Os03g02030 were the mitochondrial isoforms (McIntosh and Henry, 2008), it is still quite difficult to identify the exact location of rice FPGS expression by only comparing their sequences to *Arabidopsis* sequences.

It was clear for the rice FPGS genes that they should not encode cytosolic isoforms because they contain signal peptides (compared to *Arabidopsis* cytosolic FPGS). However, we cannot distinguish the differences between plastid and mitochondria isoforms as both rice FPGS were predicted to contain both mitochondria and plastid signal sequences (from prediction software discussed in chapter 3). Thus, their subcellular localization should be determined experimentally to clarify this point.

Comparison of the entire amino acid sequences of *L.casei*, human, *Arabidopsis*, wheat and rice FPGS allows us to perceive the evolutionary relationships among these species (Figure 4.5). It is clearly seen that rice FPGS protein sequences are most closely related to wheat FPGS.

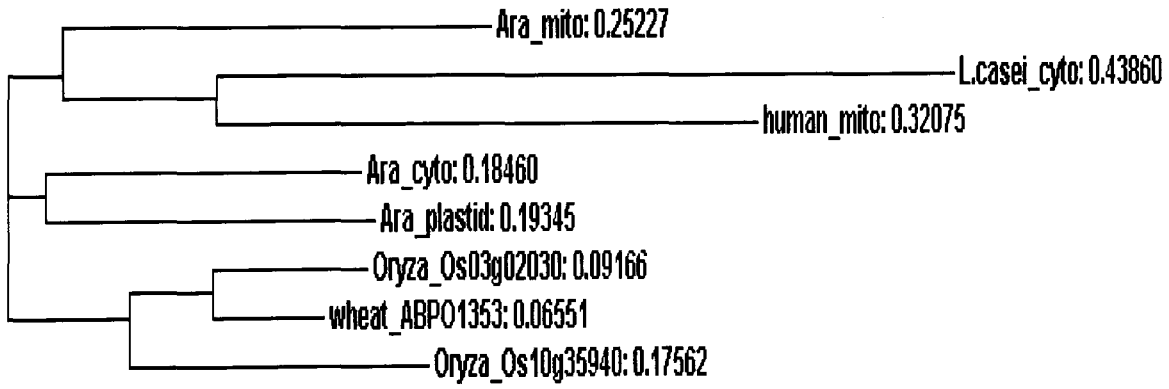


Figure 4.5 Phylogenetic analysis of FPGS proteins from five species

The phylogram was created in EBI CLUSTAL W tool (www.ebi.ac.uk/clustalw). Ara_mito represents *Arabidopsis* FPGS from At3g10160 gene, Ara_cyto represents *Arabidopsis* FPGS from At3g55630 gene, Ara_plastid represents *Arabidopsis* FPGS from At5g05980 gene, Oryza_Os03g02030 represents rice FPGS from LOC_Os03g02030 gene, Oryza_Os10g35940 represents rice FPGS from LOC_Os10g35940 gene, wheat_ABPO1353 represents wheat FPGS from accession no ABPO1353 in NCBI database, human_mito represents human FPGS from accession no Q05932 in UniProt database and L.casei_cyto represents *L.casei* FPGS from accession no P15925 in UniProt database. The numbers behind the name are tree graph distances.

4.3.2 The position of T-DNA insertion on *FPGS* Os03g02030 and *FPGS* Os10g35940 genes

Rice FPGS putative protein sequences are encoded by two distinct *FPGS* genes, LOC_Os03g02030 and LOC_Os10g35940. To characterise FPGS function during rice development, T-DNA insertion lines were characterised. Plant FST line number A16772 disrupting *FPGS* Os03g02030 gene was generated on *japonica* rice *Oryza sativa* cv. Dongjin background and FST line number D02773 disrupting *FPGS* Os10g35940 gene was generated on *japonica* rice *Oryza sativa* cv. Hwayoung background. Lines were ordered from Postech by French Rice Functional Genomics Centre (Montpellier). The screening of T-DNA insertional mutants was performed.

Information about T-DNA insertion lines were obtained from the OryGenes website (http://orygenesdb.cirad.fr). *FPGS* Os03g02030 mutant was created using the pGA2715 vector whilst *FPGS* Os10g35940 mutant was created using the pGA2772 vector. Insert position and the T-DNA map are shown in Figure 4.6. The *FPGS Os03g02030* gene contains 16 exons, with the T-DNA inserted within the 2nd exon (Figure 4.6A). The *FPGS* Os10g35940 gene contains 15 exons, with the T-DNA inserted within the 6th intron (Figure 4.6B).

4.3.3 Characterization of the knock out lines

4.3.3.1 Southern analysis

In order to identify and confirm gene replacement, Southern blot analysis was performed on both FPGS mutant lines by the French Rice Functional Genomics Centre (Montpellier). Hygromycin specific probe was used to indicate a number of T-DNA insertions whilst gene specific probe could distinguish wild type, homozygous or heterozygous T-DNA plants.

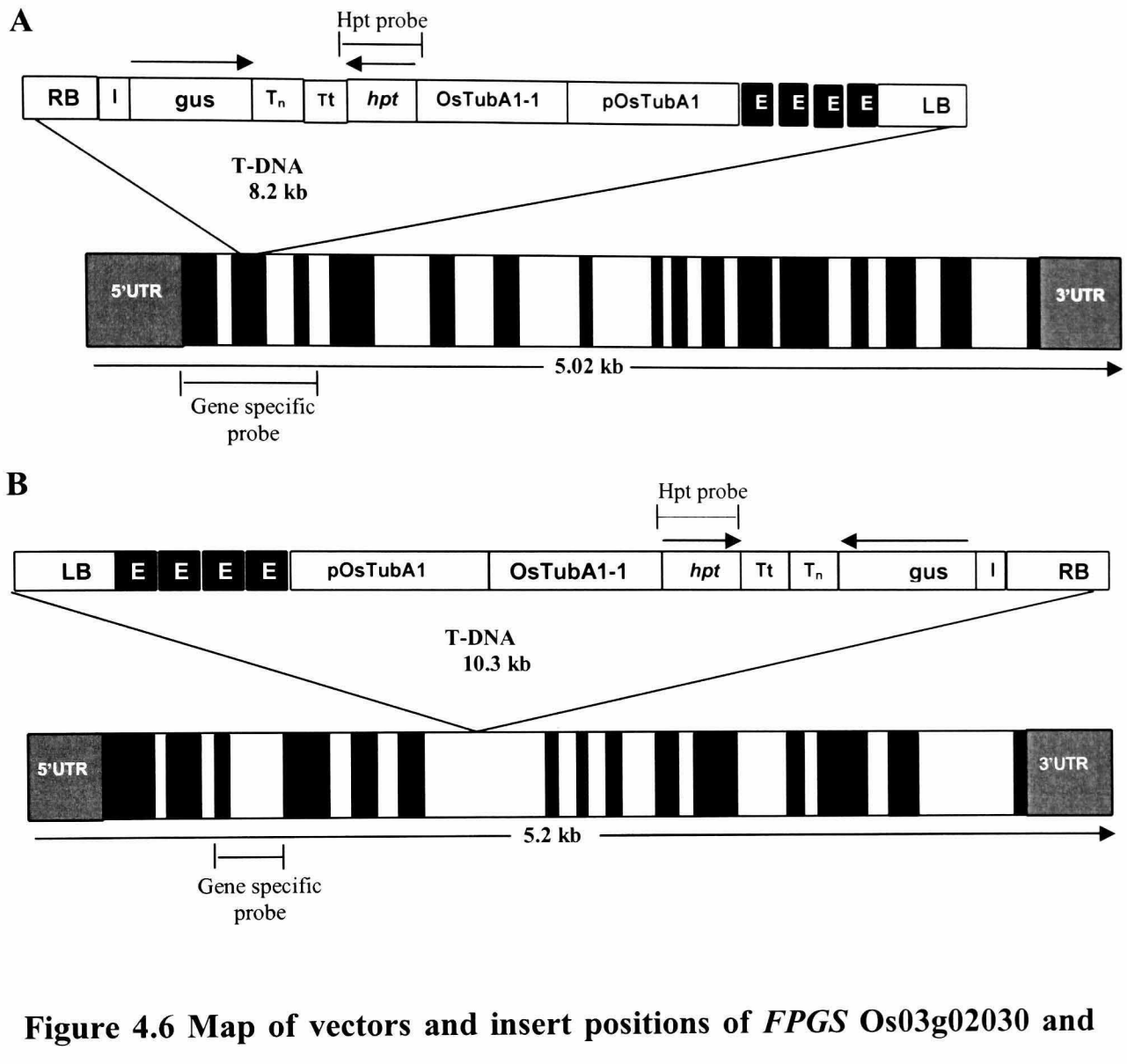


Figure 4.6 Map of vectors and insert positions of *FPGS* Os03g02030 and *FPGS* Os10g35940 knockout lines

A) Map of 8.2 kb T-DNA insertion from the pGA2715 vector and the insert position on *FPGS* Os03g02030 gene in the 2nd exon. B) Map of 10.3 kb T-DNA insertion from the pGA2772 vector and the insert position on *FPGS* Os10g35940 gene in the 6th intron. RB and LB represent the right and left borders of T-DNA. I, OsTubA1 intron 2 carrying three putative splicing acceptor and donor sites; gus, beta-glucuronidase (GUS) reporter gene; Tn, nopaline synthase (nos) terminator; Tt, OsTubA1 terminator; hpt, hygromycin phosphotransferase selectable marker gene; OsTubA1-1, intron 1 of rice alpha tubulin gene, pOsTubA1, rice alpha tubulin promoter; E, CaMV35S enchancer element. Black bars on the gene represent the exons and white bars represent the introns. The arrows indicate the direction of gene expression. The dashes show the position of hygromycin (Hpt) and gene-specific probes used in the southern analysis. The exon and intron sequences of both genes were shown in appendix 4. Map of T-DNA was modified from Jeong *et al.* (2002).

4.3.3.1.1 Southern result for *FPGS* Os03g02030 knock out lines

Nine plants from the A16772 FST target were analysed by Southern analysis. To generate the Southern probes, the hpt_for and hpt_rev primers were used to amplify the 966 bp hygromycin probe and the 03g_for and 03g_rev primers were used to amplify the 777 bp gene specific probe (see appendix 1 for the primer sequences and appendix 4 for probe sequences). Both amplified probes were used in the Southern analysis (Figure 4.6A for probe position on the gene). DNAs from wild type and mutant plants were isolated and restricted with the *Eco*RI enzyme (see Figure 4.7 for the restriction site on the gene flanking sequences).

A 5500 bp fragment was expected from the digested wild type DNA hybridised by the gene specific probe (Figure 4.8B). For the hygromycin probe, only plants carrying T-DNA provided a band (Figure 4.8A). The hygromycin probe revealed that only one band was detected indicating a single T-DNA insert in the A16772 FST target line (Figure 4.8A). The gene specific probe revealed different hybridized band patterns between wild type and T-DNA containing plants (Figure 4.8B). Compared to Dongjin wild type, plant number 2 and 4 were wild type, plant number 3, 6, 7 and 14 were heterozygous and plant number 5 and 15 were homozygous for the T-DNA insertion (circles in Figure 4.8B). Hence, the progeny of plant number 5 was chosen for further analysis. (see next section).

4.3.3.1.2 Southern result for *FPGS* Os10g35940 knock out lines

For *FPGS* Os10g35940 gene, DNAs from wild type and 34 individual plants from the FST line number D02773 were analysed by Southern analysis. To generate the Southern probes, the 10g_for and 10g_rev primers were used to amplify the 563 bp gene specific probe (see appendix 1 for the primer sequences and appendix 4 for probe sequences). The same hygromycin probe used with *FPGS* Os03g02030 mutants was used (Figure 4.6B for probe position on the gene).

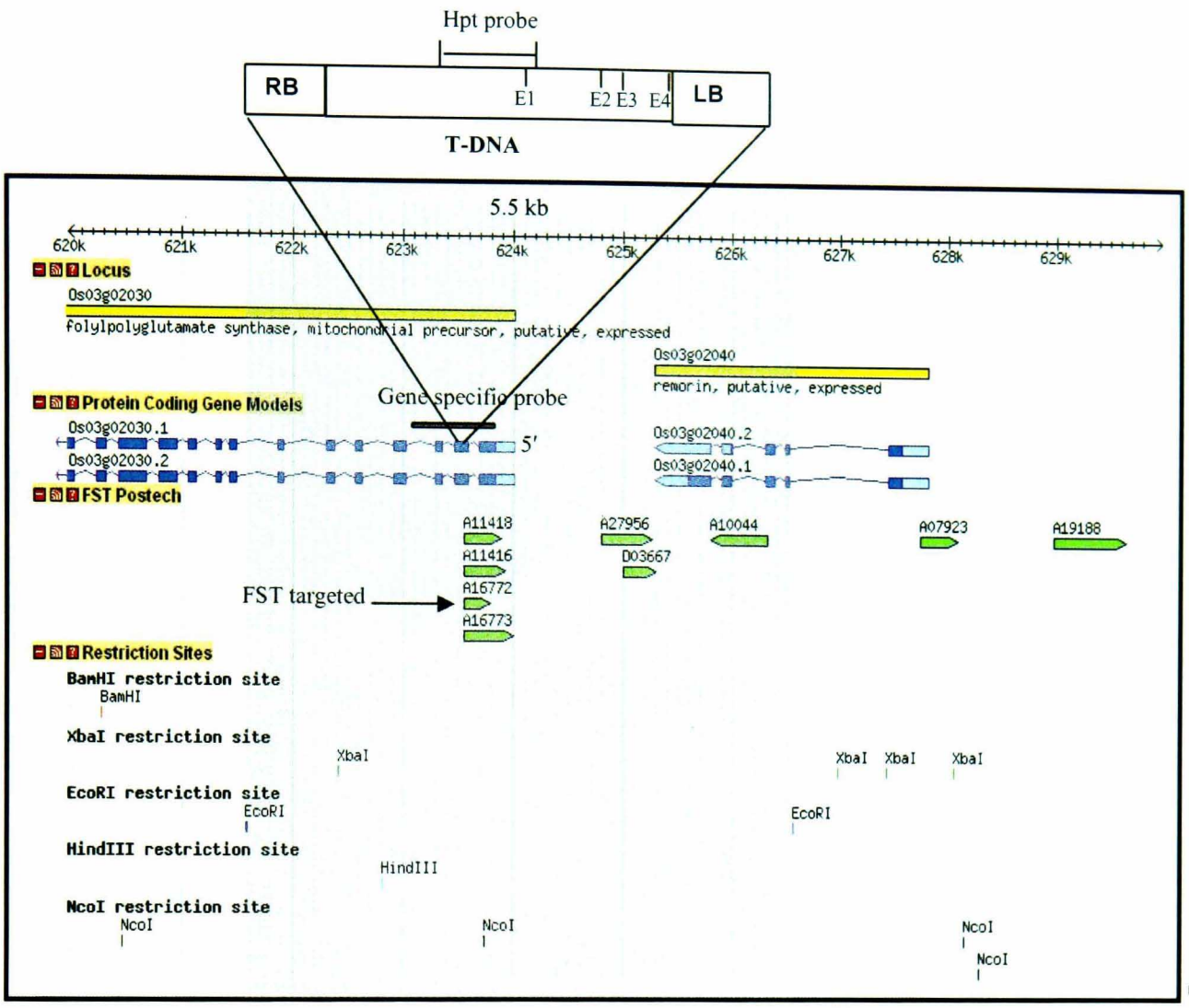


Figure 4.7 Map of restriction sites on *FPGS* Os03g02030 gene and the T-DNA insert

Map of the FST targeted A16772 and the insert position on *FPGS* Os03g02030 gene in the 2nd exon. The several restriction sites on the wild type rice chromosome are shown. For the southern analysis, *Eco*RI enzyme was used. The *FPGS* 03g02030 gene is cut a single site by *Eco*RI in the 7th intron. Another *Eco*RI site is outside this *FPGS* gene. The T-DNA insert from the pGA2715 vector has four *Eco*RI sites, E1, E2, E3 and E4 at 3769 bp, 6151 bp, 6784 bp and 8194 bp, respectively. When genomic DNA of the wild type rice was digested by *Eco*RI enzyme, the 5500 bp fragment was detected by using the gene-specific probe. This diagram was modified from OrygeneDB website (http://orygenesdb.cirad.fr/index.html). The probe details were provided by Delphine Mieuvet and Nadège Lanau, French Rice Functional Genomics Centre.

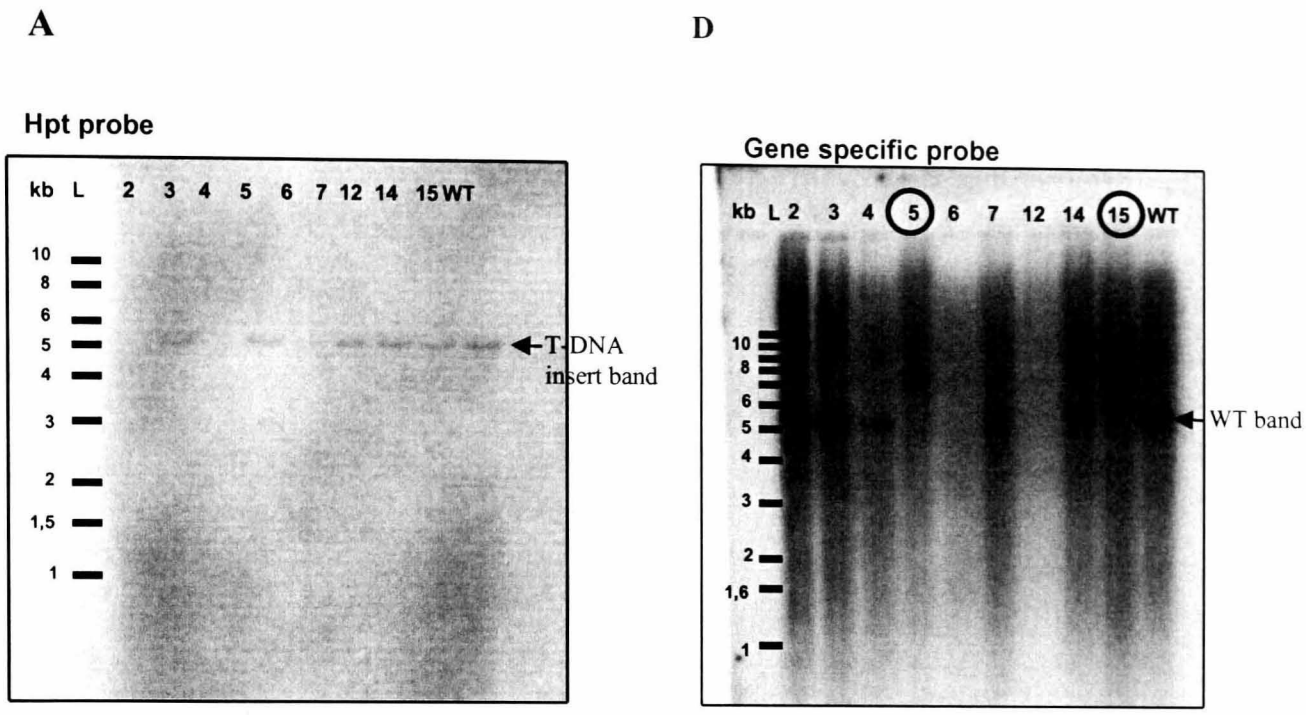


Figure 4.8 Southern analysis of *FPGS* Os03g02030 knockout lines

A) Hybridization with hygromycin specific probe. Only plants carrying T-DNA insert provided band. B) Hybridization with gene specific probe. The expected band size for the wild type fragment was 5500 bp, whilst the pattern of bands for the T-DNA inserted plant would be different. L represents DNA ladder; WT represents Dongjin wild type. The circles indicate homozygous plants. This data was generated by Delphine Mieuvet, Nadège Lanau and Helen Parker at French Rice Functional Genomics Centre.

DNAs from the wild type and mutant plants were isolated and restricted with *Eco*RI enzyme. Figure 4.9 shows the restriction sites on the *FPGS* Os10g35940 gene flanking sequence and the position of the gene specific probe. A 8100 bp fragment was expected from the digested wild type DNA hybridised by the gene specific probe.

The hygromycin probe revealed at least two bands indicating at least two T-DNA inserts were found in this mutant line (Figure 4.10A). The gene specific probe revealed different hybridized band patterns between wild type and T-DNA containing plants (Figure 4.10B). Compared to Hwayoung wild type, plant number 1, 2 and 3 were wild type and plant number 4, 5, 6, 7, 8 and 9 had at least two copies of inserts which were heterozygous plants. Unfortunately,

only heterozygous plants were found. To reconfirm by PCR, the progeny of those plants were genotyped (see next section).

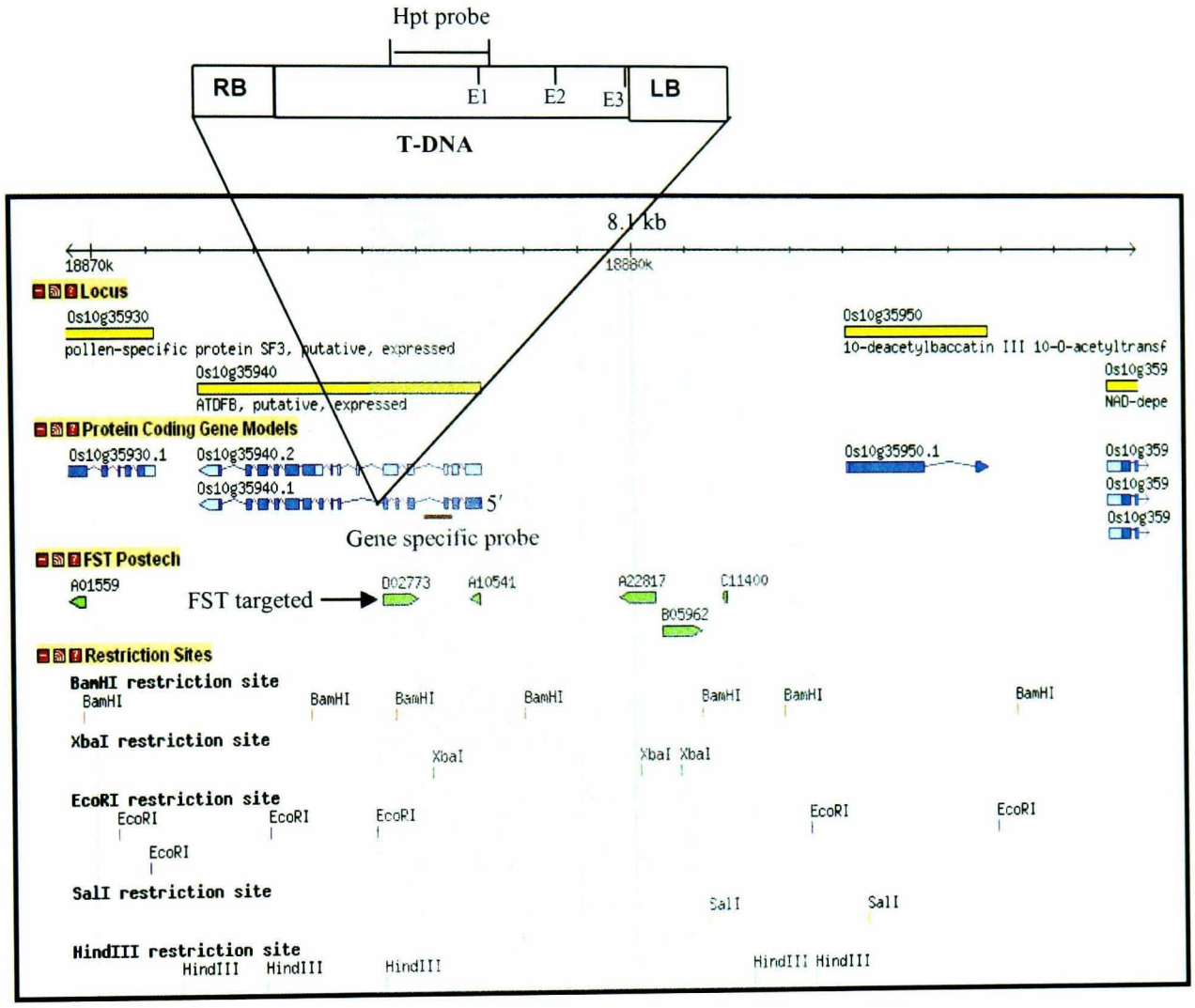


Figure 4.9 Map of restriction sites on *FPGS* Os10g35940 gene and the T-DNA insert

Map of the D02773 FST targeted and the insert position on *FPGS* Os10g35940 gene in 6th intron. The several restriction sites on the wild type rice chromosome are shown. For the southern analysis, *Eco*RI enzyme was chosen. The T-DNA insert from the pGA2772 vector has three *Eco*RI sites, E1, E2 and E3 at 3769 bp, 6151 bp and 10244 bp, respectively. When genomic DNA of the wild type rice was digested by *Eco*RI enzyme, the 8100 bp fragment was detected by using the gene-specific probe. This diagram was modified from OrygeneDB website. The probe details were provided by Delphine Mieuvet and Nadège Lanau, French Rice Functional Genomics Centre.

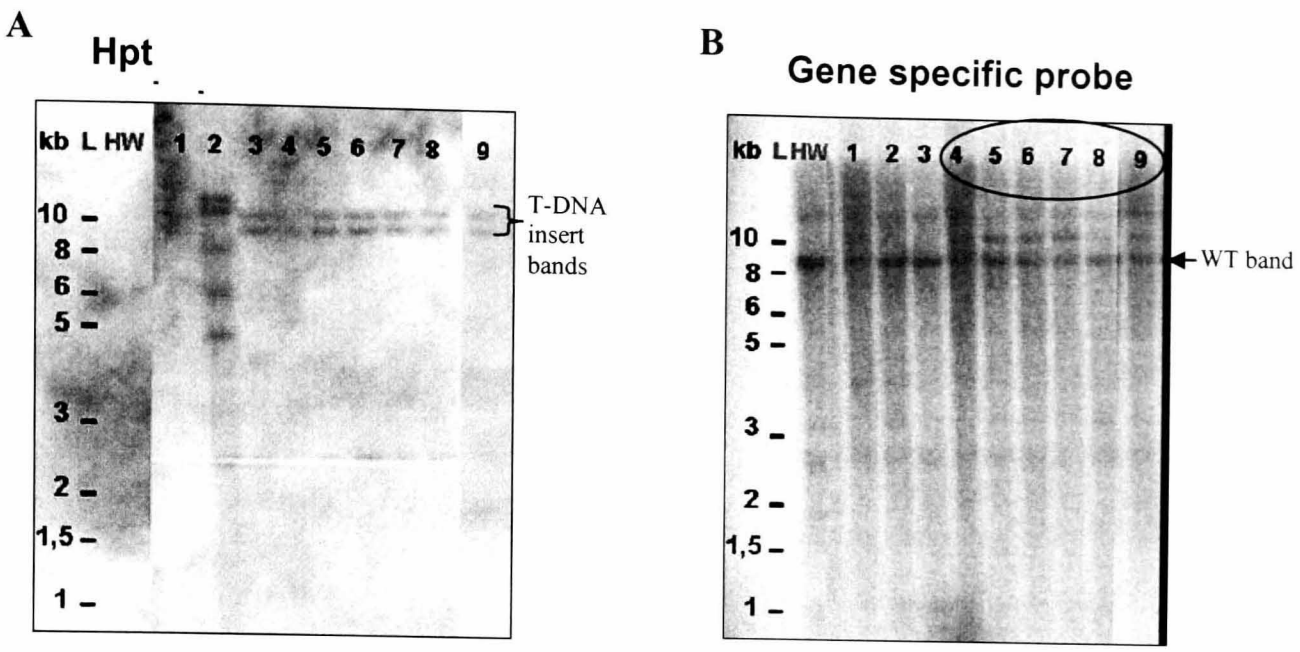


Figure 4.10 Southern analysis of *FPGS* Os10g35940 knockout lines

A) Hybridization with hygromycin specific probe. Only plant carrying T-DNA provided band. At least two T-DNA insertions were detected. B) Hybridization with gene specific probe. The expected band size for the wild type fragment was 8100 bp, whilst the heterozygous plants provided the different band pattern. L represents DNA ladder; HW represents Hwayoung wild type. The circle indicates heterozygous plants. This data was generated by Delphine Mieuvet at French Rice Functional Genomics Centre, Montpellier.

4.3.3.2 PCR and RT-PCR analysis

To confirm the original Southern blot results, PCR was performed using *FPGS* gene-specific primers and T-DNA primers. The progeny of plant number 5 of FST number A16772 (*FPGS* Os03g02030) and the progeny of all plants of FST number D02773 (*FPGS* Os10g35940) were extracted for DNA. PCR was performed using the following primers (see appendix 1 for primer sequences). 3gE1_for and 3gE4_rev primers were used to amplify a partial sequence of the wild type *FPGS* Os03g02030 gene. The PCR product of 991 bp was expected. 10gE3_for and 10gE7_rev primers were used to amplify a partial sequence of the wild type *FPGS* Os10g35940 gene. The PCR product of 1629 bp was expected.

For the presence of the T-DNA, 3gE1_for and RB_rev primers were used for screening the T-DNA insertion mutant of *FPGS* Os03g02030 gene. The expected PCR product size was 667 bp. 10gE3_for and LB_rev primers were used for screening the T-DNA insertion mutant of *FPGS* Os10g35940 gene. The expected product size was 1183 bp. Figure 4.11A and B shows the diagram of designed primers on the *FPGS* Os03g02030 and *FPGS* Os10g35940 genes, respectively.

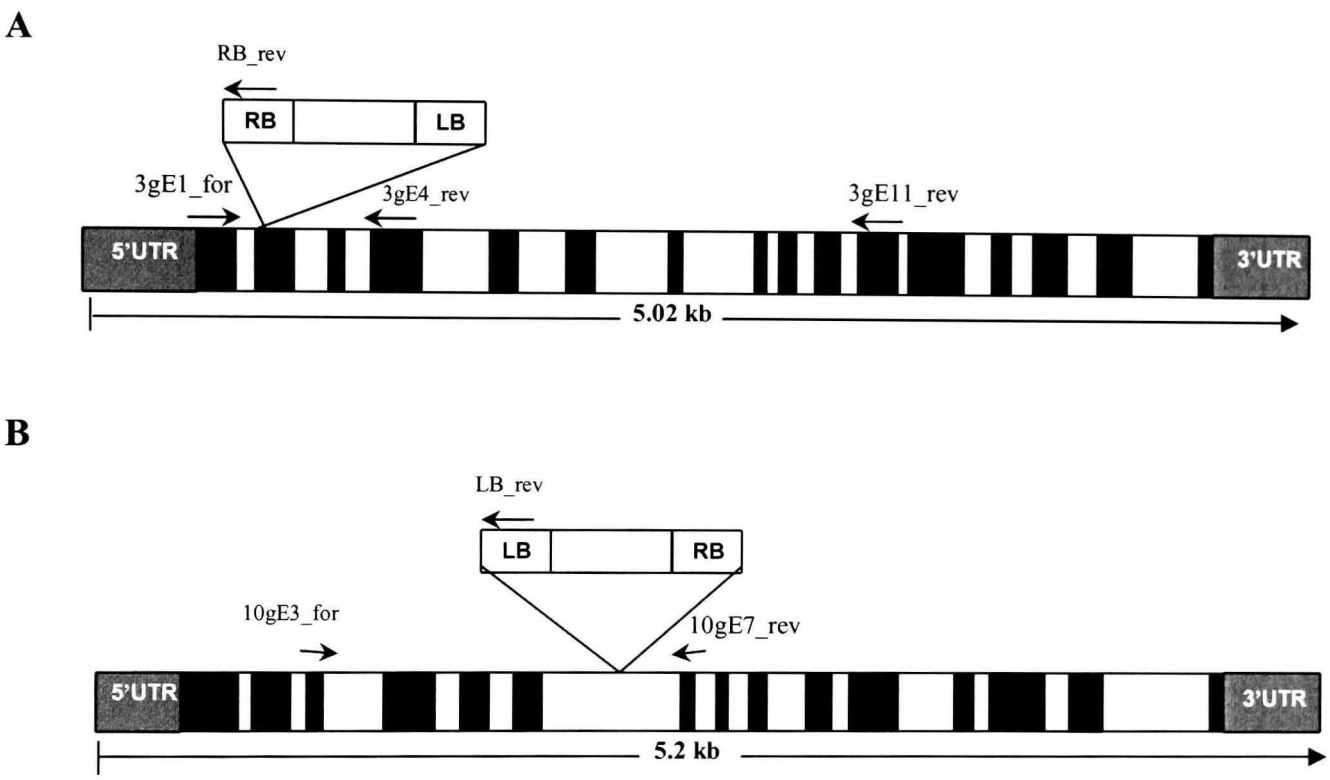


Figure 4.11 Diagram of primer positions on both rice *FPGS* genes

A: *FPGS* Os03g02030 genomic diagram containing 16 exons (black bars) and 15 introns (white bars). T-DNA was inserted into the 2nd exon. 3gE1_for, 3gE4_rev, 3gE11_rev and RB_rev are PCR primers. The PCR product size from 3gE1_for and 3gE4_rev primers was 991 bp and that from 3gE1_for and RB_rev primers was 667 bp. 3gE1_for and 3gE11_rev primers were used in the RT-PCR which the product size was 1016 bp.

B: *FPGS* Os10g35940 genomic diagram containing 15 exons (black bars) and 14 introns (white bars). T-DNA was inserted into the 6th intron. 10gE3_for, 10gE7_rev and LB_rev are PCR primers. The PCR product size from 10gE3_for and 10gE7_rev primers was 1629 bp and that from 10gE3_for and LB_rev primers was 1183 bp. 10gE3_for and 10gE7_rev primers were also used in the RT-PCR which the product size was 458 bp. The arrows indicate the direction of primers and gene expression. UTR, untranslated region; RB, right border of T-DNA; LB, left border of T-DNA.

4.3.3.2.1 PCR and RT-PCR result of *FPGS* Os03g02030 knock out line

The DNA from the second generation (F_2) of the plant number 5 disrupted on *FPGS* Os03g02030 gene was subjected to PCR. The result is shown in Figure 4.12A compared to Dongjin wild type.

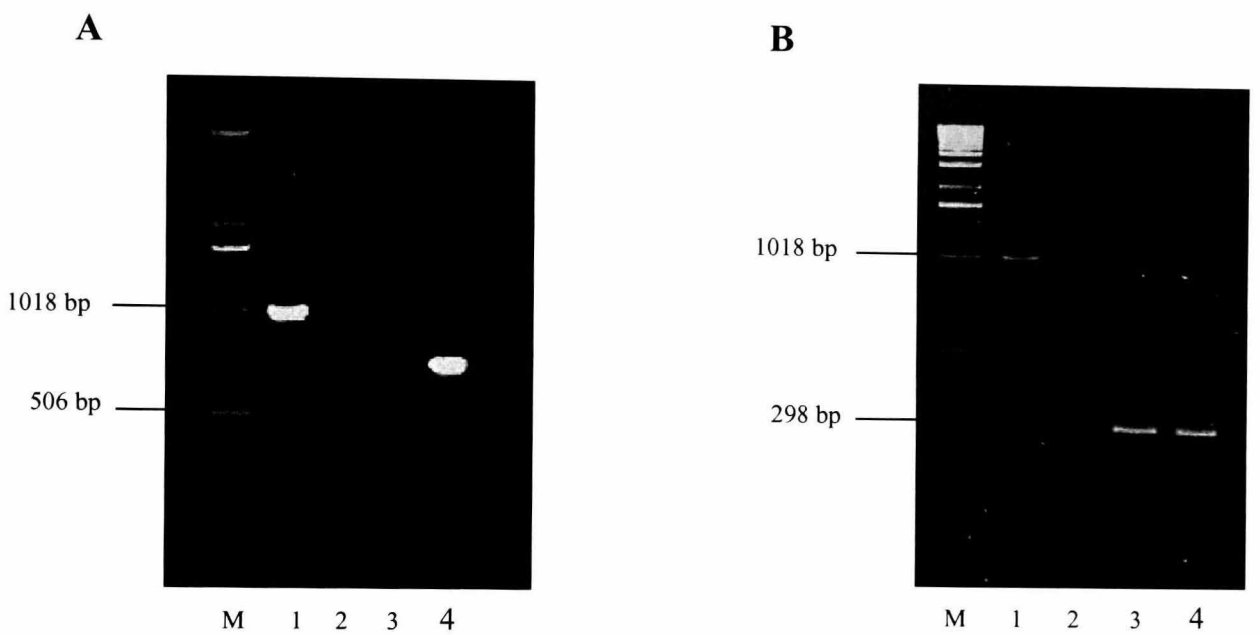


Figure 4.12 PCR and RT-PCR analysis of *FPGS* Os03g02030 lines

DNA and RNA extracted from the third leaves (2 week-old) of Dongjin wild type and *FPGS* Os03g02030 homozygous knock out line number 5 were analyzed by PCR (A) and RT-PCR (B). A: Lane 1 and 2 were the result from using 3gE1_for and 3gE4_rev primers and the product size of 991 bp was expected. The positive band on this primer pair confirmed the wild type genotype. Lane 3 and 4 were the result from using 3gE1_for and RB_rev primers and the expected product size was 667 bp. The positive band on this primer pair suggested the presence of the T-DNA insertion. Dongjin wild type DNA was applied in the PCR reaction of lane 1 and 3 and *FPGS* Os03g02030 homozygous knock out line DNA was applied in the PCR reaction of lane 2 and 4.

B: Lane 1 and 2 were the RT-PCR result from 3gE1_for and 3gE11_rev primers which the expected product size was 1016 bp. This primer pair was used to confirm the mRNA expression of *FPGS* Os03g02030 gene. Lane 3 and 4 were the RT-PCR result from the rice actin primers which the expected band size was 276 bp. Actin bands demonstrated an equal amount of loaded cDNA. Dongjin wild type cDNA was applied in the RT-PCR reaction of lane 1 and 3 and *FPGS* Os03g02030 homozygous knock out line cDNA was applied in the RT-PCR reaction of lane 2 and 4. M represents 1 kb ladder standard.

Lane 1 was the result from Dongjin wild type DNA using 3gE1_for and 3gE4_rev primers. The 991 bp band indicated the wild type plant. Compared to the plant number 5 mutant, this band was not detected from the same primers (lane 2). Lane 3 was the result from Dongjin wild type DNA using 3gE1_for and RB_rev primers. This primer pair was used to detect the 667 bp band expected in the plant carrying T-DNA. No band was detected in lane 3 indicating no T-DNA insertion in Dongjin wild type, whilst only one specific band was detected in lane 4. Lane 4 was the result from the plant number 5 mutant. Hence, it indicated that plant number 5 contained a homozygous T-DNA insertion within the *FPGS Os03g02030* gene

In addition, RT-PCR was also performed to monitor the impact of the T-DNA insertion on *FPGS Os03g02030* mRNA abundance using cDNA specific primers, 3gE1_for and 3gE11_rev (see appendix 1 for primer sequences and Figure 4.11A for primer position). The expected band size was 1016 bp indicating the normal expression of this *FPGS* gene in Dongjin wild type (lane1, Figure 4.12B). Lane 2 showed that *FPGS Os03g02030* mRNA was not detected in the T-DNA line whereas transcripts were detected for the actin positive control (lane 4). Hence, line number 5 represented a loss of function allele for *FPGS* Os03g02030 gene and would be chosen for further phenotypic and qPCR analysis referred to as *fpgs*03g in the rest of the thesis.

Moreover, the existence of T-DNA insertion was tested in *fpgs*03g mutant as the T-DNA tagging vector, pGA2715, employed to generate the knock out line contained a GUS gene (Figure 4.6). The mutant rice plants were examined for constitutive GUS expression in order to confirm the stability of the T-DNA insertion. The third (F_3) generation of *fpgs*03g was checked for GUS expression in young leaf, root, flower, dough seed and mature seed. Blue staining indicated GUS expression in rice tissues which could be detected in the tip of the young leaf and in the aleuronic layer of the dough stage seed (Figure 4.13). Hence, *fpgs*03g still contained T-DNA vector in F_3 generation resulting in a loss-of-function phenotype for the *FPGS* Os03g02030 gene.

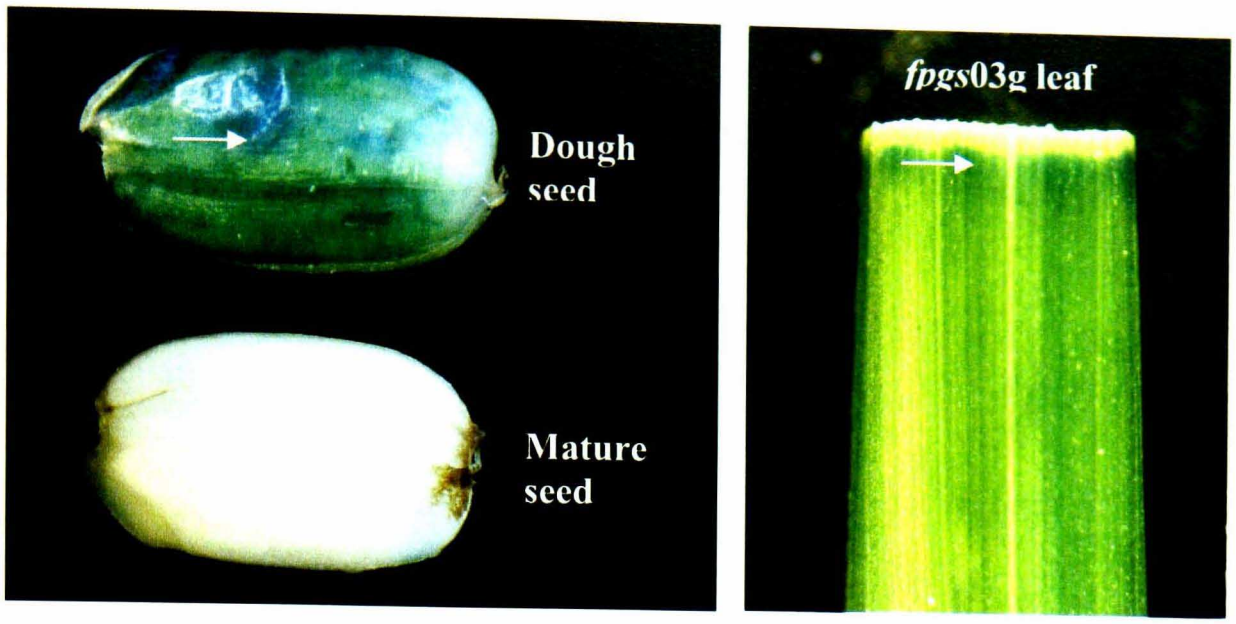


Figure 4.13 GUS staining for the homozygous *FPGS* 03g02030 line

*fpgs*03g plant was checked GUS expression in leaf (right picture) and seeds (left picture). The arrows indicate blue staining from GUS expression.

4.3.3.2.2 PCR and RT-PCR result of *FPGS* Os10g35940 knock out line

The DNA from the second generation (F_2) of all *FPGS* Os10g35940 T-DNA insertion lines was used to performed PCR using two primer pairs. 10gE3_for and 10gE7_rev primers were used to indicate the wild type *FPGS* Os10g35940 gene with 1629 bp PCR product. In total 80 plants were screened (Figure 4.14). Compared to Hwayoung wild type, 5 out of these 80 plants which were plant numbers 3, 20, 37, 62 and 63, did not give the wild type *FPGS* Os10g35940 band. Possibly, they might have a T-DNA insert. These 5 plants were selected for further analysis.

To test the existence of T-DNA in the 5 selected plants, 10gE3_for and LB_rev primers were used. Unfortunately, no positive band was detected suggesting an unexpected T-DNA orientation on chromosome. Hence, both LB and RB primers were then used to test the orientation of T-DNA insert (details of all primers used, see the caption of Figure 4.15). Although all LB and RB primers were used, the positive band indicating the insertion on the *FPGS* gene was not detected.

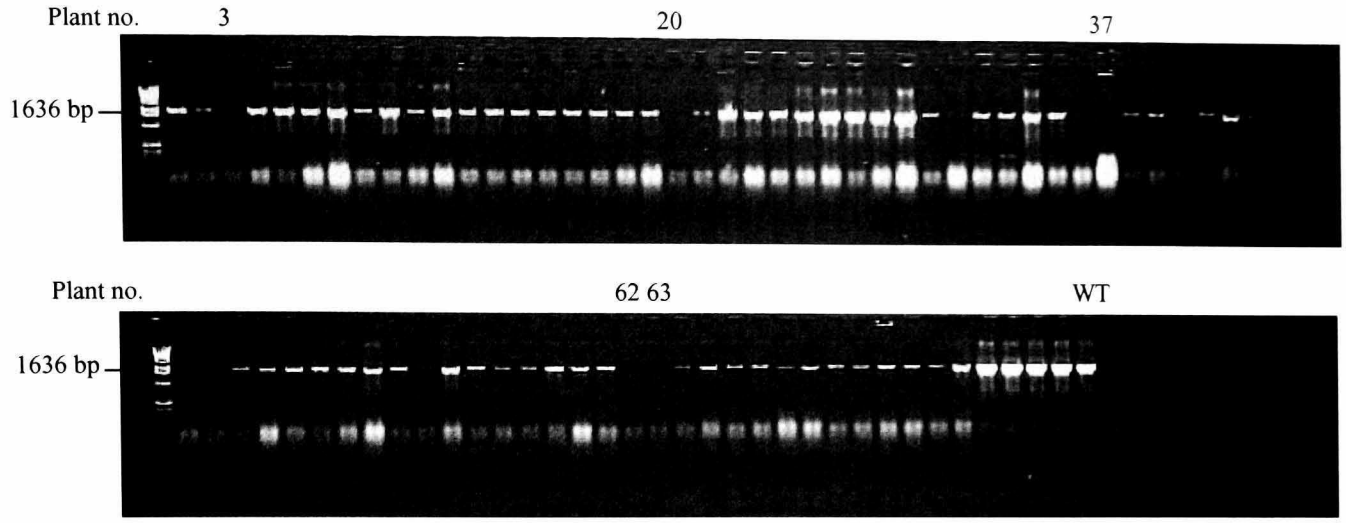


Figure 4.14 PCR analysis for the T-DNA insertion of *FPGS* Os10g35940 lines

DNA extracted from the third leaves (2 week-old) of Hwayoung wild type and *FPGS* Os10g35940 lines were analyzed by PCR. A: The PCR result from 10gE3_for and 10gE7_rev primers with the expected band size was 1629 bp. The positive band indicated the existence of normal *FPGS* Os10g35940 gene. M represents 1 kb ladder standard. N represents the negative control using pGWB7 vector lacking *FPGS* Os10g35940 gene. B represents the negative control using water as the sample.

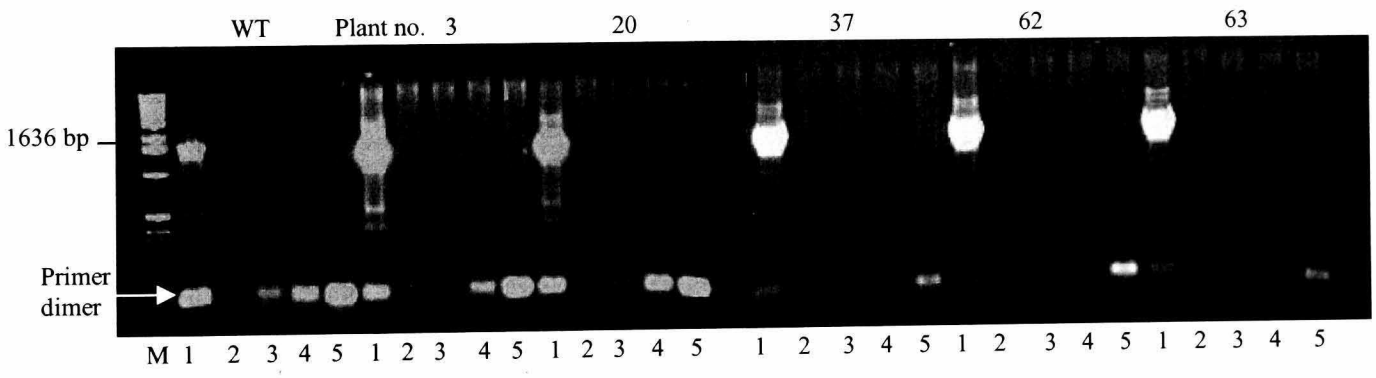


Figure 4.15 PCR analysis for T-DNA orientation on *FPGS* Os10g35940 gene

The DNA from Hwayoung and selected *FPGS* Os10g35940 mutants were analyzed by PCR using 5 primer pairs indicated by the number 1-5. 1: the PCR result from 10gE3_for and 10gE7_rev primers with the expected band size was 1629 bp indicating the normal *FPGS* Os10g35940 gene existence. 2: the PCR result from RB_for and 10gE7_rev primers. 3: the PCR result from 10gE3_for and RB_rev primers. 4: the PCR result from LB_for and 10gE7_rev primers. 5: the PCR result from 10gE3_for and LB_rev primers. Positive band from only the first primer pair was detected. M represents 1 kb ladder standard.

To confirm whether the selected five plants contain T-DNA insert, the hygromycin gene was amplified from the same set of DNA using hpt_for and hpt_rev primers (see Appendix 1 for the primer sequences).The expected band size was 368 bp. All tested 5 plants provided the positive band (Figure 4.16). Hence, all selected 5 plants contain the T-DNA insert but with unknown position and orientation of the insert.

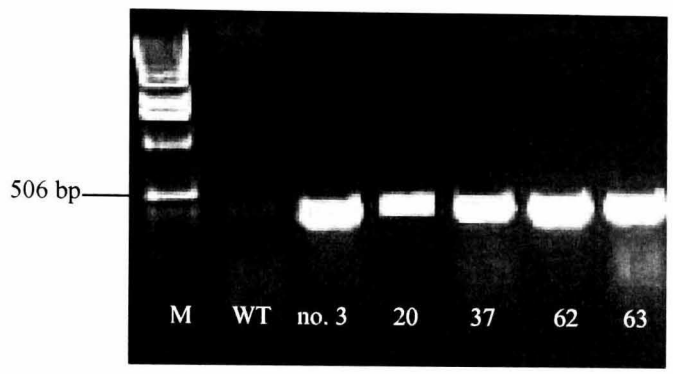


Figure 4.16 PCR analysis for hygromycin gene in *FPGS* Os10g35940 lines

DNA of Hwayoung wild type (WT) and mutant plant number 3, 20, 37, 62 and 63 were PCR identified for the hygromycin gene using hpt_for and hpt_rev primers. The expected band size was 368 bp. All selected 5 mutant plants provided the positive band. M represents 1 kb ladder standard.

The selected 5 mutant plants thus contained the T-DNA insert but the position and orientation of the insert on the target gene was unclear. To test whether these plants are true knock out lines for *FPGS* Os10g35940 gene, RT-PCR was performed to check the expression of this gene. Unfortunately, all selected plants showed the expression of *FPGS* Os10g35940 gene at a similar level to the wild type plant (Figure 4.17). Hence, the T-DNA insertion event in these lines has not knocked out or knocked down the expression of the *FPGS* Os10g35940 gene. Those plants would not be able to be used in the further observation. Hence, the phenotypic and expression analysis would continue in only the homozygous plant of *FPGS* Os03g02030 gene.

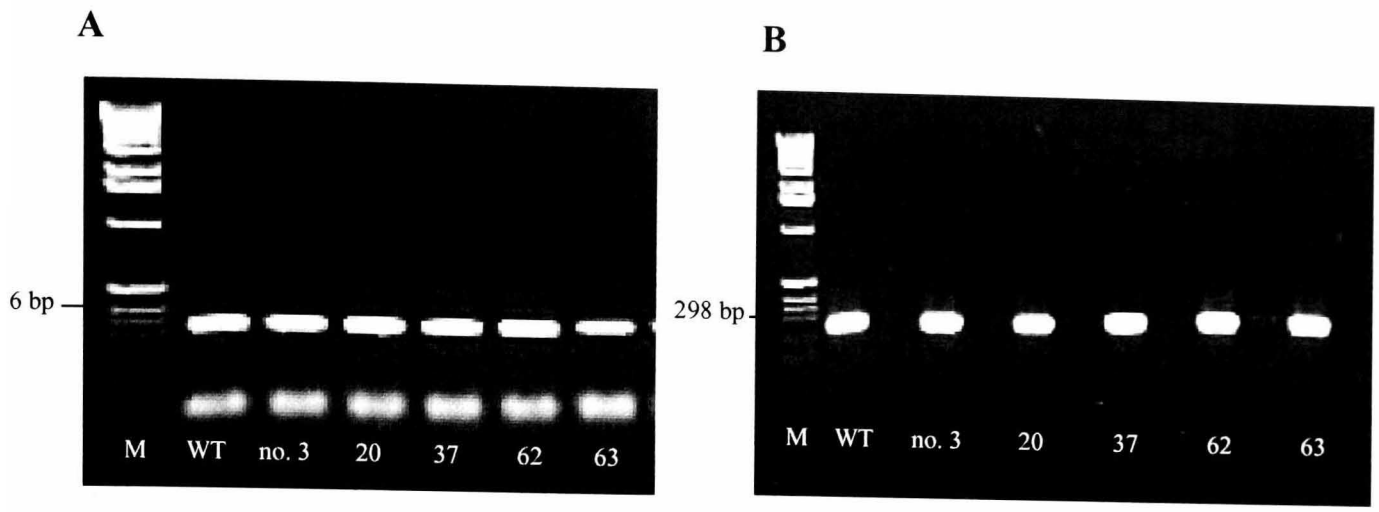


Figure 4.17 RT-PCR analysis of selected *FPGS* Os10g35940 lines

A) cDNA of Hwayoung wild type (WT) and selected 5 plants of *FPGS* Os10g35940 lines were amplified for the expression of *FPGS* Os10g35940 gene using 10gE3_for and 10gE7_rev primers. All selected 5 mutant plants showed the similar level of expression in wild type plant. B) cDNA of Hwayoung wild type (WT) and selected 5 plants of *FPGS* Os10g35940 lines were amplified for the actin gene expression using the rice actin primers which the expected band size was 276 bp. Actin bands demonstrated an equal amount of loaded cDNA. M represents 1 kb ladder standard.

4.3.4 Phenotypic analysis of *FPGS* Os03g02030 knock out plant

To determine the function of the *FPGS* Os03g02030 gene, the phenotype of the *fpgs*03g mutant line was characterised throughout the rice life cycle. Twenty *fpgs*03g mutant plants were grown in soil in controlled conditions and their phenotypes were monitored at every stage of development. All 20 mutant plants showed the same phenotype. From the seedling stage until the stem elongation stage, the leaves and shoot phenotypes of the knock out mutant appeared wild type (Figure 4.18A). However, during seed development, the *fpgs*03g mutant exhibited slower grain filling and later ripening (Figure 4.18B). Observations made every day noted that the knock out rice line was 1 week late in reaching the heading stage compared to wild type and was delayed at every subsequent step, reaching the mature stage approximately 2 weeks later than Dongjin wild type. Despite the seed development phenotype of the knock out plants, the mature seeds were wild type-like in appearance and ability to germinate (Figure 4.18C and D). The germination rate of the *fpgs*03g mutant

was about 95% compared to Dongjin wild type. The number of seeds per ear (approximately 60 seeds) and ears per plant of the *fpgs*03g mutant were also similar to wild type. Hence, the developmental delay exhibited by the *fpgs*03g mutant during the seed setting stage suggested that the *FPGS Os03g02030* gene played a key role during the seed filling stage.

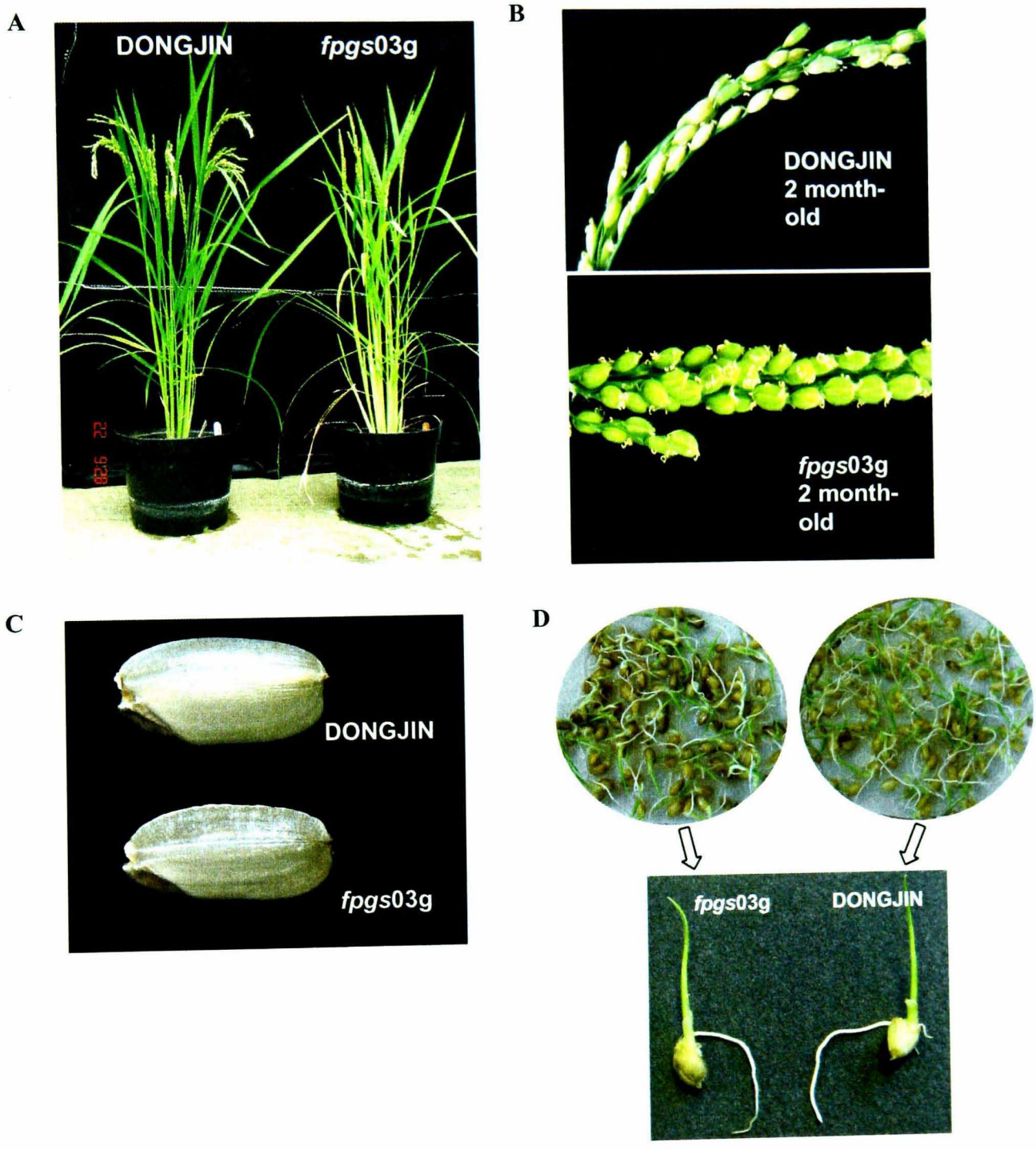


Figure 4.18 The plant phenotypes of *fpgs*03g and Dongjin wild type.

A: 2-month old plants; B: 2-month old seeds; C: hulled mature dry seeds; D: 4 -day old germinated seeds

4.3.5 Quantification of folate levels in *fpgs*03g mutant

To directly determine whether knocking out *FPGS Os03g02030* gene function affected the abundance of polyglutamylated folates and total folate levels, folate quantification was performed using liquid chromatography/negative ion electrospray ionisation tandem mass spectrometry (LC-MS/MS) in collaboration with Riza Abilgos-Ramos. The concentration of folate derivatives in rice tissues are summarized in table 4.2. The predominant form of mono-glutamylated folate detected in rice seeds (S) and leaf (L) was 5-methyltetrahydrofolate. With respect to polyglutamylated folates, only tetra- and penta-glutamylated forms were detected in leaf and seed samples.

Table 4.2 Total folate levels and folate derivatives in Dongjin wild type and *fpgs*03g mutant

Rice Variety *(n = 3)*	Folate Form (µg/100g)					
	5-CH_3 H_4PteGlu	5&10-CHO H_4PteGlu	5-CH_3 H_4PteGlu$_4$	5-CH_3 H_4PteGlu$_5$	5-CHO H_4PteGlu$_5$	Total Folate
Dongjin (L)	5.16 ± 0.72	1.62 ±0.24	0.33± 0.01	0.7± 0.40	4.36± 0.47	12.17± 0.40
*fpgs*03g (L)	3.88 ± 0.14	0.79± 0.18	0.67± 0.01	0.7 ± 0.01	2.65 ± 0.1	8.69± 0.10
Dongjin (S)	3.21± 0.70	1.87±0.28	0.33± 0.01	0.3 ± 0.004	0.74±0.01	6.45± 1.00
*fpgs*03g (S)	2.13±0.30	1.72±0.61	0.11±0.001	0.1±0.00	0.22±0.01	4.28± 0.20

5-CH_3-H_4PteGlu = 5-methyltetrahydropteroylmonoglutamated;
5-CHO-H_4PteGlu = 5-formyltetrahydropteroylmonoglutamated;
5-CH_3-H_4PteGlu$_4$ = 5-methyltetrahydropteroyltetraglutamate;
5-CH_3-H_4PteGlu$_5$ = 5-methyltetrahydropteroylpentaglutamate;
5-CHO-H_4PteGlu$_5$ = 5-formyltetrahydropteroylpentaglutamate;
(L) represents a leaf sample and (S) represents a mature seed sample
The mean ± SD of the folate contents was presented with n = 3
No statistical analysis was possible in this experiment.

The level of the major polyglutamylated folate form, 5-CHO-H_4PteGlu$_5$, in *fpgs*03g leaf and seed tissues was lower than the wild type. Similarly, 5-CH_3-H_4PteGlu$_5$ levels were decreased in seed tissue. The monoglutamate folates (methyl- and formyltetrahydrofolate) were reduced in only seed of the knock out line. Overall, there was a 33% and 28% decrease in total folate levels in seed and leaf samples of the knock out line compared to the wild type,

respectively. Hence, loss of *FPGS 03g02030* gene expression resulted in reduction of both mono- and polyglutamylated folate content.

4.3.6 Transcript abundance of folate biosynthesis genes in the *fpgs*03g mutant

To examine the effect of disrupting the *FPGS* Os03g02030 gene on other folate-related genes, RT-qPCR was performed on cDNA prepared from mRNA isolated from leaf and grain of *fpgs*03g and Dongjin wild type. Figure 4.19A shows mRNA abundance in grain of thirteen genes encoding the enzymes in the folate biosynthesis pathway. The mRNA abundance of almost every biosynthetic gene (*GTPCHI, ADC synthase, ADC lyase, HPPK/DHPS*) increased in the *fpgs*03g knock out line compared to the wild type. The transcripts of most genes increased in the knock out except the *FPGS Os10g35940* gene which was expressed at a similar level to the wild type and *GGH* which decreased approximately 2-fold. In leaf (Figure 4.19B), the transcript abundance of all genes increased in the knock out similar to grain except the *FPGS Os10g35940* gene which was increased approximately 2-fold.

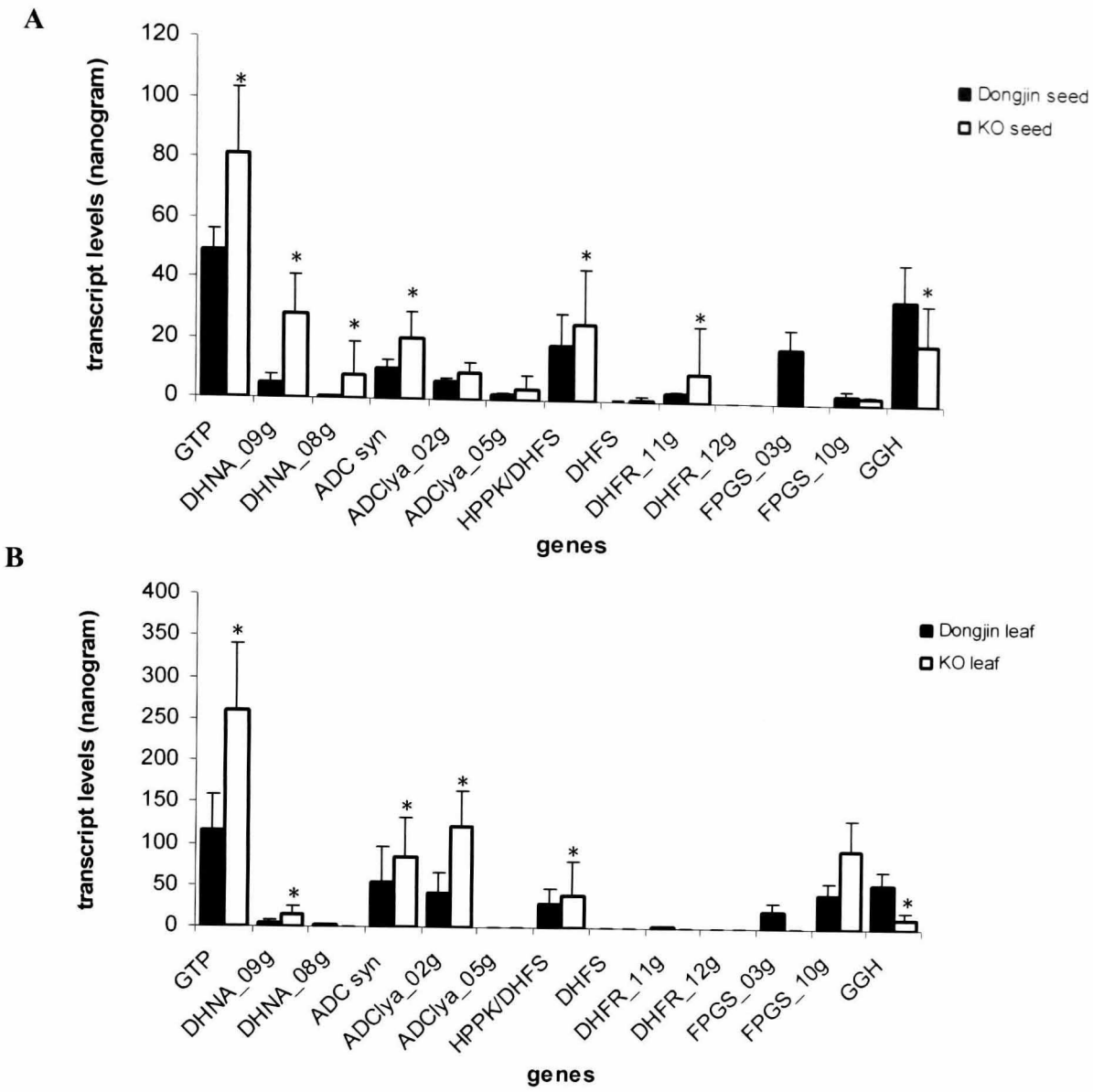


Figure 4.19 Expression profiles of folate-related genes in *fpgs*03g mutant and Dongjin wild type

Dry mature seeds (A) and leaves (B) of Dongjin and *fpgs*03g mutant were RT-qPCR proflied. For seed samples, 40 dry mature seeds from 10 different plants were pooled per analysis. For leaf samples, the third green leaf from 5 plants was pooled per analysis. The transcript values are provided in nanograms representing the amount of mRNA for a particular gene relative to the actin transcript. All folate-related genes including *GTPCHI* LOC_Os04g56710 (GTP), *DHNA* LOC_Os09g38759 (DHNA_09g), *DHNA* LOC_Os08g44210 (DHNA_08g), *ADC synthase* LOC_Os06g48620 (ADC syn), *ADC lyase* LOC_Os02g17330 (ADClya_02g), *ADC lyase* LOC_Os05g15530 (ADClya_05g), *HPPK/DHPS* LOC_Os07g42632 (HPPK/DHPS), *DHFS* LOC_Os12g42870 (DHFS), *DHFR* LOC_Os11g29390 (DHFR_11g), *DHFR* LOC_Os12g26060 (DHFR_12g), *FPGS* LOC_Os03g02030 (FPGS_03g), *FPGS* LOC_Os10g35940 (FPGS_10g) and *GGH* LOC_Os05g44130 (GGH) are showed in vertical bars. Each bar represents the mean of triplicate determination and the vertical line indicates the standard deviation. Asterisks represent the significant difference ($P<0.05$) of transcript abundance compared between knock out and Dongjin wild type within tissues.

4.4 DISCUSSION

4.4.1 Disrupting the *FPGS* Os03g02030 gene affects rice grain filling

Since polyglutamylated folates exist predominantly in most eukaryotes (Shane, 1995; Scott *et al.*, 2000), the polyglutamylation process, catalysed by FPGS enzyme, is essential for cell growth and development. In rice, two putative FPGS genes exist, *FPGS* Os03g02030 and *FPGS* Os10g35940. Transcript abundance profiling (chapter 3) suggested *FPGS* Os03g02030 gene might function predominant in grain; whilst *FPGS* Os10g35940 gene functions mainly in leaf.

To examine the functional importance of rice FPGS genes, T-DNA insertion mutants were characterised. The single T-DNA insertion in the *FPGS* Os03g02030 gene was identified by Southern (Figure 4.8) and PCR (Figure 4.12A) analysis and revealed a homozygous genotype with the total absence of *FPGS* Os03g02030 transcript expression (Figure 4.12B). This mutant was referred to as *fpgs*03g.

Characterisation of the *fpgs*03g mutant revealed that plants lacking *FPGS* Os03g02030 gene product exhibit a delayed seed development phenotype. Seed development was impaired from the heading stage, leading to slow grain filling and late ripening of the mature grains (Figure 4.18B). Molecular profiling of the *fpgs*03g knock out line revealed a decrease in polyglutamylated folates in seed (0.43 µg/100g) compared to the wild type (1.37 µg/100g) (Table 4.2). Hence, a major proportion of polyglutamylation process in seed seems to be dependent on the *FPGS* Os03g02030 gene. This could confirm the importance of the polyglutamylation process catalysed by *FPGS* Os03g02030 gene during seed development. Since embryo development starts at the post-anthesis stage (Counce *et al.*, 2000) and requires a high level of nucleotide synthesis, impaired folate polyglutamylation during seed developing is likely to impact on the rate of DNA synthesis, causing a delay in embryo development. Alternatively, the delayed grain phenotype exhibited by *fpgs03g* may reflect the impact on other folate-dependent processes such as the glycine-serine

interconversion, and methionine biosynthesis (Somerville, 2001; Wingler *et al.*, 1997; Heineke *et al.*, 2001; Moffatt and Weretilnyk, 2001; Gallardo *et al.*, 2002).

However, germination, the developmental stage reported to generate the greatest demand for folates (Jabrin *et al.*, 2003, Gambonnet *et al.*, 2001), was not affected in the *fpgs*03g mutant (Figure 4.18C and D). This suggests that the remaining rice *FPGS* gene Os10g35940 can compensate for the lack of Os03g02030 activity during germination. Alternatively, folate polyglutamylation activity may not be essential during embryo development such that enzymes normally dependent on polyglutamylated folates are able to substitute them with mono-glutamated folate forms. However, these possibilities will require the characterisation of the phenotype of plants lacking both rice *FPGS* genes.

FPGS Os10g35940 transcript abundance appeared to increase only in the *fpgs*03g mutant leaf. This appeared to compensate for any loss of FPGS activity due to disruption of the *FPGS* Os03g02030 gene in the vegetative stage development. Although the *FPGS* Os03g02030 gene was also expressed in the wild type leaf, FPGS activity from *FPGS* Os03g02030 gene may be less important since lack of this gene did not affect the leaf phenotype or polyglutamylated folate levels in the mutant leaf. The *fpgs*03g mutant leaf retained a high level of polyglutamated folate forms (4.02 µg/100g) compared to the wild type leaf (5.39 µg/100g) (Table 4.2). Hence, the *FPGS* Os10g35940 gene appeared to function predominantly in leaf.

In wild type, approximately 40% of rice folate forms detected in leaf and 20% of folate in seed exist either as tetra or penta forms of the 5-CH_3- and 5&10-CHO-H_4Pte (Table 4.2). This result is in agreement with previous studies from various plant species which revealed a significant proportion of plant folates exist in a polyglutamylated form (Storozhenko *et al.*, 2007; Orsomando *et al.*, 2005; Diaz de la Garza *et al.*, 2004). Pteroylpentaglutamate was reported as the most predominant form of polyglutamate in the mitochondria, to provide for

the high affinity of the SHMT enzyme involving in photorespiration (Rébeillé *et al.*, 1994). Between tissues, tetra- and pentaglutamylated folate forms were produced in leaf 74% higher than in grain of Dongjin wild type (Table 4.2). This is likely to be due to the huge demand for the polyglutamylated folate forms as cofactors for several key enzymes especially SHMT and GDC enzymes during the photorespiratory cycle in the mitochondria (Bourguignon *et al.*, 1988; Besson *et al.*, 1993). The absence of a leaf phenotype in the *fpgs*03g mutant suggests that the *FPGS* Os10g35940 gene product may localise and function in the mitochondria.

To examine the above hypothesis, a knock out in the rice *FPGS* Os10g35940 gene is required. If this hypothesis is true, the phenotype of the *fpgs*10g rice mutant should be similar to the discovery in *Arabidopsis* (Payam Mehrshahi, PhD thesis 2008, University of Nottingham) either embryo or seedling lethal. The T-DNA insertion mutant of rice *FPGS* Os10g35940 gene used in this study was shown to be ineffective. This may be due to rearrangement of T-DNA on the chromosome. Gene transformation with T-DNA is a random process; hence, the number of integrated genes is uncontrollable and the position of insertion is unpredictable (Hiei *et al.*, 1997). Possibly, in a range from one to six copies anywhere within the genome could happen. Other more specific approaches such as RNA silencing are suggested to generate a knocked out mutant of this gene.

LC-MS/MS metabolite profiling in this study revealed that 5-formylpentaglutamylated folate was the most abundance form of polyglutamylated folates in Dongjin and the *fpgs*03g mutant leaves. Basically, methyl derivatives of folate account for 45-65% of the total folate level in higher plants (Cossins, 2000) mostly in the cytosol fraction, whilst the formyl derivatives are produced mostly in organelles with 44% of the 5-formyl form dominant in mitochondria (Chen *et al.*, 1997; Chan and Cossins, 2003). This finding is similar to observations in rat liver (Horne *et al.*, 1989) and pancreas mitochondria (Horne and Holloway, 1997). 5-formyl substituted derivative have been suggested to be the most stable folate derivative which are not

directly involve in any pathway of one-carbon metabolism (Stover and Schirch, 1992). It could be a storage form (Chen *et al.*, 1997) or a regulatory factor of the photorespiration process through the inhibition of mitochondrial SHMT enzyme (Roje *et al.*, 2002). The above reasons might explain the high levels of 5-formylpentaglutamylated folate in rice leaves. However, folate derivatives could be interchanged into various forms of methylene THF, methenyl THF and 10-formyl THF (Jabrin *et al.*, 2003) *in vivo*. These interchange reactions are catalysed by methylenetetrahydrofolate dehydrogenase, methenyltetrahydrofolate cyclohydrolase and methylene tetrahydrofolate reductase enzymes which can be found in one-carbon metabolism in all compartments, the cytosol, mitochondria and chloroplast (Kirk *et al.*, 1995; Neuburger *et al.*, 1996). Hence, even though only a few monoglutamate and polyglutamate folate forms are predominant in the mature leaf and the mature seed, during development they could be changed into other folate derivatives by enzymes in one-carbon metabolism.

In summary, single *fpgs*03g gene knock out plant revealed a role for the *FPGS* Os03g02030 gene during seed development, whilst the *FPGS* Os10g35940 gene appears to function mainly during leaf development. Although lacking one FPGS enzyme encoded by the *FPGS* Os03g02030 gene, rice leaf tends to compensate by increasing the mRNA level of another *FPGS* gene in order to maintain the normal leaf phenotype and provide sufficient amount of cofactors for enzymes in one-carbon metabolism. In the mutant leaf, polyglutamated folate forms account for 46% of the total folate level which is no different to 44% in wild type leaf. In contrast, polyglutamated folate forms in the mutant grain account for only 10% of the total folate level, compared to the wild type grain of about 21%, together with the impairment of seed setting stages. Hence, these observations strongly suggest that lacking polyglutamylation from the *FPGS* Os03g02030 product causes defect in the grain filling and reduces polyglutamylated folate in grain.

4.2.2 Folate biosynthesis gene expression in rice is subject to feedback regulation in folate homeostasis

Metabolic and expression profiling from this study suggested folate polyglutamylation is involved in retention and homeostasis of folate pool. Same conclusion has been previously reported in a number of organisms (Desouza *et al.*, 2000; Lin *et al.*, 1993; Lin and Shane, 1994). LC-MS/MS folate profiling of the *fpgs*03g mutant revealed a reduction in folate abundance (Table 4.2). Considering the transcript levels of all key biosynthesis genes in leaf and grain (Figure 4.19), *GTPCHI, ADC synthase, ADC lyase* and *HPPK/DHPS* mRNA levels increased significantly in the *fpgs*03g mutant. A correlation between the reduction in folate abundance in tissues of the *fpgs*03g mutant and an increased expression of a number of folate biosynthesis genes was proposed. Mutant plants appear to compensate for reduced polyglutamylation by increasing its substrate, monoglutamate folates, through increasing mRNA levels of folate synthesis genes. However, little has been reported about the mechanism of regulation in the plant folate biosynthesis pathway. The HPPK/DHPS enzyme was suggested to be a key regulatory part of the entire folate biosynthetic pathway as DHPS enzyme is strongly feedback-inhibited by dihydropteroate and also dihydrofolate and tetrahydrofolate monoglutamate (Mouillon *et al.*, 2002). This is in agreement with this study which showed that the expression of this enzyme significantly increased in the *fpgs*03g mutant to maintain the monoglutamate folate pool. These results can confirm the observations from previous studies that the expression of folate biosynthesis enzymes correlates with the folate pool and the one-carbon unit demands in various physiological situations (Basset *et al.*, 2002, 2004a, 2004b; Jabrin *et al.*, 2003).

However, the reduction in monoglutamate folates (Table 4.2) may be a result of reduction of folate breakdown, as polyglutamylated protein-bound folates are less susceptible to oxidative breakdown (Suh *et al.*, 2001). Interestingly, the reduction in *GGH* transcript levels in the *fpgs*03g knockout plant might result in the reduction of folate breakdown with only small changes in

polyglutamylated folate levels in the *fpgs*03g leaf, presumably due to an increase in the expression of the second rice FPGS isoform (Os10g35940) (Figure 4.19B). This finding may indicate a mechanism of folate homeostasis in leaf tissue. Disruption of *FPGS* Os03g02030 gene being compensated for another functional rice FPGS was showed together with the reduction of the vacuolar folate deglutamylation in maintaining the cellular polyglutamylated folate pools.

This chapter only provided the preliminary data in the control of folate synthesis in rice system. Future studies are suggested to focus on determining the molecular basis of this regulatory mechanism which will aid manipulation of folate abundance in plants. The exact mechanism of regulation is needed to manipulate to confirm involvement of other transcriptional and translational regulatory factors.

Finally, metabolic engineering is a further aspect. Loss of GGH activity may reduce the folate turnover rate and lead to an increase of cellular polyglutamylated folate pool. Taken with the function of *FPGS* genes revealed in this chapter could support selecting the *FPGS* Os03g02030 gene in overexpression study in rice grain. Introducing this gene which plays the key role during rice seed development might be more powerful than overexpression another gene which expressed mostly in rice leaf in terms of increasing bioavailability and stability of folate directly in rice endosperm.

CHAPTER 5 TRANSGENIC MANIPULATION OF FOLATE POLYGLUTAMYLATION

5.1 INTRODUCTION

Metabolic engineering of biosynthetic pathways provides a strategy to increase micronutrient levels in plants in order to improve food quality. This strategy has recently been used to improve folate levels in rice by over 100 fold (Storozhenko *et al*, 2007). However, the impact of this increase in terms of bioavailability and stability of folate in grains is unclear. In this chapter, two types of folate binding proteins were expressed in rice to help stabilise folate and decrease its oxidation rate: cow's milk folate binding protein and rat liver glycine N-methyltransferase enzyme. In parallel, two rice *FPGS* genes (chapter 4) were over-expressed in an attempt to enhance retention of folates in plant cells as polyglutamylated folates.

5.1.1 Folate binding proteins

Folate binding proteins (FBP) have been described mostly in mammalian cells (Henderson, 1990). FBPs have been identified from various tissues including placenta, liver and extracellular fluids. FBP plays an important role in the assimilation, distribution and retention of folate in cells throughout the body (Henderson, 1990). FBPs are divided into 3 major groups; high-affinity; membrane-associated; and cytoplasmic FBPs. These three groups differ in folate affinity and functions. High-affinity FBP represents a major class having a high preference for various forms of reduced-folate compounds. The second major class are membrane-associated FBPs which are different to the first class in structure but have a lower folate affinity. Their main function is to transport folate across cell membranes. The last class contains cytoplasmic-binding proteins which have high affinity for only some specific reduced-folate

compounds (Henderson, 1990). In this thesis, only high-affinity FBPs will be reviewed and considered.

High-affinity FBPs exist as two main forms, membrane-bound (mFBPs) and soluble (sFBPs) (Elnakat and Ratnam, 2004). Membrane-bound forms are found in carcinoma cell culture, placenta, proximal kidney tubules, choroid plexus, small intestine and liver. mFBPs accumulate primarily on the cell surface (Lacey *et al.*, 1989). Their C-terminal sequences contain 31 hydrophobic amino acids that integrate those proteins into cell membranes (Lacey *et al.*, 1989). mFBPs play an important role in the transport of folate into cells with a slow transport, but high turnover, rate (Kamen and Capdevila, 1986). Interestingly, when the 31 hydrophobic amino acids at the C-terminal of the proteins are removed, they function as soluble FBP forms (Antony *et al.*, 1981; Antony *et al.*, 1989).

Soluble FBP forms (sFBPs) are abundant in human, bovine and goat milk and umbilical cord serum, but normal human serum contains only low levels (Waxman and Schreiber, 1973). sFBP in milk is encoded by the same gene as mFBP but without a signal sequence (Svendsen *et al.*, 1984). The signal sequence of mFBP is cleaved during the solubilization step, converting it to the soluble form (Elwood 1989). Placenta mFBP showed 100% protein homology to human milk FBP (Sadasivan *et al.*, 1994, Svendsen *et al.*, 1982).

In milk, 5-methyltetrahydrofolate is the principal molecule to bind milk FBP. Approximately 60-100% of total milk folate is in the bound form (Mason and Selhub, 1988). This helps protect bound-folate complexes from oxidation during digestion. It has also been reported that FBP increases folate absorption in the gut (Jones *et al.*, 2003). Nevertheless, bound folates are released temporarily from FBP by the acidic condition in the stomach then rebind in the basic condition of the small intestine (Henderson, 1990). The resistance to proteolysis of the pancreatic enzyme of FBP from goat milk resulted in its survival along the length of the small intestine (Salter and Mowlem, 1983). The absorption of bound folate and free folate are crucial to analyze in order to

examine the bioavailability of the bound-folate form. From several studies, the bound-folate form is absorbed in the rat small intestine twice as fast as free folate (Colman *et al.*, 1981) and it occurs mainly in the ileum while free forms are absorbed more readily in the jejunum (Mason and Selhub, 1988). In neonates, bound folate can be absorbed via an enterocytotic process, rather than through the dissociation or degradation of FBP. FBP is subject to proteolytic digestion inside enterocytes to release bound folate for export into the portal circulation (Mason and Selhub, 1988).

5.1.2 Cow's milk folate binding protein (cFBP)

cFBP mediates delivery of 5-methyltetrahydrofolate to the interior of intestinal cells during folate absorption. Milk is responsible for 10-15% of the daily folate intake especially in high milk consumption countries (Verwei *et al*, 2004). In suckling animals, this protein regulates folate bioavailability since milk is the only dietary source (Henderson, 1990).

cFBP has high binding affinity for folate molecules. The equilibrium dissociation constant (K_D) of cFBP, measured *in vitro* using the equilibrium-dialysis assay, was reported to be 3 nM for folate and 5 nM for 5-methyltetrahydrofolate at pH 7.2 at 37°C (Salter, 1981). From an *in vitro* gastrointestinal model, cFBP could bind up to 79% of 5-methyltetrahydrofolate monoglutamate in pH7 and pH3 condition (Verwei *et al*, 2004). Recently, Nygren-Babol *et al.* (2005) measured the K_D of cFBP at pH 7.4 using surface plasmon resonance technology (SPR, Biacore) and reported that folic acid and 5-methyltetrahydrofolate have the highest affinities for FBP at approximately 20 and 160 pmol/L, respectively. Thus, this binding protein has very high binding affinity for the major folate form, 5-methyltetrahydrofolate monoglutamate.

The native molecule of cFBP is a 222-amino acid-monomer consisting of 8 disulfide bonds with N-terminal carbohydrate residues (Salter, 1981). The molecular weight on SDS-PAGE is between 30 and 35 kDa depending on the degree of glycosylation (Salter *et al.*, 1981). Even though no crystal structure is

available, functional domains can be observed based on the amino acid sequence. Sequence alignments revealed that the "AKHHK" region is conserved in mammalian FBPs (Figure 5.1). This has been proposed to have a folate binding function as the motif could form a pocket, with two lysine (K) residues binding to the glutamic acid moieties of folate molecules (Sadasivan *et al.*, 2000).

```
CLUSTAL W (1.83) multiple sequence alignment

FOLR1_BOVIN_FR-alpha_       MAWQMT-QLLLLALVAAAWGAQAPRTPRARTDLLNVCMDAKHHKAEPGPE 49
FOLR1_HUMAN_FR-alpha_       MAQRMTTQLLLLLVWVAVVGEAQTRIAWARTELLNVCMNAKHHKEKPGPE 50
FOLR2_HUMAN_FR-beta_        MVWKWMPLLLLLVCVATMCSAQD------RTDLLNVCMDAKHHKTKPGPE 44
                            *. :    ****   .:  .          **:*******:*****  :****

FOLR1_BOVIN_FR-alpha_       DSLHEQCSPWRKNACCSVNTSIEAHKDISYLYRFNWDHCGKMEPACKRHF 99
FOLR1_HUMAN_FR-alpha_       DKLHEQCRPWRKNACCSTNTSQEAHKDVSYLYRFNWNHCGEMAPACKRHF 100
FOLR2_HUMAN_FR-beta_        DKLHDQCSPWKKNACCTASTSQELHKDTSRLYNFNWDHCGKMEPACKRHF 94
                            *.**:** **:*****:..** * *** * **.***:***:* *******

FOLR1_BOVIN_FR-alpha_       IQDTCLYECSPNLGPWIREVNQRWRKERVLGVPLCKEDCQSWWEDCRTSY 149
FOLR1_HUMAN_FR-alpha_       IQDTCLYECSPNLGPWIQQVDQSWRKERVLNVPLCKEDCEQWWEDCRTSY 150
FOLR2_HUMAN_FR-beta_        IQDTCLYECSPNLGPWIQQVNQSWRKERFLDVPLCKEDCQRWWEDCHTSH 144
                            ****************::*:* *****.*.********: *****:**:

FOLR1_BOVIN_FR-alpha_       TCKSNWHKGWNWTSGYNQCPVKAAHCRFDFYFPTPAALCNEIWSHSYKVS 199
FOLR1_HUMAN_FR-alpha_       TCKSNWHKGWNWTSGFNKCAVGAACQPFHFYFPTPTVLCNEIWTHSYKVS 200
FOLR2_HUMAN_FR-beta_        TCKSNWHRGWDWTSGVNKCPAGALCRTFESYFPTPAALCEGLWSHSYKVS 194
                            *******:**:**** *:*.. *    *. *****:.**: :*:******

FOLR1_BOVIN_FR-alpha_       NYSRGSGRCIQMWFDPFQGNPNEEVARFYAEN--PTSGSTPQGI------ 241
FOLR1_HUMAN_FR-alpha_       NYSRGSGRCIQMWFDPAQGNPNEEVARFYAA---AMSGAGPWAAWPFLLS 247
FOLR2_HUMAN_FR-beta_        NYSRGSGRCIQMWFDSAQGNPNEEVARFYAAAMHVNAGEMLHGTGGLLLS 244
                            ***************. ************      :*    .

FOLR1_BOVIN_FR-alpha_       ----------
FOLR1_HUMAN_FR-alpha_       LALML-LWLLS 257
FOLR2_HUMAN_FR-beta_        LALMLQLWLLG 255
```

Figure 5.1 Alignment of cFBP amino acid sequences from human and bovine

The amino acids of mature cFBP were shown. The conserved sequence for folate binding site was shown in the box. The sequence for mature cFBP was 241 bp in length different from human cFBP in the C-terminal. The bovine FBP has 75% amino acids homology with human FBPs.

5.1.3 Rat liver's glycine N-methyltransferase (GNMT)

GNMT is another folate binding protein which has high affinity for folate compounds. Normally, this enzyme plays a role in the methylation process to

catalyze the methylation of glycine using S-adenosylmethionine (SAM) to form sarcocine and S-adenosylhomocysteine (SAH) (Blumenstein and Williams, 1960) (Figure 5.2).

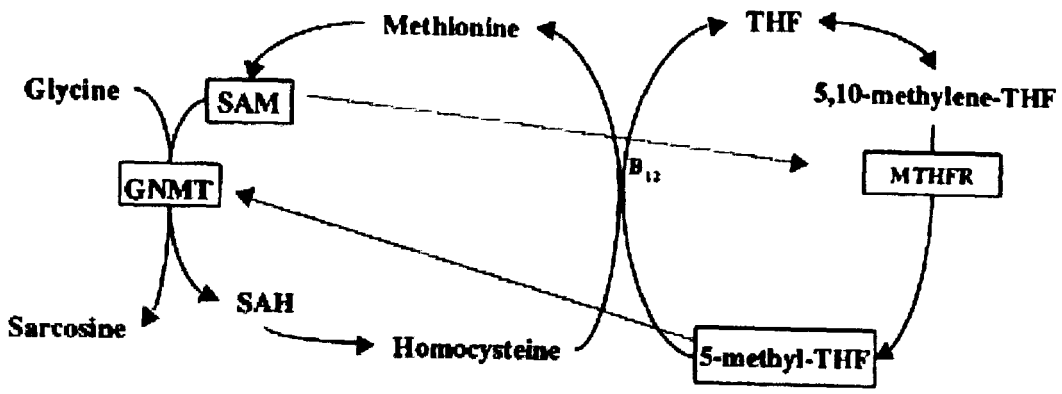


Figure 5.2 GNMT related metabolism

The GNMT enzyme is required in the transmethylation reaction using folate derivatives synthesised from folate metabolism. MTHFR is a methylenetetrahydrofolate reductase enzyme required in methionine synthesis which converts 5,10-methylene-tetrahydrofolate (THF) to 5-methyl-THF. This picture was taken from Rowling and Schalinske (2001).

The function of the GNMT enzyme can be inhibited by 5-methyltetrahydrofolate pentaglutamate (Wagner *et al.*, 1985). 5-Methyltetrahydrofolate pentaglutamate binds strongly to this enzyme with K_D of 0.13 μM (Yeo and Wagner, 1992). However, the binding affinity of GNMT to other folate derivatives is still unknown.

In addition, its binding site for folate is distinct from SAM and SAH. The folate binding site is in the large central hole of the tetrameric molecule detected by crystallography of recombinant GNMT (Pattanayek *et al.*, 1998) (Figure 5.3). Overall, the structure of GNMTs from rat, mouse and human are very similar and has the conserved region for folate binding (Luka *et al.*, 2007). The molecular weight of each GNMT subunit is approximately 31-33 kDa based on SDS-PAGE, whilst, based on gel filtration the native molecule is

130,000 – 132,000 kDa (Ogawa and Fujioka, 1982; for review see Ogawa *et al.*, 1998).

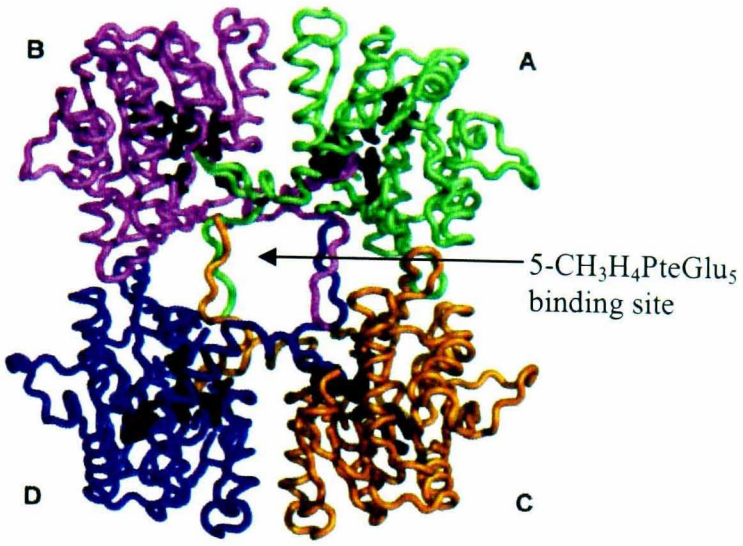


Figure 5.3 GNMT structure

GNMT consists of four identical subunits; A; B; C and D. The central hole of the molecule is the binding site of 5-methyltetrahydrofolate pentaglutamate which is an enzyme inhibitor (Luka *et al.*, 2007).

5.2 AIMS AND OBJECTIVES

This chapter attempts to enhance folate bioavailability by expressing cow's milk folate binding protein (cFBP) and rat liver's glycine N-methyltransferase (GNMT), in rice endosperm.

However, prior to expressing two folate binding proteins in rice, the binding affinity to various forms of folate derivatives was determined as previous studies reported only their binding affinity to folic acid and 5-methyltetrahydrofolate monoglutamate, but not polyglutamated folate forms.

The first part of this chapter describes synthesizing recombinant folate binding proteins in *E.coli* to perform binding affinity assays. The second part describes ongoing studies to engineer rice to express FPGS and FBP (cFBP and GNMT) in rice endosperm.

5.3 RESULTS AND DISCUSSION

PRODUCTION OF RECOMBINANT FOLATE BINDING PROTEINS

In an attempt to determine the binding affinity of truncated cFBP and GNMT to polyglutamylated folate molecules, an *E.coli* expression system was selected to synthesize recombinant proteins.

cFBP and GNMT sequences were first engineered to remove signal peptides in their expressed sequences. Figure 5.4 and 5.5 shows the complete cDNA sequences of cFBP and GNMT, respectively, indicating where the first amino acid starts. These nucleotide sequences were the templates used to amplify cDNA fragments. The cFBP recombinant protein started at 20th codon until the end of the cDNA sequence since the first 19 codons encoded the signal sequence. In contrast, GNMT has no signal sequence. The amino acid sequences of those two proteins are shown in Figure 5.6 and 5.7. Both recombinant proteins were translated until reaching their normal stop codon.

5.3.1 Amplification of cFBP and GNMT coding sequences

Total RNAs, used for amplification of cFBP and GNMT coding sequences, were extracted from cow udder and rat liver tissue, respectively. cDNAs without membrane bound signal sequence were amplified using three primer pairs (as shown in appendix 1) to generate attB1 (5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTC3') and attB2 (5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCTA3') gateway compatible sites for each gene. For the details of templates and PCR condition, see section 2.3.5 in chapter 2. A single band of the PCR product corresponding to *cFBP* and *GNMT* genes was obtained as shown in Figure 5.8. DNA sequencing was then performed on both genes to confirm the sequences.

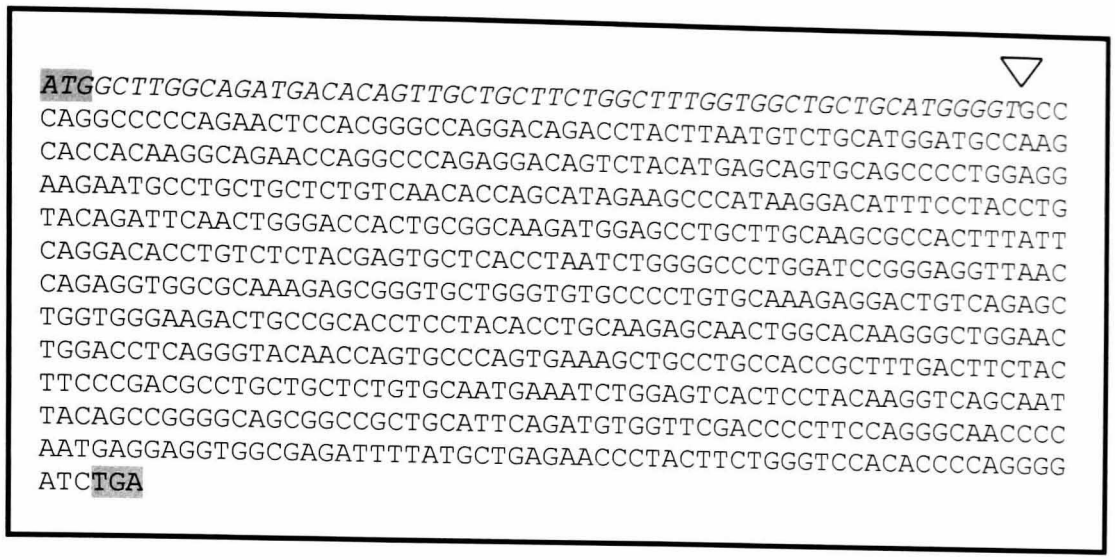


Figure 5.4 The cFBP complete cDNA sequence

The nucleotide sequence of cFBP was shown and its signal sequence for this protein was shown in italic. The sequence was 669 bp in length. The triangle indicates the first nucleotide sequence selected for protein translation starting point. Grey-highlighted letters are start (ATG) and stop (TGA) codons.

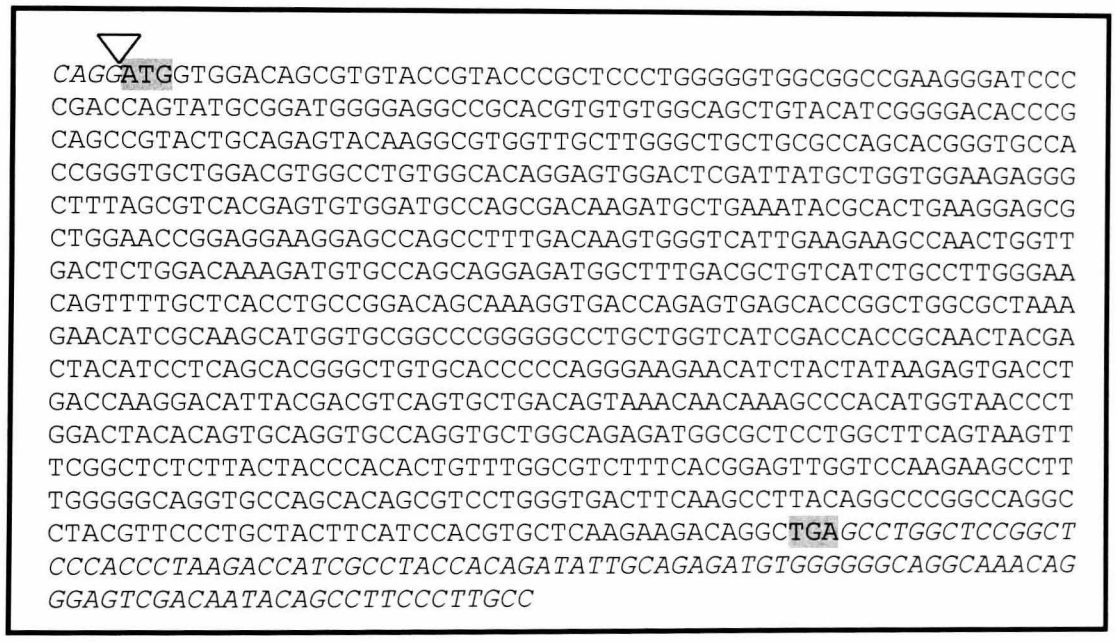


Figure 5.5 The GNMT complete cDNA sequence

The nucleotide sequence of GNMT was shown and non-coding sequence was shown in italic. The sequence was 882 bp in length. The triangle indicates the first nucleotide sequence selected for protein translation starting point. Grey-highlighted letters are start (ATG) and stop (TGA) codons.

```
MAWQMTQLLLLALVAAAWGAQAPRTPRARTDLLNVCMDAKHHKAEPGPEDSLHEQCSP
WRKNACCSVNTSIEAHKDISYLYRFNWDHCGKMEPACKRHFIQDTCLYECSPNLGPWI
REVNQRWRKERVLGVPLCKEDCQSWWEDCRTSYTCKSNWHKGWNWTSGYNQCPVKAAC
HRFDFYFPTPAALCNEIWSHSYKVSNYSRGSGRCIQMWFDPFQGNPNEEVARFYAENP
TSGSTPQGI*
```

Figure 5.6 The amino acid sequence of cFBP

The grey-highlighted letters show the signal sequence on this protein and the normal black letters shows the mature protein sequence.

```
MVDSVYRTRSLGVAAEGIPDQYADGEAARVWQLYIGDTRSRTAEYKAWLLGLLRQHGCHRVLDVACGTG
VDSIMLVEEGFSVTSVDASDKMLKYALKERWNRRKEPAFDKWVIEEANWLTLDKDVPAGDGFDAVICLG
NSFAHLPDSKGDQSEHRLALKNIASMVRPGGLLVIDHRNYDYILSTGCAPPGKNIYYKSDLTKDITTSV
LTVNNKAHMVTLDYTVQVPGAGRDGAPGFSKFRLSYYPHCLASFTELVQEAFGGRCQHSVLGDFKPYRP
GQAYVPCYFIHVLKKTG*
```

Figure 5.7 The amino acid sequence of GNMT

The whole amino acid sequence of GNMT is shown. There is no signal sequence predicted in this sequence.

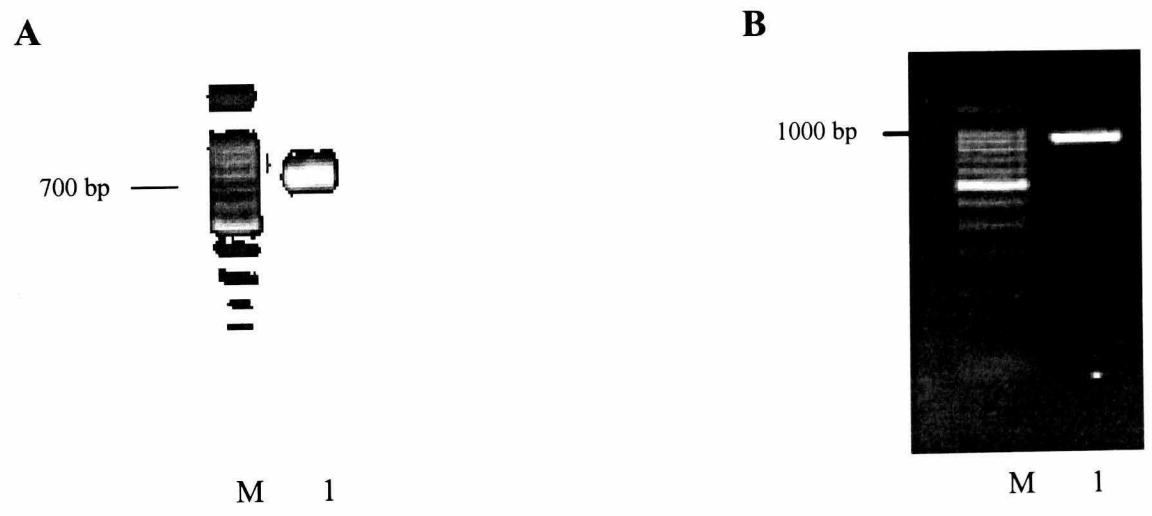


Figure 5.8 Amplification of the cFBP and GNMT fragments by RT-polymerase chain reaction (PCR)

A: The PCR product of cFBP (lane 1). The expected band size of cFBP was 662 bp. B: The PCR product of GNMT (lane 1). The expected band size of GNMT was 882 bp. They were purified and separated alongside the size markers (M; 100 bp ladder) on a 1% agarose gel containing ethidium bromide.

5.3.2 Generating the expression vector

The *E.coli* expression vector, pDEST17, was used for cFBP and GNMT recombinant protein production (see the vector map in appendix 2) (Figure 5.9). This vector contains a T7 promoter which leads to the synthesis of a large amount of mRNA and the accumulation of expressed proteins in the bacterial cell (Baneyx, 1999). This vector also contains an N-terminal polyhistidine (6xHis) fusion tag allowing convenient purification using a single step of metal affinity chromatography. The pDEST vector is also compatible for the Gateway cloning approach. The site-specific recombination can occur between attR sites of the destination vector (pDEST) and attL sites of the entry vector containing the gene of interest via the simple LR reaction (Gateway manual, Invitrogen). These sites are the universal sites on the Gateway vectors. This provides a rapid and efficient way to move the target gene directionally into multiple vector systems (Earley *et al.*, 2006). This technology leads to the increase of cloning efficiency and the specific gene orientation. No restriction digestion is required as in some cases the genes of interest contain almost all common restriction sites which is difficult to clone via traditional cloning approaches.

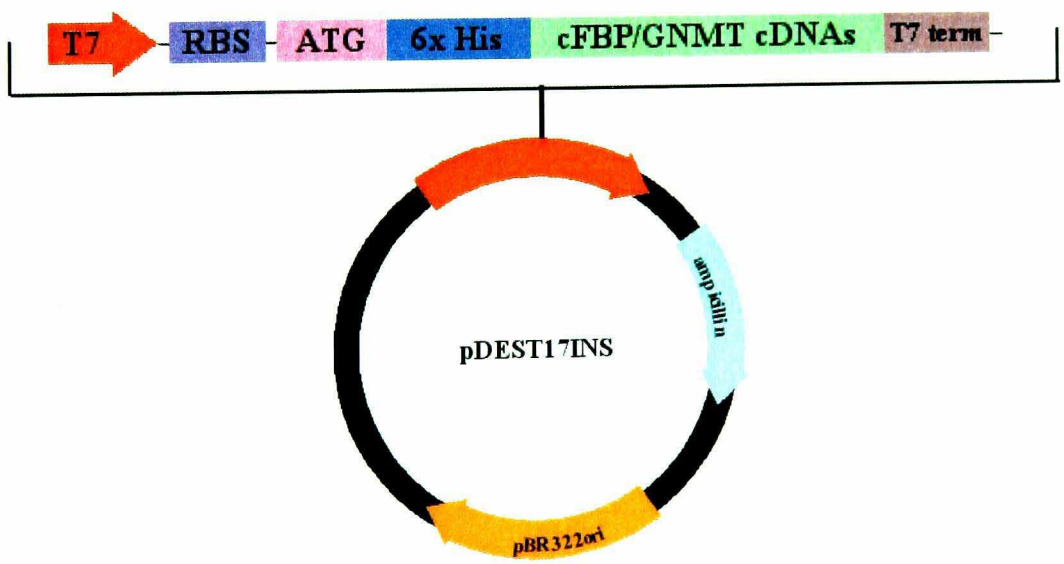


Figure 5.9 pDEST17 containing cFBP or GNMT cDNAs

5.3.3 Optimization of the protein expression condition

The choice of host strain is another factor affecting protein expression efficiency. cFBP and GNMT clones were transformed into Origami B (DE3) *E.coli* cells. This strain carries a mutation in the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes which greatly enhances the formation of disulfide bonds in the cytoplasm (Bassette *et al.*, 1999) because in *E.coli* cells, cysteine residues are normally reduced by thioredoxin reductase and glutaredoxin (Sørensen and Mortensen, 2005). cFBP is rich in disulfide bonds containing at least 8 disulfide bonds per molecule (Sadasivan *et al.*, 2000).

As recombinant proteins over-expressed in the cytoplasm of *E.coli* mostly accumulate as insoluble aggregates known as inclusion bodies, a refolding step is required to obtain soluble folded proteins. Inclusion bodies can be solubilized by strong detergents such as urea and guanidium hydrochloride which denatures the protein structure. Dialysis, dilution or on-column refolding methods are required to gain soluble protein with biological function. However, there is a possibility that the resolubilization methods can affect the integrity of refolding proteins and lead to a defect in protein structure and function. The recovery yield is also poor (for review, see Baneyx, 1999; Sørensen and Mortensen, 2005). Hence, the optimization of cultivation temperature and the concentration of IPTG were performed in an attempt to reduce protein aggregation and gain more soluble protein.

The amount of soluble recombinant protein secreted by *E.coli* was highly dependent on the cultivation temperature. Reducing the cultivation temperature is a well known technique to limit aggregation of proteins (Schein, 1989). It was effective in solubility improvement of several recombinant proteins such as human interferon, bacterial luciferase and rice lipoxygenase using the *cspA* promoter system (Vasina and Baneyx, 1997).

In small scale synthesis of both FBPs, the induction for gene expression was carried out in a 5 ml culture at a final concentration of 0.1 mM and 0.5 mM IPTG for 24 hours. Two different temperatures; 20°C and 37°C were tested.

For the induction at 20°C, culture pellets were collected at 3 and 24 hours after induction as the bacteria growth rate is slow at low temperature. For the induction at 37°C, the culture pellets were collected at 1 and 6 hours after induction. All pellets were analyzed by SDS-PAGE and proteins were detected by western blotting using an antibody recognising the 6x His tag. Figure 5.10 shows the western blot results indicating the best condition for protein expression. The most suitable condition discovered to express soluble cFBP was at 20°C with 0.1 mM IPTG (Figure 5.10A). At this temperature the maximal recombinant cFBP synthesis was obtained after 24 hours cultivation (Figure 5.10A), whereas induction at 37°C did not provide any soluble protein (as no specific bands detected by western blot; data not shown). In contrast, the best condition for GNMT was induction for 6 hours at 37°C with 0.1 mM IPTG (Figure 5.10B). However, during cell culture, it was observed that at 20°C, the bacteria cells grow slowly due to a decrease in replication, transcription and translation rate of cells (Shaw and Ingraham, 1967). Hence, when the lower temperature is used, a longer period of cultivation is required to increase the biomass of bacteria. Nevertheless, it is not guaranteed that reducing the cultivation temperature could increase the solubility of all proteins.

The Origami B (DE3) strain is derived from a *lacZY* mutant of BL21 strain containing a deletion in the *lacZY* which allows the precise control of the expression rate of recombinant proteins with isopropyl-beta-D-thiogalactopyranoside (IPTG). Hence, the concentration of IPTG was optimized to obtain maximal protein expression. In the trial experiment, 1 mM and 5 mM were used but the cells were died (data not shown). Hence, only 0.1 mM and 0.5 mM were considered. However, using 0.5 mM IPTG made cells look unhealthy and they died very soon. Only 0.1 mM IPTG could be used and soluble protein produced.

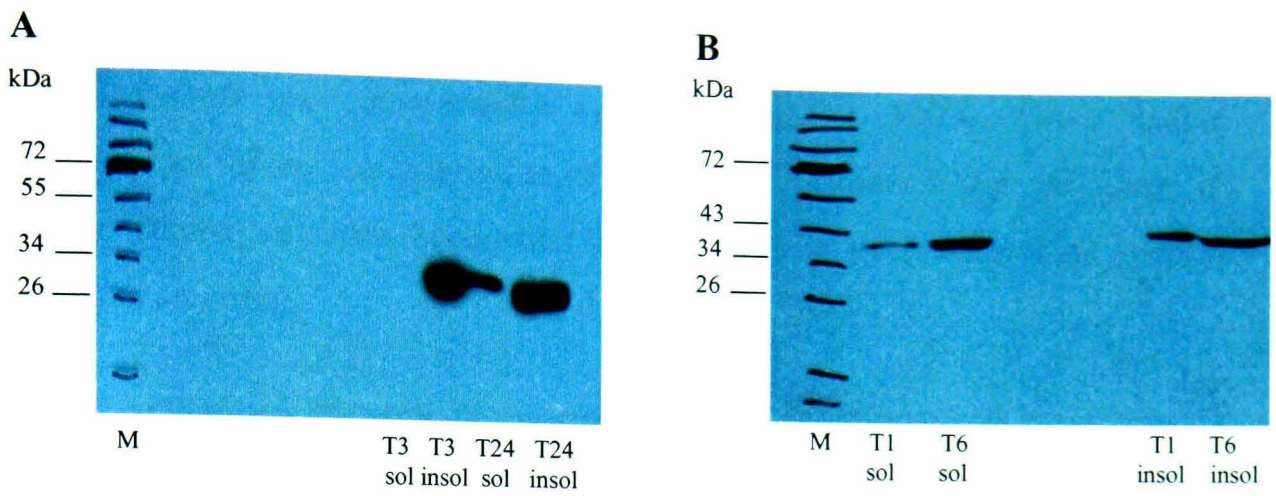


Figure 5.10 Western blot analysis for cFBP and GNMT

cFBP expression induced at 20°C with 0.1 mM IPTG provided soluble protein 24 hours after induction (A), while, GNMT expressed at 37°C with 0.1 mM IPTG resulted in soluble protein 6 hours after induction (B). M represents protein marker; T1, T3, T6 and T24 represents samples at 1, 3, 6 and 24 hours after induction; sol represents soluble part and insol represents insoluble part. Western blots were using anti-6xHis polyclonal antibody.

Recombinant expression of mouse, rat and human GNMTs succeeded in the pET expression vector by induction using 1 mM IPTG at 20 °C for 17-18 hours (Luka and Wagner, 2003). The authors showed that expressing GNMT at a low temperature for a longer period provided soluble proteins, but not at 37°C, which is in contrast to the current experiment. However, their expression condition resulted in 20-40 mg of recombinant GNMTs in a litre of culture, a yield which was much higher than the current experiment. It might also depend on the ability to express proteins in a pET vector and BL21 (DE3) *E.coli* strain which differ from the Origami B strain. The higher concentration of IPTG used by Luka and Wagner (2003) may be another case in induction of protein expression which BL21 (DE3) cells might have higher resistance to high IPTG concentration unlike the Origami B cell.

Besides the variation of the host strains, the expression vector, the induction temperature and IPTG concentration used in expression of soluble recombinant proteins, other strategies can be applied to obtain a higher yield. The target proteins can be further engineered to have fusion partner proteins. Maltose-binding protein (MBP), thioredoxin and glutathione S-transferase have been

widely used as a fusion protein in the production of many proteins and these fusion vectors are commercially available (Baneyx, 1999). These fusion proteins can induce the expression of the soluble proteins in the cytoplasm of *E.coli*. Moreover, other expression systems such as yeast expression system could be an option.

5.3.4 Purification of intracellular proteins

Ni-NTA agarose was used to purify 6xHis-tagged FBP recombinant proteins expressed in *E.coli* (Invitrogen, UK) (see section 2.5.6 in chapter 2 for the purification protocol).

After SDS-PAGE and western blot analyses, all positive western blot fractions were pooled and concentrated into a 1 ml volume using an Amicon Ultra-4 filter device with a 10 kD cut off (Millipore, UK). Although the Ni-NTA column is an affinity column, there were still some contaminating proteins (Figure 5.11A). However, the purification condition was already optimized by lowering the pH of the wash buffer to pH 4 and 6 (normally used at pH 8) and increasing the concentration of imidazole in the wash and elution buffers as well as using a new Ni-NTA column to further purify the eluate. However, this still could not improve purity. Figure 5.11A shows the Coomassie blue staining gel of the pool of GNMT fraction. On the SDS gel after purification, the size of the subunit was confirmed to be 34 kDa by western blotting (Figure 5.11B). Total soluble protein obtained was ~1-2 mg per litre of cell culture.

Using the same methods as GNMT, cFBP was also purified and all fractions that gave signals using western detection were pooled and concentrated into a 1 ml volume. After purification, cFBP exhibited a size of 28 kDa on SDS gel (Figure 5.12). Total soluble protein obtained was ~0.5-1 mg per litre of cell culture. Both purified cFBP and GNMT recombinant proteins can be used in further testing of folate binding affinity.

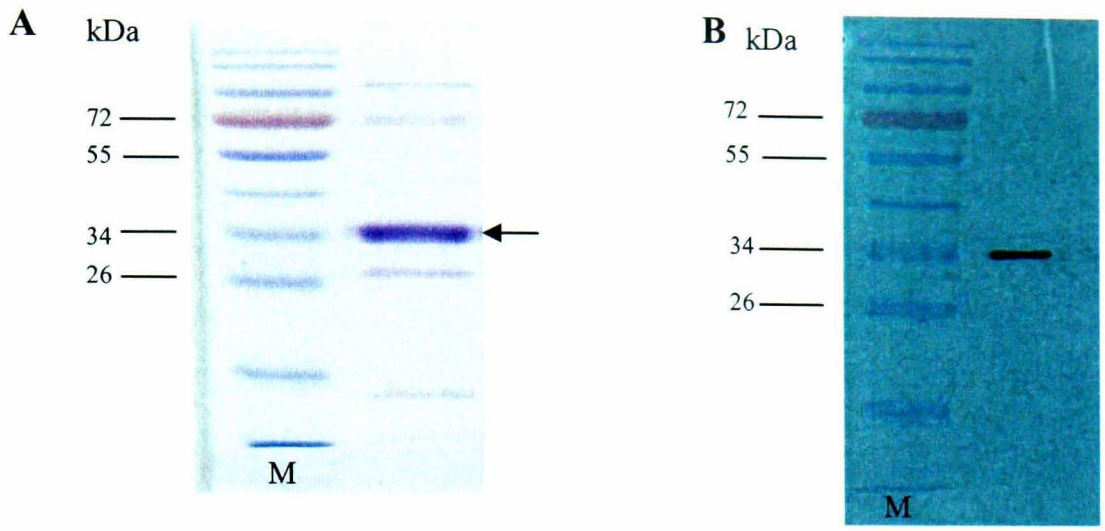


Figure 5.11 Purification of GNMT

Eluted fractions were analyzed by Coomassie blue stained SDS-PAGE (A) and western blot using anti-GNMT (B). Arrows indicate 34 kDa band size of GNMT. M represents the protein molecular weight marker. Numbers denote the kDa values of marker proteins in lane M.

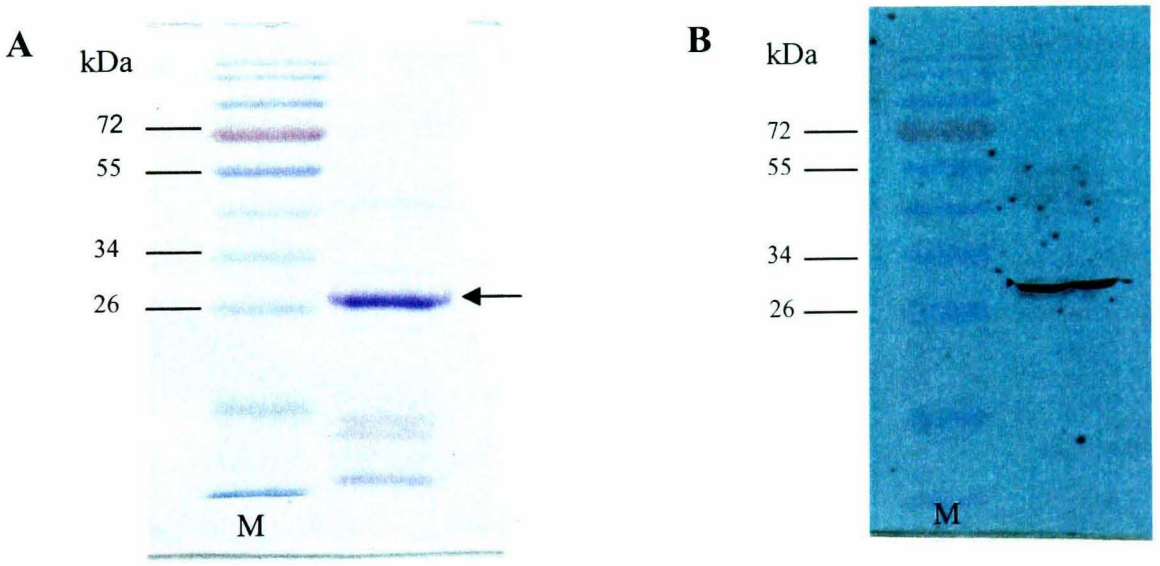


Figure 5.12 Purification of cFBP

Eluted fractions were analyzed by Coomassie blue stained SDS-PAGE (A) and western blot using anti-cFBP (B). Arrows indicate approximately 28 kDa band size of cFBP. M represents the protein molecular weight marker. Numbers denote the kDa values of marker proteins in lane M.

Even though the purification condition was optimized by lowering the pH of the wash buffer and increasing the imidazole concentration in wash and elution buffers in order to increase stringency, some contaminating proteins still remained in the eluate fractions. On SDS gel, some minor bands were observed besides the major bands of cFBP and GNMT (Figure 5.11A and 5.12A). Those proteins might be other His-rich proteins in the cell lysate which can attach strongly on the Ni column. However, those bands account for less than 50% of total proteins compared to the major protein bands. Hence, the possibility of interference on the folate binding assay of those contaminated proteins would be negligible.

Although the Ni-column is an affinity column, other purification methods might be required. From previous reports, rat recombinant GNMT was produced and purified to homogeneity using at least two steps of purification. Ogawa *et al* (1997) purified rat GNMT by DEAE-cellulose column followed by ion-exchange chromatography on DE-52 column. Luka and Wagner (2003) also purified rat GNMT by ammonium sulphate precipitation and ion-exchange chromatography. In this chapter, ammonium sulphate precipitation was firstly tried before applying the lysate onto affinity column. There was no different result from applying the cell lysate directly onto the Ni-column (data not shown). Hence, it seems that ion-exchange chromatography is the method of choice to purify recombinant GNMT. It can be suggested that the ion-exchange column could be used in the last step of GNMT purification. For cFBP purification, there is no report of recombinant cFBP production and purification. Hence, the experiment in this chapter was the first trial of recombinant production of cFBP.

Although the recombinant cFBP and GNMT were partially purified, western blot analysis using antibody against histidine tag and against cFBP or GNMT confirmed that the majority of the protein in the purified portions was represented by these proteins (Figure 5.11B and 5.12B).

Comparing the subunit mass of GNMT produced in this chapter to previous reports, and calculation using their deduced amino acid sequence, showed they were slightly different. One reason is both recombinant GNMT and cFBP contained a 6xHis tag at the N-terminal which would increase the molecular mass approximately 2.6 kDa, accounting for the differences in the observed molecular weight.

The calculation of the peptide mass using the PeptideMass software available at the Expasy website (www.expasy.ch) of recombinant rat GNMT was 33 kDa. From the previous reports of production of recombinant rat GNMT, its subunit mass was 32 kDa when expressed in *E.coli* (Ogawa *et al.*, 1997; Luka and Wagner, 2003). The slightly differences between various experiments might be because either the differences in the amino acid composition at the N- or C-terminal or post-translational modification of expressed GNMT. N-terminal sequencing is needed to confirm whether the recombinant proteins contain all amino acid residues reported in the native proteins. Recombinant GNMT expressed by Ogawa *et al.* (1997) and Luka and Wagner (2003) observed that the first methionine residue at the N-terminal was removed *in vivo* in *E.coli* compared to that expressed in rat liver cells (Ogawa *et al.*, 1987). Nevertheless, the recombinant GNMT structure and enzymatic properties were identical to the native rat liver GNMT.

In the case of recombinant cFBP, only the mature protein sequence was expressed in this chapter, without a signal peptide. The estimated subunit mass was 28 kDa, which was not significantly different from the predicted size of approximately 26 kDa when adding the 6xHis tag. However, the molecular weight of FBP from cow's milk was reported to be 35-40 kDa due to glycosylation of the protein (Salter *et al.*, 1981; Iwai *et al.*, 1983).

Given these differences in the properties of recombinant cFBP and GNMT versus their native proteins, an examination of the binding properties on folate derivatives, especially of the polyglutamylated folate forms, is required.

5.3.5 Folate binding assay

In an attempt to profile the kinetics of folate binding activity of recombinant mammalian folate binding proteins, optical biosensor technology was investigated. The kinetic analysis of small molecule interactions with proteins has previously been examined with surface plasmon resonance (SPR) biosensors (Karlsson and Fält, 1997). This technology provides many advantages over traditional assays such as ELISA or radioimmunoassay (Finglas and Morgan, 1994). For example, labelling is not needed on the binding molecules and very low amounts of sample (in the scale of micrograms) with low purity are required (Myszka, 1997). BIAcore® has developed commercial biosensors since 1990 with the key advantage of allowing real-time analysis of molecular interactions (Jönsson *et al.*, 1991).

The basis of the biosensor involves one reactant (referred to as the ligand) being covalently attached to the sensor surface and a second reactant (referred to as analyte) is introduced into the buffer above the sensor surface. The interaction between analyte and ligand is then monitored in real-time and a signal response is generated (Myszka, 1997). This is reported as the refractive index changes over time (referred to as resonance units; RU) which is proportional to a change in molecular mass of the interacted molecules on the sensor surface (Stenberg *et al.*, 1991).

Typically, the BIAcore® sensor surface is coated with a carboxymethyldextran matrix providing a flexible anchor for protein ligand immobilisation (Löfas and Johnsson, 1990). With specific ligands, various immobilisation methods are provided depending on their biochemical properties, such as amine coupling, thiol-disulphide exchange, streptavidin or avidin-biotin coupling. However, when selecting the coupling method, a uniform orientation is required for the immobilised reactants and no effect on the conformation of their binding sites (Löfas and Johnsson, 1990).

In the case of folate derivatives, direct amine coupling to the activated dextran is not an option because the amine group in folate is essential for binding to

FBP (Nygren *et al.*, 2003). Hence, folates are converted into their hydroxysuccinimidyl derivative prior to coupling to the amino group on the dextran surface using ethylenediamine (Nygren *et al.*, 2005). A previous study has reported the affinity of interaction between bovine FBP and folic acid as well as monoglutamate folates determined using SPR technology (Nygren *et al.*, 2005). Although folic acid showed the highest affinity to bovine FBP (Table 5.1), among natural folate derivatives, tetrahydrofolate exhibited the highest affinity over 5-methyl and 5-formyltetrahydrofolate (Nygren *et al.*, 2005). These data shows that differences exist between folate derivatives on their binding ability to FBP *in vitro*.

Table 5.1 Affinity and rate constant of FBP-folate interaction determined by BIAcore® SPR technology (taken from Nygren *et al.*, 2005)

folate	K_a ($M^{-1} s^{-1}$)	K_d (s^{-1})	K_D (pM)
PteGlu	$(1.0 \pm 0.4) \times 10^{5}$	$(1.3 \pm 0.4) \times 10^{-5}$	20 ± 9
(6*R*)-5-CH_3-H_4folate	$(3.0 \pm 0.4) \times 10^{5}$	$(4.7 \pm 1.7) \times 10^{-5}$	160 ± 50
(6*S*)-H_4folate	$(3.8 \pm 0.3) \times 10^{5}$	$(9.7 \pm 3.0) \times 10^{-5}$	250 ± 60
(6*S*)-5-CH_3-H_4folate	$(2.8 \pm 0.3) \times 10^{5}$	$(5.7 \pm 1.0) \times 10^{-4}$	2000 ± 200
(6*S*)-5-HCO-H_4folate	$(1.2 \pm 0.2) \times 10^{5}$	$(3.2 \pm 1.0) \times 10^{-3}$	12000 ± 3000

[a] Mean values ± standard error obtained from three or more independent measurements on different surfaces.

Recent efforts have been made using the same approach as the above report to characterise recombinant cFBP and GNMT. The binding affinity and rate constants of interactions between FBP and polyglutamylated folate derivatives are of great interest. However, due to technical issues, access to the Biacore facility has not been possible.

TRANSGENIC EXPRESSION OF RICE *FPGS* GENES AND MAMMALIAN FOLATE BINDING PROTEINS IN RICE ENDOSPERM TISSUE

In an attempt to enhance and retain the levels of folate in rice endosperm, cFBP, GNMT and FPGS coding sequences have been fused to the rice endosperm specific glutelin-1 promoter. Glutelin is a major seed storage protein representing 80% of total protein during seed development (Okita *et al.*, 1989) which is expressed most strongly in the endosperm 17 days after flowering (Usuda *et al.*, 2009).

The glutelin-1 promoter has been successfully used to express several genes in rice seed, including phytoene synthase, lycopene β-cyclase and phytoene desaturase for golden rice generation (Ye *et al.*, 2000) and the human lysozyme gene in rice grains (Huang *et al.*, 2002). Given such successful examples, the glutelin-1 promoter was used to express *FPGS* and *FBP* genes in rice endosperm.

5.3.6 Gene cloning

Amplification of the rice glutelin-1 promoter was initially performed. The glutelin-1 promoter was generated by PCR amplification using rice genomic DNA as a template. The specific oligonucleotide primers shown in appendix 1 were used to generate the promoter part. A single band of the PCR product corresponding to the glutelin-1 promoter (for gene sequence, see Figure 5.13) was obtained as shown in figure 5.14 and the size of the product was about 900 base pair as expected. The PCR product was purified and DNA sequencing performed to confirm the correct sequence. The correct insert was then cloned into the pENTR11 vector using NcoI and SalI sites.

The two rice *FPGS* genes (for cDNA sequences, see appendix 3) were also amplified from Nipponbare rice cDNA as shown in Figure 5.15. The expected bands of 1,608 bp *FPGS* Os03g02030 and 2,057 bp *FPGS* Os10g35940 were

obtained. Sequences were confirmed and the gene fragments were transferred into pENTR11 at SalI and NotI sites downstream of the glutelin promoter.

```
TTCTGTAGTACAGACAAAACTAAAAGTAATGAAAGAAGATGTGGTGTTAGAAAAGGAAACAATATCATGA
GTAATGTGTGAGCATTATGGGACCACGAAATAAAAAGAACATTTTGATGAGTCGTGTATCCTCGATGAGC
CTCAAAAGTTCTCTCACCCCGGATAAGAAACCCTTAAGCAATGTGCAAAGTTTGCATTCTCCACTGACAT
AATGCAAAATAAGATATCATCGATGACATAGCAACTCATGCATCATATCATGCCTCTCTCAACCTATTCA
TTCCTACTCATCTACATAAGTATCTTCAGCTAAATGTTAGAACATAAACCCATAAGTCACGTTTGATGAG
TATTAGGCGTGACACATGACAAATCACAGACTCAAGCAAGATAAAGCAAAATGATGTGTACATAAAACTC
CAGAGCTATATGTCATATTGCAAAAAGAGGAGAGCTTATAAGACAAGGCATGACTCACAAAAATTCACTT
GCCTTTCGTGTCAAAAAGAGGAGGGCTTTACATTATCCATGTCATATTGCAAAAGAAAGAGAGAAAGAAC
AACACAATGCTGCGTCAATTATACATATCTGTATGTCCATCATTATTCATCCACCTTTCGTGTACCACAC
TTCATATATCATAAGAGTCACTTCACGTCTGGACATTAACAAACTCTATCTTAACATTTAGATGCAAGAG
CCTTTATCTCACTATAAATGCACGATGATTTCTCATTGTTTCTCACAAAAAGCATTCAGTTCATTAGTCC
TACAACAACATGGCATCCATAAATCGCCCCATAGTTTTCTTCACAGTTTGCTTGTTCCTCTTGTGCGATG
GCTCCCTAGCCCAGCAGCTATTAGGCCAGAGCACTAGTCAATGGCAGAGTTCTCGTCGTG
```

Figure 5.13 The coding sequence of rice glutelin-1 promoter

The complete *japonica* rice glutelin-1 promoter sequence was 900 bp in length. The gene code from NCBI database is M28156.

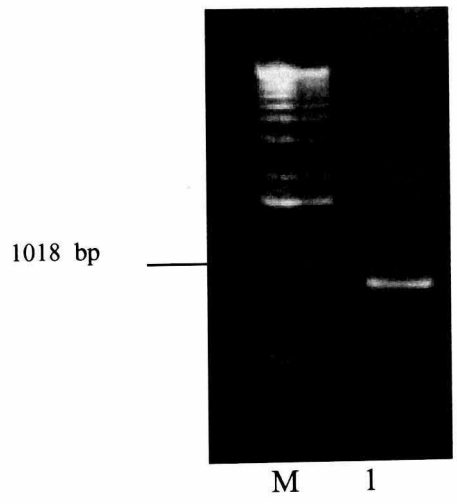


Figure 5.14 Amplification of the glutelin-1 fragment by polymerase chain reaction

The PCR products of glutelin-1 promoter (lane 1) were purified and separated alongside the size markers (M; 1 kb ladder) on a 1% agarose gel containing ethidium bromide.

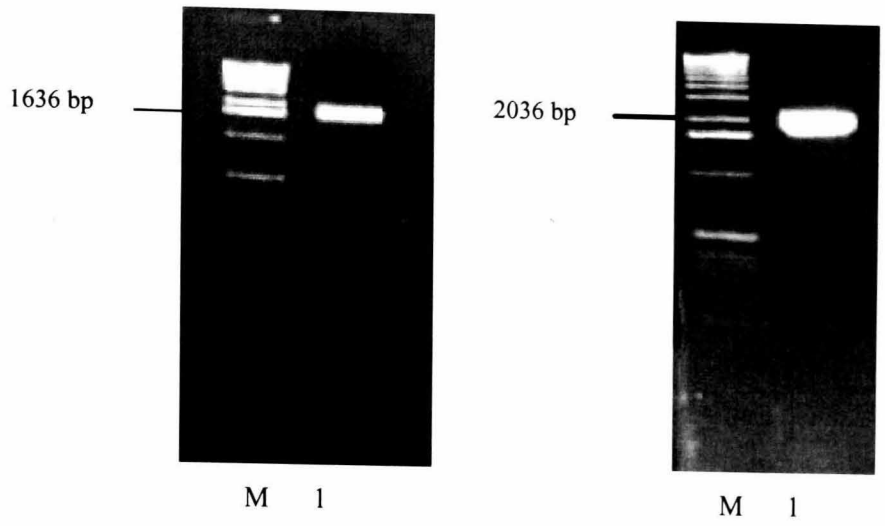


Figure 5.15 Amplification of *FPGS* Os03g02030 and *FPGS* Os10g35940

The PCR products of *FPGS* Os03g02030 (lane 1, left picture) and *FPGS* Os10g35940 (lane 1, right picture) were purified and separated alongside the size markers (M; 1 kb ladder) on a 1% agarose gel containing ethidium bromide. The expected band size of *FPGS* Os03g02030 was 1,608 bp and *FPGS* Os10g35940 of 2,057 bp.

cFBP and GNMT fragments from the same clones used to generate protein expression clones were used (Figure 5.8). *cFBP* and *GNMT* coding sequences were then cloned downstream of the glutelin promoter in pENTR11 in the same approach as *FPGS*.

Using gateway compatible sites (attL) in pENTR11 vector, the glutelin promoter and genes of interest could be transferred into the binary vector, pGWB7. The final working clones are shown in Figure 5.16. The direction of the inserts was checked by PCR using promoter and gene specific primers and sequencing. Then, pGWB7 containing *FPGS* Os03g02030, *FPGS* Os10g35940, *cFBP* and *GNMT* clones were transformed into the EHA105 *Agrobacterium* strain prior to rice transformation.

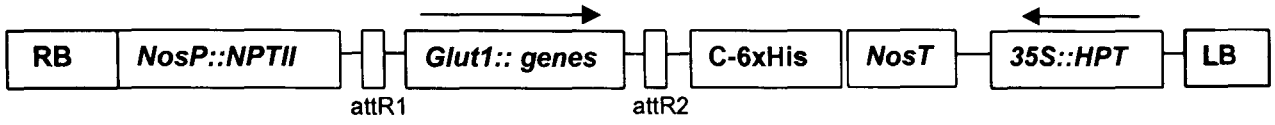


Figure 5.16 Diagram of pGWB7 vector containing rice glutelin promoter (Gt1) and interested genes

The structure of pGWB7 is presented between the right (RB) and left (LB) borders. In the plants, neomycin phosphotransferase II (NPTII) driven the expression by nopaline synthase promoter (NosP) confers kanamycin resistance and hygromycin phosphotransferase (HPT) driven the expression by the cauliflower mosaic virus (CaMV) 35S prompter (35S) confers hygromycin resistance. The nopaline synthase terminator (NosT) was used to stop expression of both selectable genes. This vector contains hexahistidine tag (C-6xHis) at the C-terminal of the inserted genes. The genes of interest were *cFBP, GNMT, FPGS* Os03g02030 and *FPGS* Os10g35940 driven the expression by glutelin-1 promoter (Glut1). attR1 and attR2 are the Gateway compatible sites. This diagram was modified from Nakagawa *et al.* (2007).

5.3.7 Rice transformation

Agrobacterium-mediated transformation in rice has been successfully carried out for the past 10 years in both *japonica* and *indica* rice. The advantages of this system over other systems involves the high efficiency of transformation events, the stability of the integration of the foreign genes (Chan *et al.*, 1993) and the precise gene orientation with minimal rearrangements of DNA during the transformation process (De Buck *et al.*, 2000). Many attempts have since improved the transformation protocol for rice (Hiei and Komari, 2008).

In this chapter, the transformation protocol was obtained from Maarten Volckaert (Dominique's lab, Ghent University) based on the protocol of Hiei *et al.* (1994). This method has previously been used successfully in the construction of high-folate rice (Storozhenko *et al.*, 2007). Hence, it was applied in this study.

Rice transformation was performed in collaboration with members of the Dominique Van Der Straeten group (Ghent University). *Agrobacterium*-mediated transformation was conducted using the Nipponbare *japonica* rice variety. The *Agrobacterium* infected rice embryogenic callus and the selection methods were performed (see section 2.4, chapter 2). The whole process took around 10-12 months to get the transgenic rice seeds. The transformation process is summarized in Figure 5.17.

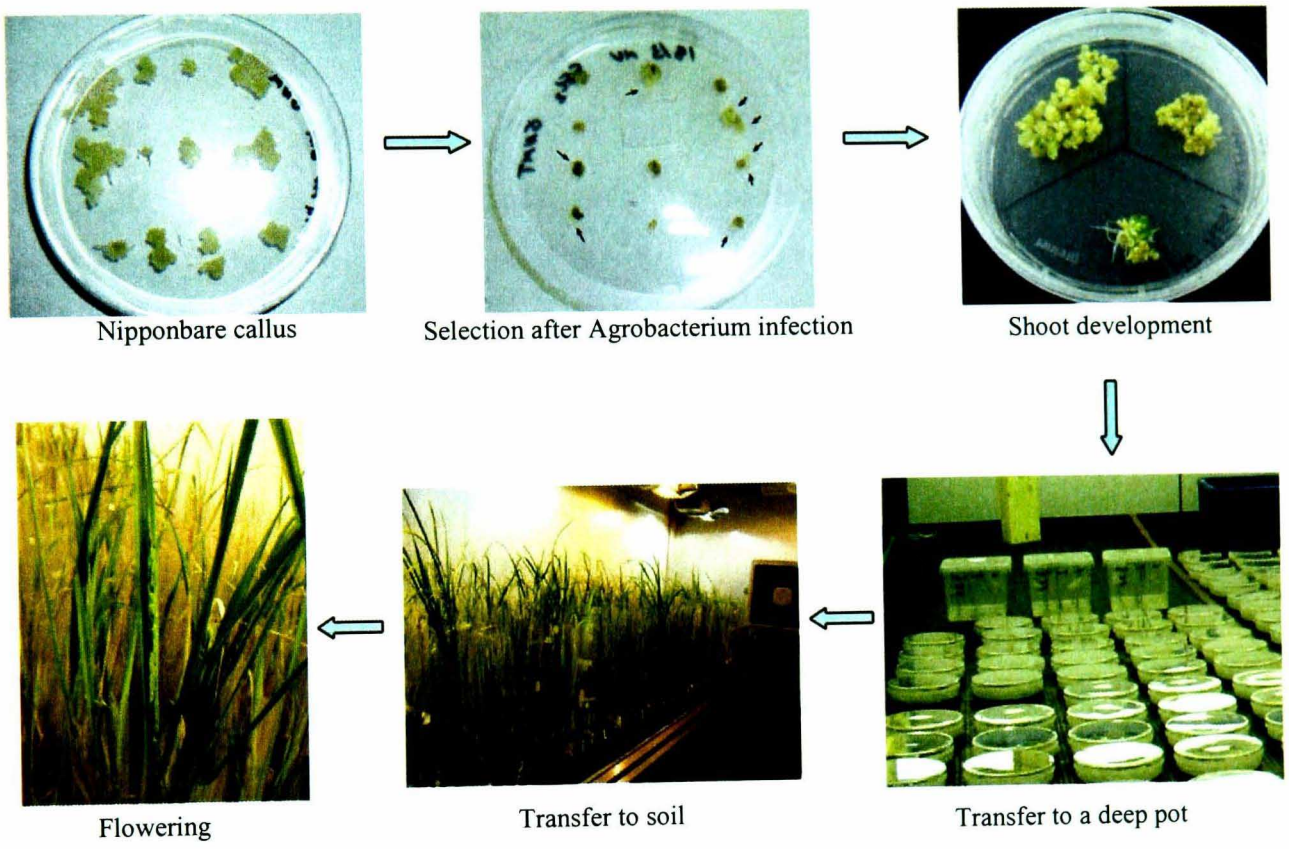


Figure 5.17 The process of rice transformation

Rice transformation was started with embryogenic callus. *Agrobacterium*-mediated transformation was performed and infected callus was selected on hygromycin antibiotic plates. The shooting callus was transferred to the bigger pot and to soil waiting for flowering and seed filling. Pictures were taken by Maarten Volckaert (Ghent University).

5.3.8 Genotyping of transgenic rice

A total of 69 primary transformant hygromycin-resistant plants (T_0) were obtained after Agrobacterium infection and calli selection. Molecular analysis of transgenic T_1 rice plants was first performed in order to select the true transgenic lines containing each of the four transgenes. PCR was performed on DNA from T_1 seeds using gene-specific primers as shown in appendix 1. Rice glutelin promoter was used as the forward primer named riceglu_for. The reverse primers were the gene-specific primers which are cFBPrt_rev, GNMTrt_rev, 3gE4_rev10gE7_rev for *cFBP*, *GNMT*, *FPGS* Os03g02030 and *FPGS* Os10g35940 constructs, respectively. The expected band size from PCR of *cFBP*, *GNMT*, *FPGS* Os03g02030 and *FPGS* Os10g35940 constructs was 700, 920, 475 and 1050 bp, respectively. The results of genotyping are shown in Figure 5.18 indicating the successful integration of genes of interest.

Genotyping confirmed the success of rice transformation for all 4 constructs. In total, 20 transgenic lines contained *cFBP*, 13 lines for *GNMT*, 9 lines for *FPGS* Os03g02030 and 8 lines for *FPGS* Os10g35940. Plants are currently being cultivated in the glass house to collect the self-pollinated seed (T_2).

5.3.9 Transgenic rice analysis

To confirm the over-expression event, the transcript levels of FPGS and FBPs were measured in dough T_2 seeds (17 days post-anthesis) by relative RT-qPCR and their levels of expression presented relative to that in wild type control (Figure 5.19). All the selected transgenic lines showed an increased expression of the target transgene but with huge variation. However, this result confirmed the successful of expression of these transgenes in rice endosperm, when driven by the rice glutelin promoter.

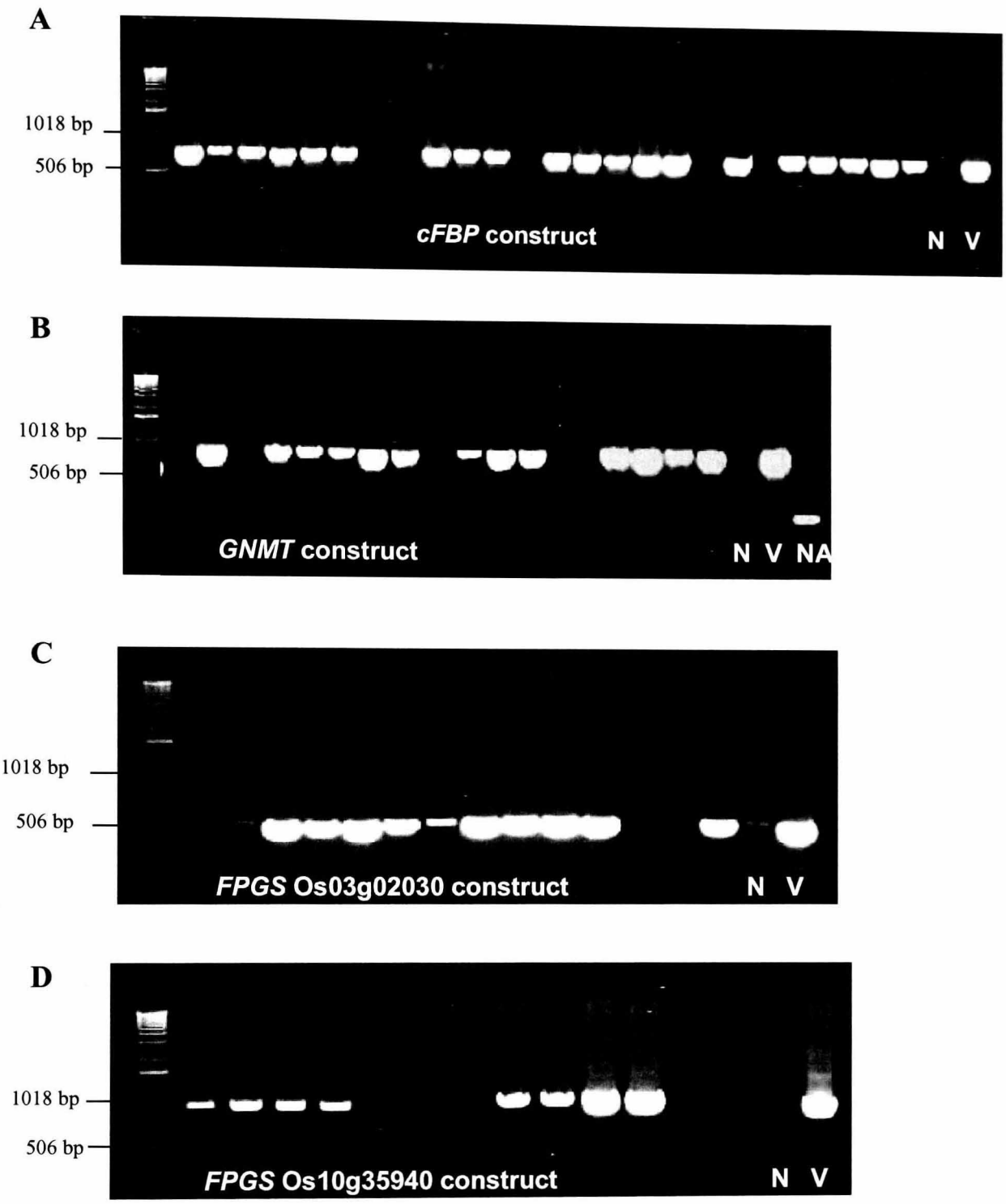
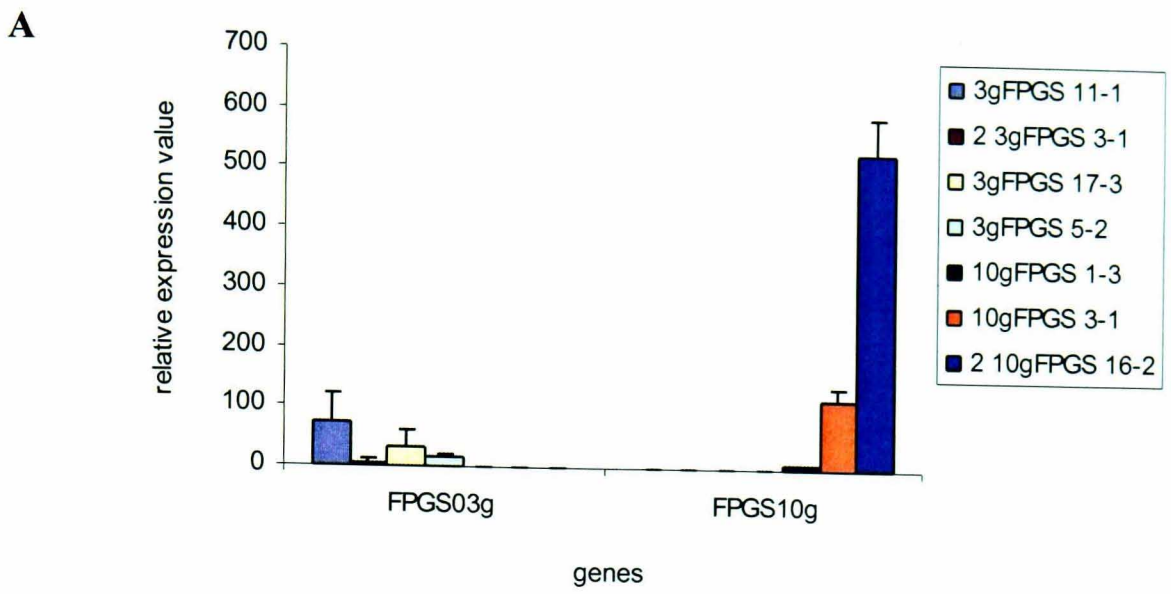


Figure 5.18 Genotyping of transgenic rice lines

Young leaves of transgenic lines for each gene were collected for DNA extraction and PCR analysis. Using the rice glutelin forward and gene-specific reverse primers in PCR indicated the insertion of the binary vector. The primer sequences are shown in appendix 1. **A**: positive lines for 20 out of 25 lines for *cFBP* construct which the expected band size was 700 bp; **B**: positive lines for 13 out of 17 lines for *GNMT* construct which the expected band size was 920 bp; **C**: positive lines for 9 out of 14 lines for *FPGS* Os03g02030 construct which the expected band size was 475 bp and **D**: positive lines for 8 out of 13 lines for *FPGS* Os10g35940 construct which the expected band size was 1050 bp. N represents negative control (Nipponbare wild type DNA). V represents positive control (binary vector contained genes of interest). NA represents the control of Nipponbare wild type DNA using actin primer.



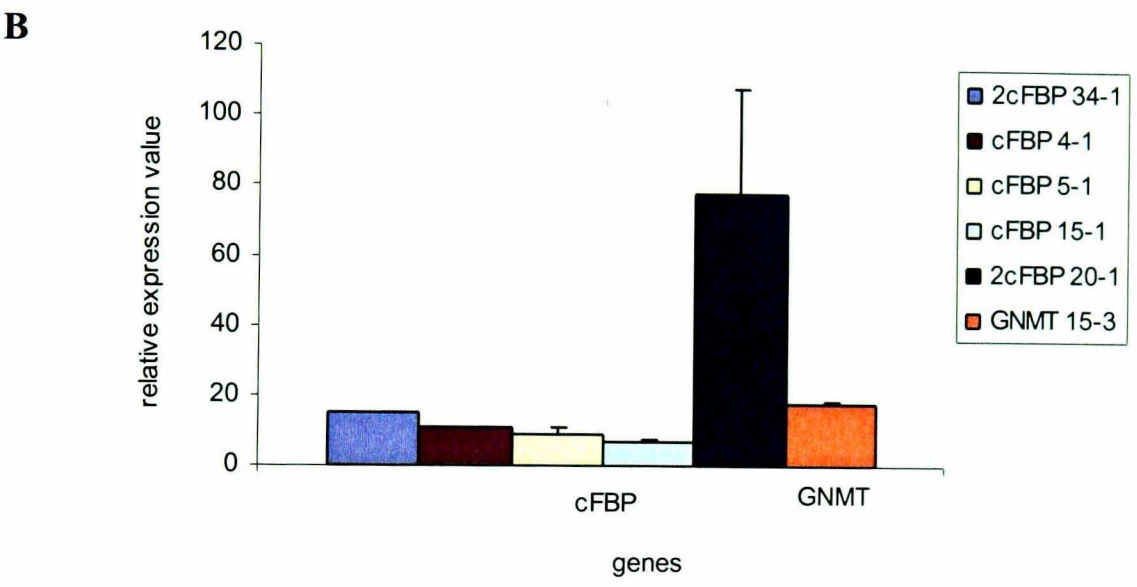


Figure 5.19 Relative transcript levels of transgenes in rice endosperm

At least 5 dough seeds (at 17 days post-anthesis) of selected transgenic lines were collected for relative RT-qPCR analysis. Gene specific primers, as shown in appendix, were used to monitor the expression of transgenes. FPGS transcript levels are shown in figure A, whilst cFBP and GNMT transcript levels are shown in figure B. The transcript levels are expressed relative to the transcript level of the same gene in Nipponbare wild type plants grown under the same conditions. Each sample was normalised to actin.

Folate concentration was also quantified by LCMS/MS (by Riza Albigos-Ramos, personal communication) in selected transgenic rice lines. Table 5.2 shows the concentration of folate forms and total folate concentration in 15 transgenic lines.

Over-expression of both rice FPGS isoforms in rice endosperm appear to enhance total folate approximately 5 folds compared to wild type rice. The majority of polyglutamylated folate was present as 5-methylhexaglutamated folate, which interestingly does not exist in Nipponbare wild type. In 3gFPGS and 10gFPGS lines, monoglutamate folate levels were no different to those in wild type indicating that over-expression of FPGS isoforms specifically increases amounts of polyglutamylated folates only. This result supports the major function of FPGS enzymes which could produce more polyglutamylated folate in over-expression rice lines.

Expression of the folate binding proteins, cFBP and GNMT, resulted in an increase in total folate of approximately 7-10 fold compared to wild type. These FBPs over-expressing lines thus increased total folate more than the FPGS over-expressed lines by also increasing the amount of monoglutamate folate. Monoglutamate folate levels in FBPs lines were increased up to 5-8 fold compared to wild type and FPGS over-expression lines. This could imply that FBPs might have a higher affinity for monoglutamate folate than polyglutamylated folate forms. However, a study of binding affinity of recombinant FBPs to a variety of folate molecules is still required to support this point of view.

Enhancement of folate by as much as 100 times has been reported in transgenic rice (Storozhenko *et al.*, 2007) by over-expression of *Arabidopsis GTPCHI* and *ADC synthase* genes in rice endosperm. However, 89% of this enhanced folate was as 5-methyltetrahydrofolate monoglutamate folate which is different to the enhancement of polyglutamylated folates reported in the present study. Due to the possibility of folate loss during storage and cooking procedures, folate stability should be tested in those transgenic rice lines. Polyglutamylated rich

rice lines might exhibit better retention of folate than monoglutamate rich rice lines due to the ability of polyglutamated folate to be retained within cells.

5.4 SUMMARY

- The folate binding protein coding sequences of cFBP and GNMT have been isolated from cow and rat sources.
- 6xHis-tagged recombinant folate binding proteins were successfully expressed and partially purified in *E.coli*.
- Folate binding affinity assay will be formed using BIAcore® technology to determine binding affinity to polyglutamylated folate derivatives.
- Transgenic rice lines containing *cFBP*, *GNMT* and rice *FPGS* transgenes have been successfully generated using *Agrobacterium*-based rice transformation.
- Enhancing transcript levels of *FPGS* Os03g02030 and *FPGS* Os10g35940, *cFBP* and *GNMT* can increase folate concentration in rice endosperm in both mono- and polyglutamylated folate forms.

Table 5.2 Amount of mono- and polyglutamylated folate forms in transgenic *FPGS* and *FBPs* mature seeds (mean ± SD; *n=3*) compared to Nipponbare wild type seeds

Transgenic lines	Folate Form (µg/100g)							
	5-CH_3 H_4PteGlu	5/10-CHO H_4PteGlu	5-CH_3 H_4PteGlu$_4$	5-CH_3 H_4PteGlu$_5$	5/10-CHO H_4PteGlu$_5$	5-CH_3 H_4PteGlu$_6$	5/10-CHO H_4PteGlu$_6$	Total Folate
3g*FPGS* 11-1	13.21 ± 3.1	0.30 ± 0.1	0.48 ± 0.05	0.85 ± 0.1	0.63 ± 0.1	14.15 ± 4.1	4.81 ± 0.5	34.43 ± 2.3
23g*FPGS*3-1	21.81 ± 2.3	0.36 ± 0.04	0.50 ± 0.1	1.10 ± 0.1	0.38 ± 0.04	34.43 ± 3.5	5.16 ± 0.6	63.74 ± 1.2
3g*FPGS*17-3	13.91 ± 1.3	0.14 ± 0.01	0.50 ± 0.03	0.82 ± 0.1	0.38 ± 0.	32.23 ± 1.2	3.13 ± 0.2	51.11 ± 0.5
3g*FPGS*5-2	18.41 ± 2.5	0.15 ± 0.03	0.49 ± 0.1	0.81 ± 0.0	0.30 ± 0.02	24.11 ± 7.0	2.09 ± 0.1	46.36 ± 1.4
10g*FPGS*1-3	13.31 ± 0.7	0.28 ± 0.0	0.56 ± 0.05	0.97 ± 0.2	0.23 ± 0.03	31.31 ± 4.5	1.54 ± 0.3	48.20 ± 0.8
10g*FPGS*3-1	11.71 ± 0.9	0.78 ± 0.1	0.53 ± 0.1	1.13 ± 0.4	0.31 ± 0.0	30.27 ± 3.9	3.07 ± 0.2	48.25 ± 0.8
210g*FPGS*16-2	8.11 ± 3.0	0.19 ± 0.0	0.59 ± 0.1	1.04 ± 0.2	0.39 ±0.1	43.19 ± 12.3	2.83 ± 0.9	56.34 ± 2.4
2cFBP20-1	23.01 ± 6.0	0.69 ± 0.04	0.99 ± 0.1	2.38 ± 0.3	-	21.94 ± 4.3	6.72 ± 0.5	55.73 ± 1.9
cFBP5-1	50.11 ± 5.4	0.74 ± 0.01	0.64 ± 0.2	1.22 ± 0.2	0.28 ± 0.2	26.19 ± 4.7	3.98 ± 1.5	83.16 ± 3.1
cFBP4-1	38.81 ± 3.5	0.88 ± 0.03	0.67 ± 0.3	1.37 ± 1.1	0.65 ± 0.4	25.79 ± 1.0	5.60 ± 1.9	73.77 ± 1.2
cFBP15-1	13.21 ± 5.6	0.69 ± 0.04	1.70 ± 0.6	1.80 ± 0.1	0.17 ± 0.01	29.24 ± 8.3	5.19 ± 2.9	52.00 ± 2.5
2GNMT25-2	51.56 ± 6.1	5.43 ± 1.9	4.24 ± 0.9	0.97 ± 0.1	-	30.19 ± 12.9	4.86 ± 0.8	97.25 ± 3.3
GNMT15-3	62.61 ± 11.3	8.38 ± 2.8	0.71 ± 0.5	1.01 ± 0.4	0.67 ± 0.2	-	4.99 ± 0.7	78.37 ± 2.7
GNMT18-3	40.16 ± 4.9	5.18 ± 0.8	0.81 ± 0.03	0.92 ± 0.1	2.12 ± 1.6	34.35 ± 17.1	-	83.54 ± 4.1
GNMT4-2	86.61 ± 7.2	5.43 ± 2.4	2.03 ± 0.9	2.30 ± 0.9	-	22.73 ± 7.6	4.69 ± 0.4	119.10 ± 3.3
Nipponbare	11.09 ± 0.5	1.69 ± 0.04	0.71 ± 0.5	-	-	-	-	13.49 ± 0.4

CHAPTER 6 GENERAL DISCUSSION

6.1 INTRODUCTION

Folate deficiency represents one of a major nutritional challenge in many countries worldwide especially in Asia and Africa countries. The consequences of low folate intake include birth defects, megaloblastic anemia and increased risk of vascular diseases. Plants are the main sources of folate in human diets because human and animals lack important enzymes to synthesize folate *in vivo*. The daily requirement, 400 μg for normal adults and up to 600 μg for pregnant and lactating women, is required to prevent the severe consequences. However, some staple fruits and crops are poor in this vitamin. Stable foods such as rice contains very low amounts of folate. White rice (which is the most popular form of rice for human consumption) contains less than 80 μg per 100 g folate. To reach the folate requirement, higher amounts of rice need to be consumed to reach the minimum requirement. Dietary deficiency could be reduced using folic acid-fortified foods or by folic acid supplementation. However, these solutions have only partially worked in developed countries. Given the significant recurrent costs, it is hard to implement in poorer countries. Crop biofortification represents a long term solution and provides an alternative way to improve nutritional qualities. Given that it forms a major part of the South East Asian and African diet, rice is an important target to improve the nutritional status.

This thesis attempts to cover the knowledge of folate biosynthesis in rice. Four major issues are discussed below:

- Folate biosynthesis and rice development
- Investigating the importance of folate polyglutamylation during rice development
- The existence of a folate homeostasis mechanism in rice

- Improvement in abundance and bioavailability of folate through exploitation of natural variation and genetic engineering approaches

6.1.1 Folate biosynthesis and rice development

All 10 steps involved in the plant folate biosynthesis pathway have been described and characterised in detail (Bekaert *et al.*, 2007). Moreover, the expression of several enzymes involved in this pathway has been correlated with the folate pool and the one-carbon unit demands in various physiological situations (Basset *et al.*, 2002, 2004a, 2004b; Jabrin *et al.*, 2003). The abundance of transcripts encoding key enzymes in the folate biosynthesis pathway appears to be correlated with the demand for folate-dependent metabolites in one-carbon metabolism during rice development. Increased transcript abundance of folate biosynthesis genes correlated with high folate abundance in mature rice seed. Up-regulation of transcript levels of all key genes in tetrahydrofolate synthesis, *GTPCHI*, *ADC synthase* and *HPPK/DHPS*, was observed in the high folate rice variety, Moroberekan (60 μg/100 g) compared to other low folate rice varieties, Shan Yon 63, Mongur, IR72 and Nipponbare (Figure 3.12). Hence, the variation of total folate concentration appeared to be controlled at the transcriptional level, in agreement with a recent study in potato (Goyer and Navarre, 2009).

Transcript analysis has revealed that there is contrasting demand for folate during rice development. The d*e novo* folate synthesis machinery is active in all rice tissues based on the presence of mRNA for folate biosynthesis genes (Figure 3.9). The differences in mRNA abundance between rice tissues may indicate contrasting folate requirement during each developmental stage. The demand for folate appears to increase following germination. This is in agreement with previous studies in pea that reported low folate in the first day following imbibition which then increased to the maximum level 7 days after germination (Gambonnet *et al.*, 2001; Chan and Cossins, 2003). An increase in the level of folate-dependent enzymes; for example, methionine synthase

catalysing the last step of methionine synthesis and S-adenosylmethionine synthase catalysing the synthesis of S-adenosylmethionine from methionine (Gallardo *et al.*, 2002), indicates the essential role of folate during germination. Methyl requirement is also reported in seedling and developing plant depending on the degree of methylation processes by methyltransferase enzyme (Moffatt and Weretilnyk, 2001; Chan and Cossins, 2003). High concentration of several methylated compounds reflects the methyl requirement and 5-methyltetrahydropteroylpolyglutamates are a major source of methyl groups (Hanson and Roje, 2001). For example, pectin, chlorophyll, lignin and choline are required during cell wall synthesis and cell expansion processes (Giovanelli *et al.*, 1985; Zhong *et al.*, 1998). Reduction of methylation activity due to a reduction of folate results in a pleiotropic phenotype with stunted and wavy leaves (Tanaka *et al.*, 1997). Folate is synthesized preferentially in photosynthetic leaves and in light conditions (Jabrin *et al.*, 2003; Chen *et al.*, 1997; Chan and Cossins, 2003) which correlates with an increase of mRNA abundance for all folate biosynthesis genes expressed in rice leaves (Figure 3.9). Most biosynthetic enzymes were only gradually accumulated in the cytosol of etiolated leaves compared to green leaves (Chen *et al.*, 1997; Jabrin *et al.*, 2003). For example, the accumulation of two folate-dependant enzymes, SHMT and GDC, involved in photorespiration during greening correlates to the high folate concentration (Oliver *et al.*, 1990; Gambonnet *et al.*, 2001).

After the vegetative stage of rice, demand for folate appears to drop after anthesis during the seed filling stages. However, folate is required for nucleotide synthesis during embryo development which occurs for a few days after anthesis (Gambonnet *et al.*, 2001). mRNA level of folate-related genes appears to reach a maximum at this stage (Figure 3.9). During the seed filling stage, a reduction of mRNA abundance exhibits the lowest demand for folate. Instead, starch and seed storage protein preferentially accumulate at this stage, given by increased transcript levels of genes involving in their biosynthetic pathway (Zhu *et al.*, 2003).

At the maturation stage, an increased accumulation of folate biosynthesis genes was observed again in dry rice mature seed suggesting the accumulation of folate and folate-related genes in preparation for germination. Blocking folate synthesis during germination in *Arabidopsis* using sulphanilamide drugs did not block the germination process suggesting that sufficient amounts of folate in mature seed is available to support seedling establishment (Gambonnet *et al.*, 2001).

6.1.2 Investigating the importance of folate polyglutamylation during rice development

Most folate-dependent enzymes prefer polyglutamylated folate forms for their activities (Shane, 1995; Scott *et al.*, 2000; Strong *et al.*, 1990; Cichowicz and Shane, 1987). In all organisms, polyglutamylated folate is produced by folylpolyglutamate synthetase (FPGS). In *Arabidopsis*, three *FPGS* genes encode distinct mitochondrial, cytoplasmic and plastidial FPGS isoforms (Ravanel *et al.*, 2001). Interestingly, only two *FPGS* genes, *FPGS* Os03g02030 and *FPGS* Os10g35940, were identified in rice from comparative genomics which encode two putative FPGS protein sequences. Both FPGS putative protein sequences are predicted to be dual targeted to mitochondria and chloroplasts. Rice FPGS isoforms retain 50-57% protein identity with the three *Arabidopsis* FPGS isoforms and contain all functional motifs for ATP, tetrahydrofolate (THF) and glutamate molecules as reported previously in other eukaryotes (Sun *et al.*, 2001).

Transcript analysis suggests that *FPGS Os03g02030* appears to function predominantly in grains whilst *FPGS Os10g35940* is the major isoform in leaves (Figure 3.8). Disruption of *FPGS* Os03g02030 caused delayed seed filling and maturity (Figure 4.18B) and a reduction of polyglutamylated folate content in mutant seed (Table 4.2). Impaired folate polyglutamylation in seed might impact nucleotide synthesis and cell division causing a delay in embryo development which may impact other one-carbon processes; such as glycine and methionine synthesis (Gambonnet *et al.*, 2001) which require high amounts

of folate during cell proliferation (Jabrin *et al.*, 2003). Reduced rates of cell division could result from a defect in thymidylate synthesis and dihydrofolate formation (McNulty *et al.*, 1995). Hence, *FPGS* Os03g02030 gene appears to play a key role during rice seed development.

The development of leaf and mature seed as well as germination ability was not affected in plants lacking *FPGS* Os03g02030 (Figure 4.18A, C and D). Another FPGS isoform, encoded by the *FPGS* Os10g35940 gene, appeared to compensate for the loss of FPGS 03g02030 activity during the germination and the vegetative stages, based on the increase of *FPGS* Os10g35940 transcript abundance in the *fpgs*03g mutant leaf (Figure 4.19). Hence, *FPGS* Os10g35940 gene is suggested to play a major role in the leaf. Compensation from *FPGS* Os*10g35940* gene expression appears to maintain polyglutamylated folate content in the mutant leaf with no significant reduction of 5-methyltetrahydrofolate tetraglutamate and 5-formyltetrahydrofolate pentaglutamate compared to the mutant seed (Table 4.2).

A double *FPGS* knock out plant is required to get more knowledge about the importance of polyglutamylated folate forms for plant development. Earlier observations in other eukaryotes indicate that disrupting the generation of polyglutamylated folate resulted in auxotrophies for methionine and other products of one-carbon reactions; for example, thymidine and purine (Cossins, 1987; Moran *et al.*, 1976). In *Arabidopsis*, a double knock out *FPGS* mutant resulted in the decreased levels of folate-dependent metabolites including methionine (Payam Mehrshahi, PhD thesis 2008, University of Nottingham). Supplementation with a high level of methionine could partially rescue the dwarf phenotype of the mutant, revealing the functional importance of polyglutamylation to plant metabolism and development (Payam Mehrshahi, PhD thesis 2008, University of Nottingham).

6.1.3 The existence of a folate homeostasis mechanism in rice

A significant reduction in the monoglutamate folate form in the *fpgs*03g mutant was observed suggesting that polyglutamylation may influence folate breakdown. A significant proportion of the mitochondria folate pool presents in a protein-bound form to folate-dependent enzymes (Chen *et al.*, 1997). Polyglutamylated protein-bound folates are prevented from undergoing oxidative degradation and deglutamylation by the γ-glutamyl hydrolase (GGH) enzyme (Rébeillé *et al.*, 1994). GGH is responsible for removing glutamic acid from polyglutamylated folates and has been identified in human (Yao *et al.*, 1996a), rat (Yao *et al.*, 1996b), pea leaf (Lin *et al.*, 1993), soybean (Huangpu *et al.*, 1996) and *Arabidopsis* (Orsomando *et al.*, 2005). This catabolic process is presumably under the control of the folate salvage pathway which helps maintain the homeostasis of cellular folate (Orsomando *et al.*, 2006; Schneider and Ryan, 2006). Changing rates of folate synthesis and glutamylation might affect the accumulation of short-chained folate involving in homeostasis of folate in eukaryotic cells (Chan and Cossins, 2003). In rice, the existence of such a mechanism was observed through a reduction in *GGH* transcript levels (Figure 4.19). This may represent a mechanism to compensate for the disruption of *FPGS* Os03g02030 gene and help maintain the cellular polyglutamylated folate pool. Hence, polyglutamylation and catabolism represent important factors influencing total cellular folate concentration which may account for the differences in folate abundance in specific tissues (Eisenga *et al.*, 1992; Suh *et al.*, 2000; Anguera *et al.*, 2003).

FPGS and *GGH* activities appear to play an important role during folate homeostasis in addition to regulation of the folate biosynthesis. Although mammalian and plant systems share similarities in folate homeostasis mechanisms through FPGS and GGH processes, a few differences are observed. Differences between plants and mammals exist due to the inability to synthesise folate in mammalian tissues. The homeostasis mechanism in mammalian cells is dependent mostly on folate turnover and circulation, whilst, the plant system is mostly concerned with the rate of folate biosynthesis, and

turnover in cells. The homeostasis of folate in mammals mostly works through the enterohepatic cycle. To maintain plasma levels of folate, oxidized folate is recirculated to the liver (Hilton *et al.*, 1979). Disrupting this homeostasis mechanism resulted in a rapid increase by 70% of normal serum folate levels (Steinberg *et al.*, 1979). In plant systems, there are no reports about the transportation and circulation of folate metabolites from sources to sink sites.

6.1.4 Improvement in abundance and bioavailability of folate through exploitation of natural variation and genetic engineering approaches

Biofortification provides an attractive way to improve food quality. From the knowledge about genetic variation of folate biosynthesis pathway indicated by the huge amount of transcript expression of folate biosynthesis genes, metabolic engineering is suggested. To date, up to a 100 fold increase of total folate (89% monoglutamate folate) has been observed in rice endosperm through overexpressing *Arabidopsis GTPCHI* and *ADC synthase* genes which are the key enzymes in the pterine and *p*ABA synthesis branches (Storozhenko *et al.*, 2007). The percentage of polyglutamylated folates in transgenic rice has been reported as being approximately 2.6-14%. However, bioavailability and stability of these folate pools have been questioned.

In terms of bioavailability, polyglutamylation protects folate from oxidative breakdown and enzymatic digestion (Suh *et al.* 2001, Jones and Nixon 2002). In mammalian studies, permeability of polyglutamylated folate is noticeably less and its derivatives are retained within cells (McGuire and Bertino, 1981). Hence, increasing polyglutamylated folate forms might also facilitate stability and retention of cellular folate concentration in rice cells through protein-bound mechanism. Also, mammalian folate binding protein such as cow's milk folate binding protein (cFBP) could help stabilise both mono- and polyglutamated folates and decrease their oxidation rate. To support the above statement, FBP has been reported to improve the bioavailability of food folate but the relevance to daily human folate status has been questioned (Picciano *et al.*, 2004). This thesis suggests an alternative way to enhance folate stability

through expressing high levels of polyglutamylated folate and enhancing folate-bound forms which might retain folate within rice endosperm. Currently, folate binding proteins and two rice *FPGS* genes overexpressed rice lines have been generated successfully and further studies will reveal their impact on bioavailability.

6.2 KEY QUESTIONS AND FUTURE STUDIES

6.2.1 Considerable natural variation in folate abundance exists in rice varieties. Transcription profiling reveals that folate levels are closely correlated with transcript abundance. Balam and Moroberekan seeds exhibit a very high total folate level; approximately 60-80 µg/100g. To pinpoint the regulatory genes controlling the production and accumulation of folate, QTL analysis should be carried out. Future studies should use molecular markers which can aid breeding programmes by crossing these varieties with market-based varieties to improve folate quality in popular rice.

6.2.2 To obtain more insight into the importance of the polyglutamylation process in rice, the creation of a double knock out line lacking both rice FPGS genes, *FPGS* Os03g02030 and *FPGS* Os10g35940 is required. RNAi represents an alternative option to reduce expression of both rice *FPGS* genes

6.2.3 GGH is an enzyme in the folate degradation and salvage pathway which is thought to control the ratio of monoglutamate and polyglutamate folate forms in rice tissues. In the rice *FPGS* mutant, this enzyme is down-regulated through the decrease of mRNA transcript level. However, to manipulate the down-regulation of folate synthesis through GGH activity, the knock out *GGH* mutant may be required to describe the function of GGH on folate synthesis rate. Future studies should attempt to manipulate the flux of folate derivatives using metabolic engineering to reduce the turnover of polyglutamylated folates and increase retention of polyglutamylated folates in cells.

6.2.4 Mammalian folate binding proteins could be used to generate new rice varieties with the aim of retaining high levels of mono/polyglutamylated folates. Future studies should investigate the effectiveness of such a strategy either alone or in conjunction with other metabolic engineering approaches.

6.3 SUMMARY

This thesis provides new insights about folate biology including discovering natural genetic variation and a novel homeostasis mechanism during rice development. Total variation in folate abundance appears to be largely controlled by transcriptional regulation. Without *FPGS* Os03g02030 gene, one of *FPGS* genes responsible in production of polyglutamylated folate, the defect of seed maturation was observed and suggests the major function of this gene during seed development. The compensation from another *FPGS* gene, *FPGS* Os10g35940, occurs together with the reduction of folate break down processed by GGH enzyme in order to maintain folate pool for normal physiological function during rice development. As rice contains very low folate levels and is an important staple food in most of developing countries, these biological observations could be applied further to improve folate status in rice by combining metabolic engineering and conventional breeding programmes. Creating new rice varieties with high levels of either polyglutamylated folate or folate binding proteins in order to increase folate stability and bioavailability has yet to be established in rice endosperm. This thesis has successfully cloned and transformed rice to generate new rice varieties. Nevertheless, further analysis in terms of the folate abundance and bioavailability will have to be explored in the coming years.

REFERENCES

Anguera, M.C., Suh, J.R., Ghandour, H., Nasrallah, I.M., Selhub, J. and Stover, P.J. (2003) Methenyltetrahydrofolate synthethase regulates folate turnover and accumulation. *Journal of Biological Chemistry* **278**: 29856-29862.

Antony, A.C., Utley, C., Van Horne, K.C. and Kolhouse, J.F. (1981) Isolation and characterisation of a folate receptor from human placenta. *Journal of Biological Chemistry* **256**: 9684-9692.

Antony, A.C., Verma, R.S., Unune, A.R. and LaRosa, J.A. (1989) Identification of a Mg^{2+} dependent protease in human placenta which cleaves hydrophobic folate binding proteins to hydrophilic forms. *Journal of Biological Chemistry* **264**: 1911-1914.

Appling, D.R. (1991) Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. *FASEB Journal* **5**: 2645-2651.

Assantachai, P. and Lekhakula, S. (2005) Epidemiological survey of vitamin deficiencies in older Thai adults: implications for national policy planning. *Public Health Nutrition* **10**: 65-70.

Atkins, C.A., Smith, P.M.C. and Storer, P.J. (1997) Reexamination of the intracellular localization of de novo purine synthesis in cowpea nodules. *Plant Physiology* **113**: 127-135.

Auerbach, G., Herrmann, A., Bracher, A., Bader, G., Gütlich, M., Fischer, M., Neukamm, M., Garrido-Franco, M., Richardson, J., Nar, H., Huber, R. and Bacher, A. (2000) Zinc plays a key role in GTP cyclohydrolase I. *Proc. Natl Acad. Sci. USA* **97**: 13567–13572.

Bailey, L.B. (1998) Dietary reference intakes for folate: the debut of dietary folate equivalents. *Nutrition Reviews* **56**: 294-299.

Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* **10**: 411-421.

Barlowe, C.K. and Appling, D.R. (1988) In vitro evidence for the involvement of mitochondrial folate metabolism in the supply of cytoplasmic one-carbon units. *Biofactors* **1**: 171-176.

Basset, G.J., Quinlivan, E.P., Ziemak, M.J., Díaz de la Garza, R., Fischer, M., Schiffmann, S., Bacher, A., Gregory, J.F. III and Hanson, A.D. (2002) Folate synthesis in plants: the first step of the pterin branch is mediated by a unique bimodular GTP cyclohydrolase I. *Proc Natl Acad Sci USA* **99**: 12489–12494.

Basset, G.J., Quinlivan, E.P., Ravanel, S., Rébeillé, F., Nichols, B.P., Shinozaki, K., Seki, M., Adams-Phillips, L.C., Giovannoni, J.J., Gregory, J.F. III and Hanson, A.D. (2004a) Folate synthesis in plants: the p-aminobenzoate branch is initiated by a bifunctional PabA–PabB protein that is targeted to plastids. *Proc Natl Acad Sci USA* **101**: 1496–1501.

Basset, G.J., Ravanel, S., Quinlivan, E.P., White, R., Giovannoni, J.J., Rébeillé, F., Nichols, B.P., Shinozaki, K., Seki, M., Gregory, J.F. III and Hanson, A.D. (2004b) Folate synthesis in plants: The last step of the p-aminobenzoate branch is catalyzed by a plastidial aminodeoxychorismate lyase *The Plant Journal* **40**: 453-61.

Basset, G.J., Quinlivan, E.P., Gregory, J.F. III and Hanson, A.D. (2005) Folate synthesis and metabolism in plants and prospects for biofortification. *Crop Science* **45**: 449-453.

Bassette, P.H., Aslund, F., Beckwith, J. and Georgiou, G. (1999) Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proc Natl Acad Sci USA* **17**: 13703-13708.

Batra, K.K., Wagner, J.R. and Stokstad, E.L.R. (1977) Folic acid compounds in romaine lettuce. *Canadian Journal of Biochemistry* **55**: 865-868.

Bekaert, S., Storozhenko, S., Mehrshahi, P., Bennett, M.J., Lambert, W., Gregory III, J.F., Schubert, K., Hugenholtz, J., van der Straeten, D. and Hanson, A.D. (2007) Folate biofortification in food plants. *Trends in Plant Science* **13**: 28-35.

Bender, D.A. (2003) *Nutritional Biochemistry of the Vitamins*. Cambridge: Cambridge University Press.

Besson, V., Rebeille, F., Neuburger, M., Douce, R. and Cossin, E.A. (1993) Effects of tetrahydrofolate polyglutamates on the kinetic parameters of serine hydroxymethyltransferase and glycine decarboxylase from pea leaf mitochondria. *Biochemistry Journal* **292**: 425-430.

Blount, B.C., Mack, M.M., Wehr, C.M., MacGregor, J.T., Hiatt, R.A., Wang, G., Wickramasinghe, S.N., Everson, R.B. and Ames, B.N. (1997) Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. *Proc Natl Acad Sci USA* **94**: 3290-3295.

Blumenstein, J. and Williams, G.R. (1960) The enzymatic N-methylation of glycine. *Biochemical and Biophysical Research Communications* **3**: 259-263.

Bognar, A.L., Osborne, C., Shane, B., Singer, S.C. and Ferone, R. (1985) Folylpoly-gamma-glutamate synthetase-dihydrofolate synthetase. Cloning and high expression of the *Escherichia coli folC* gene and purification and properties of the gene product. *The Journal of Biological Chemistry* **260**: 5625-5630.

Bourguignon, J., Neuburger, M. and Douce, R. (1988) Resolution and characterization of the glycine-cleavage reaction in pea leaf mitochondria. Properties of the forward reaction catalysed by glycine decarboxylase and serine hydroxymethyltransferase. *Biochemistry Journal* **255**: 169-178.

Bozzo, G.G., Basset, G.J.C, Naponelli, V., Noiriel, A., Gregory III. J.F. and Hanson, A.D. (2008) Characterization of the folate salvage enzyme *p*-aminobenzoylglutamate hydrolase in plants. *Phytochemistry* **69**: 29-37.

Brown, J.P., Scott, J.M., Foster, F.G. and Weir, D.G. (1973) Ingestion and absorption of naturally occurring pteroylmonoglutamates (folates) in man. *Gastroenterology* **64**: 223-232.

Chan, P.Y., Coffin, J.W. and Cossins, E.A. (1986) In vitro synthesis of Pteroylpoly-γ-glutamates by cotyledon extracts of *Pisum sativum* L. *Plant and Cell Physiology* **27**: 431-441.

Chan, A., Robertson, E.F., Haan, E.A., Keane, R.J., Ranieri, E. and Carney, A. (1993) Prevalence of neural tube defects in south Australia, 1966-1991: effectiveness and impact of prenatal diagnosis. *British Medical Journal* **307**: 703-706.

Chan, S.Y. and Cossins, E.A. (2003) The intracellular distribution of folate derivatives in pea leaves. *Pteridines* **14**: 17-26.

Chasin, L.A., Feldman, A., Konstam, M. and Urlaub, G. (1974) Reversion of a Chinese hamster cell auxotrophic mutant. *Proc Natl Acad Sci USA* **71**: 718-722.

Chen, L., Qi, H., Korenberg, J., Garrow, T.A., Choi, Y.J. and Shane, B. (1996) Purification and properties of human cytosolic folylpoly-γ-glutamate synthethase and organization, localization, and differential splicing of its gene. *Journal of Biological Chemistry* **271**: 13077-13087.

Chen, L., Chan, S.Y. and Cossins, E.A. (1997) Distribution of folate derivatives and enzymes for synthesis of 10-formyltetrahydrofolate in cytosolic and mitochondrial fractions of pea leaves. *Plant Physiology* **115**: 299-309.

Cherest, H., Thomas, D. and Surdin-Kerjan, Y. (2000) Polyglutamylation of folate coenzymes is necessary for methionine biosynthesis and maintenance of intact mitochondrial genome in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **275**: 14056-14063.

Christensen, K.E. and MacKenzie, R.E. (2006) Mitochondria one-carbon metabolism is adapted to the specific needs of yeast, plants and mammals. *BioEssays* **28**: 595-605.

Cichowicz, D.J. and Shane, B. (1987) Mammalian folyl-γ-polyglutamate synthethase. 1. Purification and general properties of the hog liver enzyme. *Biochemistry* **26**: 504-512.

Clifford, A.J., Heid, M.K., Muller, H.G. and Bills, N.D. (1990) Tissue distribution and prediction of total body folate of rats. *Journal of Nutrition* **120**: 1633-1639.

Cole, B.F., Baron, J.A., Sandler, R.S., Haile, R.W., Ahnen, D.J., Bresalier, R.S., McKeown,-Eyssen, G., Summers, R.W., Rothstein, R.I., Burke, C.A., Snover, D.C., Church, T.R., Allen, J.I., Robertson, D.J., Beck, G.J., Bond, J.H., Byers, T., Mandel, J.S., Mott, L.A., Pearson, L.H., Barry, E.L., Rees, J.R., Marcon, N., Saibil, F., Ueland, P.M., Greenberg, E.R. and Polyp Prevention study group (2007) Folic acid for the prevention of colorectal adenomas: a randomized clinical trial. *Journal of American Medical Association* **297**: 2351-2359.

Collakova, E., Goyer, A., Naponelli, V., Krassovskaya, I., Gregory, J.F. III, Hanson, A.D. and Shachar-Hill, Y. (2008) *Arabidopsis* 10-formyl tetrahydrofolate deformylases are essential for photorespiration. *Plant Cell* **20**: 1818-1832.

Colman, N., Hettiarachchy, N. and Herbert, V. (1981) Detection of a milk factor that facilitates folate uptake by intestinal cells. *Science* **211**: 1427-1429.

Copp, A.J. (1993) Neural tube defects. *Trends in Neurosciences* **16**: 381-383.

Cossins, E.A. (1987) *The Biochemistry of Plants*. New York: Academic Press.

Cossins, E. and Chen, L. (1997) Folates and one-carbon metabolism in plants and fungi. *Phytochemistry* **45**: 437-452.

Cossins, E.A. and Shuh, S.P.J. (1972) Pteroylglutamates of higher plant tissues. *Phytochemistry* **11**: 587-593.

Counce, P.A., Keisling, T.C. and Mitchell, A.J. (2000) A uniform, objective, and adaptive system for expressing rice development. *Crop Science* **40**: 436-443.

Cregg, J.M., Cereghino, J.L., Shi, J. and Higgins, D.R. (2000) Recombinat protein expression in *Pichia pastoris*. *Molecular Biotechnology* **16**: 23-52.

Cuskelly, G.J., McNulty, H. and Scott, J.M. (1996) Effect of increasing dietary folate on red-cell folate: implications for prevention of neural tube defects. *Lancet* **347**: 657-659.

Cybulski, R.L. and Fisher, R.R. (1976) Intramitochondrial localization and proposed metabolic significance of serine transhydroxymethylase. *Biochemistry* **15**: 3183-3187.

Cybulski, R.L. and Fisher, R.R (1981) Uptake of oxidized folates by rat liver mitochondria. *Biochimica et Biophysica Acta* **646**: 329-333.

Czeizel, A.E. and Dudás, I. (1992) Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *New England Journal of Medicine* **327**: 1832-1835.

De Buck, S., De Wilde, C., Montagu, M.V. and Depicker, A. (2000) T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by *Agrobacterium*-mediated transformation. *Molecular Breeding* **6**: 459-468.

DeSouza, L., Shen, Y. and Bognar, A.L. (2000) Disruption of cytoplasmic and mitochondrial folylpolyglutamate synthetase activity in *Saccharomyces cerevisiae*. *Archives of Biochemistry and Biophysics* **376**: 299–312.

Díaz de la Garza, R., Quinlivan, E.P., Klaus, S.M.J., Basset, G.J.C., Gregory, J.F.III and Hanson, A.D. (2004) Folate biofortification in tomatoes by engineering the pteridine branch of folate synthesis. *Proc Natl Acad Sci USA* **101**: 13720-13725.

Díaz de la Garza, R., Gregory, J.F.III and Hanson, A.D. (2007) Folate biofortification of tomato fruit. *Proc Natl Acad Sci USA* **104**: 4218-4222.

Dolk, H., Dewals, P., Gillerot, Y., Lechat, M.F., Ayme, S., Cornel, M., Cuschieri, A., Garne, E., Goujard, J., Laurence, K.M., Lillis, D., Lys, F. (1991) Heterogeneity of neural tube defects in Europe: the significance of site of defect and presence of other major anomalies in relation to geographic differences in prevalence. *Teratology* **44**: 547-559.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehna, T., Song, K. and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *The Plant Journal* **45**: 616-629.

Edwards, R. (1995) Determination of S-adenosyl-L-methionine and S-adenosyl-L-homocysteine in plants. *Phytochemical Analysis* **6**: 25-30.

Eikelboom, J.M., Lonn, E., Genest, J.Jr., Hankey, G. and Yusuf, S. (1999) Homocysteine and cardiovascular disease: a critical review of the epidemiologic evidence. *Annals of Internal Medicine* **131**: 363-375.

Eisenga, B.H., Collins, T.D. and McMartin, K.E. (1992) Incorporation of ^{3}H-Label from Folic Acid is Tissue-Dependent in Folate-Deficient Rats. *Journal of Nutrition* **122** : 977-985.

Elnakat, H. and Ratnam, M. (2004) Distribution, functionality and gene regulation of folate receptor isoforms: implications in targeted therapy. *Advanced Drug Delivery Reviews* **56**: 1067-1084.

Elwood, P.C. (1989) Molecular cloning and characterization of the human folate binding protein cDNA from placenta and malignant tissue culture (KB) cells. *Journal of Biological Chemistry* **264**: 14893-14901.

Ericson, U., Sonestedt, E., Gullberg, B., Olssan, H. and Wirfält, H. (2007) High folate intake is associated with lower breast cancer incidence in postmenopausal women in the Malmo Diet and Cancer cohort. *American Journal of Clinical Nutrition* **86**: 434-443.

Fenech, M., Noakes, M., Clifton, P. and Topping, D. (1999) Aleurone flour is a rich source of Bioavailable folate in humans. *Journal of Nutrition* **129**: 1114-1119.

Finglas, P.M., de Meer, K., Molloy, A., Verhoef, P., Pietrzik, K., Powers, H.J., van der Straeten, D., Jägerstad, M., Varela-Moreiras, G., van Vliet, T., Havenaar, R., Buttriss, J. and Wright, A.J.A. (2006) Research goals for folate and related B vitamin in europe. *European Journal of Clinical Nutrition* **60**: 287-294.

Finglas, P.M. and Morgan, M.R.A. (1994) Application of biospecific methods to the determination of B-group vitamins in food-a review. *Food Chemistry* **49**: 191-201.

Fitzgerald, M.A., McCouch, S.R. and Hall, R.D. (2009) Not just a grain of rice: the quest for quality. *Trends in Plant Science* **14**: 133-139.

Fleming, A. and Copp, A.J. (1998) Embryonic folate metabolism and mouse neural tube defects. *Science* **280**: 2107-2109.

Freemantle, S.J., Taylor, S.M., Krystal, G. and Moran, R.G. (1995) Upstream organization of and multiple transcripts from the human folylpoly-γ-glutamate synthase gene. *Journal of Biological Chemistry* **270**: 9579–9584.

Frey, L. and Hauser, W.A. (2003) Epidemiology of neural tube defects. *Epilepsia* **44**: 4-13.

Gallardo, K., Job, C., Groot, S.P.C., Puype, M., Demol, H., Vandekerckhove, J. and Job, D. (2002) Importance of methionine biosynthesis for *Arabidopsis* seed germination and seedling growth. *Physiologia Plantarum* **116**: 238-247.

Gambonnet, B., Jabrin, S., Ravanel, S., Karan, M., Douce, R. and Rébeillé, F. (2001) Folate distribution during higher plant development. *Journal of the Science of Food and Agriculture* **81**: 835-841.

Garcia-Casal, M.N., Osorio, C., Landaeta, M., Leets, I., Matus, P., Fazzino, F. and Marcos, E. (2005) High prevalence of folic acid and vitamin B12 deficiencies in infants, children, adolescents and pregnant women in Venezuela. *European Journal of Clinical Nutrition* **59**: 1064-1070.

Geisel, J. (2003) Folic acid and neural tube defects in pregnancy-a review. *Journal of Perinatal and Neonatal Nursing* **17**: 268-279.

Gerhard, G.T. and Duell, P.B. (1999) Homocysteine and atherosclerosis. *Current Opinion in Lipidology* **10**: 417-428.

Giovannucci, E. (2002) Epidemiologic studies of folate and colorectal neoplasia: a review. *Journal of Nutrition* **132**: 2350S-2355S.

Giovanelli, J., Mudd, S.H. and Datko, A.H. (1985) Quantitative analysis of pathways of methionine metabolism and their regulation in *Lemna*. *Plant Physiology* **78**: 555-560.

Godwin, H.A. and Rosenberg, I.M. (1975) Comparative studies on the intestinal absorption of [^{3}H] pteroylmonoglutamate on [^{3}H] pteroylheptaglutamate in man. *Gastroenterology* **68**: 364-373.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R.L., Dunn, M *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp japonica). *Science* **296**: 92-100.

Goyer, A. and Navarre, D.A. (2009) Folate is higher in developmentally younger potato tubers. *Journal of the Science of Food and Agriculture* **89**: 579-583.

Graham, R.D., Welch, R.M. and Bouis, H.E. (2001) Addressing micronutrient malnutrition through enhancing the nutritional quality of staple foods: principles, perspectives and knowledge gaps. *Advances in Agronomy* **70**: 77-142.

Green, J.M., Merkel, W.K. and Nichols, B.P. (1992) Characterization and sequence of *Escherichia coli pabC*, the gene encoding aminodeoxychorismate lyase, a pyridoxal phosphate-containing enzyme. *Journal of Bacteriology* **174**: 5317-5323.

Gregory, J.F. (1997) Bioavailability of folate. *European Journal of Clinical Nutrition* **51**: S54-S59.

Han, Y.H., Yon, M. and Hyun, T.H. (2005) Folate intake estimated with an updated database and its association to blood folate and homocysteine in Korean college students. *European Journal of Clinical Nutrition* **59**: 246-254.

Hanson, A.D. and Gregory, J.F. III (2002) Synthesis and turnover of folates in plants. *Current Opinion in Plant Biology* **5**: 244-249.

Hanson, A.D. and Roje, S. (2001) One-carbon metabolism in higher plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **52**: 119-137.

Heineke, D., Bykova, N., Gardeström, P. and Bauwe, H. (2001) Metabolic response of potato plants to an antisense reduction of the P-protein of glycine decarboxylase. *Planta* **212**: 880-887.

Henderson, G. B. (1990) Folate-binding proteins. *Annual Review of Nutrition* **10**: 319-335.

Hiei, Y., Ohta, S., Komari, T. and Kumashiro, T. (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *The Plant journal* **6**: 271-282.

Hiei, Y., Komari, T. and Kubo, T. (1997) Transformation of rice mediated by Agrobacterium tumefaciens. *Plant Molecular Biology* **35**: 205-218.

Hiei, Y. and Komari, T. (2008) *Agrobacterium*-mediated transformation of rice using immature embryos or calli induced from mature seed. *Nature Protocols* **3**: 824-834.

Hilton, J.G., Cooper, B.A. and Rosenblatt, D.S. (1979) Folate polyglutamate synthesis and turnover in cultured human fibroblasts. *Journal of Biological Chemistry* **254**: 8398- 8403.

Horne, D.W. and Holloway, R.S. (1997) Compartmentation of folate metabolism in rat pancreas: nitrous oxide inactivation of methionine synthase leads to accumulation of 5-methyltetrahydrofolate in cytosol. *Journal of Nutrition* **27**: 1772-1775.

Horne, D.W., Patterson, D. and Cook, R. (1989) Effect of nitrous oxide inactivation of vitamin B12-dependent methionine synthethase on the subcellular distribution of folate coenzymes in rat liver. *Archives of Biochemistry and Biophysics* **270**: 729-733.

Hossain, T., Rosenberg, I. Selhub, J., Kishore, G., Beachy, R. and Schubert, K. (2004) Enhancement of folates in plants through metabolic engineering. *Proc Natl Acad Sci USA* **101**: 5158-5163.

Huang, J., Wu, L., Yalda, D., Adkins, Y., Kelleher, S.L., Crane, M., Lonnerdal, B., Rodriguez, R.L. and Huang, N. (2002) Expression of functional recombinant human lysozyme in transgenic rice cell culture. *Transgenic Research* **11**: 229-239.

Huangpu, J., Pak, J.H., Graham, M.C., Rickle, S.A. and Graham, J.S. (1996) Purification and molecular analysis of an extracellular γ–glutamyl hydrolase present in young tissues of the soybean plant. *Biochemical and Biophysical Research Communications* **228**: 1-6.

Imeson, H., Zheng, L. and Cossins, E.A. (1990) Folylpolyglutamate derivatives of *Pisum sativum* L. Determination of polyglutamate chain lengths by high performance liquid chromatography following conversion to *p*-aminobenzoylpolyglutamates. *Plant Cell Physiology* **31**: 223-231.

Iwai, K., Tani, M. and Fushiki, T. (1983) Electrophoretic and immunological properties of folate-binding protein isolated from bovine milk. *Agricultural Biology and Chemistry* **47**: 1523-1530.

Jabrin, S., Ravanel, S., Gambonnet, B., Douce, R. and Rébeillé, F. (2003) One-carbon metabolism in plants. Regulation of tetrahydrofolate synthesis during germination and seedling development. *Plant Physiology* **131**: 1431-1439.

Jägerstad, M., Jastrebova, J. and Svensson, U. (2004) Folates in fermented vegetables-a pilot study. *Lebensmittel-Wissenschaft und-Technologie* **37**: 603-611.

Jeong, D.H., An, S., Kang, H.G., Moon, S., Han, J.J., Park, S., Lee, H.S., An, K. and An, G. (2002) T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiology* **130**: 1636-1644.

Jeong, J. and Guerinot, M.L. (2008) Biofortified and bioavailable: the gold standard for plant-based diets. *Proc Natl Acad Sci USA* **105**: 1777-1778.

Jones, M.L. and Nixon, P.F. (2002) Tetrahydrofolates are greatly stabilized by binding to bovine milk folate-binding protein. *Journal of Nutrition* **132**: 2690-2694.

Jones, M.L., Treloar, T. and Nixon, P.F. (2003) Dietary interactions influence the effects of bovine folate-binding protein on the bioavailability of tetrahydrofolates in rats. *Journal of Nutrition* **133**: 489-495.

Jönsson, U., Fägerstam, L., Ivarsson, B., Johnsson, B., Karlson, R., Lundh, K., Löfas, S., Persson, B., Roos, H., Rönnberg, I. *et al.* (1991) Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. *Biotechniques* **11**: 620-627.

Juliano, B.O. (1985) *Rice: chemistry and technology*. St Paul, MN: American Association of Cereal Chemists.

Kamen, B.A. and Capdevila, A. (1986) Receptor-mediated folate accumulation is regulated by the cellular folate content. *Proc Natl Acad Sci USA* **83**: 5983-5987.

Karlsson, R. and Fält, A. (1997) Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors. *Journal of Immunological Methods* **200**: 121-133.

Kennedy, G. and Burlingame, B. (2003) Analysis of food composition data on rice from a plant genetic resources perspective. *Food Chemistry* **80**: 589-596.

Kerr, S.J. (1972) Competing Methyltransferase Systems. *The Journal of Biological Chemistry* **247**: 4248-4252.

Kim, J.S., Lowe, K.E. and Shane, B. (1993) Regulation of one-carbon metabolism in mammalian cells. *Journal of Biological Chemistry* **268**: 21680-21685.

Kirk, C.D., Chen, L., Imeson, H.C. and Cossins, E.A. (1995) A 5,10-methylenetetrahydrofolate dehydrogenase: 5,10-methenyltetrahydrofolate cyclohydrolase protein from Pisum sativum. *Growth and Metabolism* **39**: 1309-1317.

Kohashi, M. (1980) Isolation of six unconjugated pteridines from soybean (*Glycin max L. Tsurunoko*) seeds. *Journal Biochemistry* **87**: 1581-1586.

Lacey, S.W., Sanders, J.M., Rothberg, K.G., Anderson, R.G.W. and Kamen, B.A. (1989) Complementary DNA for the folate binding protein correctly predicts anchoring to the membrane by glycosyl-phosphatidylinositol. *Journal of Clinical investigation* **84**: 715-720.

Laemmli, U.K. and Favre, M. (1973) Maturation of the head of bacteriophage T4. *Journal of Molecular Biology* **80**: 575-599.

Leech, R.W. and Payne, GG Jr. (1991) Neural tube defects: epidemiology. *Journal of Child Neurology* **6**: 286-287.

Lin, B.F., Huang, R.F.S. and Shane, B. (1993) Regulation of folate and one-carbon metabolism in mammalian cells: III. Role of mitochondrial folylpoly-γ-glutamate synthetase. *Journal of Biological Chemistry* **268**: 21674-21679.

Lin, B.F. and Shane, B. (1994) Expression of *Escherichia coli* folylpolyglutamate synthethase in the Chinese hamster ovary cell mitochondrion. *Journal of Biological Chemistry* **269**: 9705-9713.

Löfas, S. and Johnsson, B.J. (1990) A novel hydrogel matrix on gold surfaces in surface plasmon resonance sensors for fast and efficient covalent immobilization of ligands. *Journal of the Chemical Society-Chemical Communications* **121**: 1526-1528.

Lorimer, G.H. (1981) The carboxylation and oxygenation of ribulose 1,5-bisphosphate: the primary events in photosynthesis and photorespiration. *Annual Review of Plant Physiology* **32**: 349-383.

Lowe, K.R., Osborne, C.B., Lin, B.F., Kim, J.S., Hsu, J.C. and Shane, B. (1993) Regulation of folate and one-carbon metabolism in mammalian cells II. Effect of folypoly-γ-glutamate synthetase substrate specificity and level on folate metabolism and folypoly- γ-glutamate specificity of metabolic cycles of one-carbon metabolism. *The Journal of Biological Chemistry* **268**: 21665-21673.

Lucca, P., Hurrell, R. and Potrykus, I. (2002) Fighting iron deficiency anemia with iron-rich rice. *Journal of the American College of Nutrition* **21**: 184S-190S.

Lucca, P., Polett, S. and Sautter, C. (2006) Genetic engineering approaches to enrich rice with iron and vitamin A. *Physiologia Plantarum* **126**: 291-303.

Luka, Z., Pakhomova, S., Loukachevitch, L.V., Egli, M., Newcomer, M.E. and Wagner, C. (2007) 5-Methyltetrahydrofolate is bound in intersubunit areas of rat liver folate-binding protein glycine N-methyltransferase. *Journal of Biological Chemistry* **282**: 4069-4075.

Luka, Z. and Wagner, C. (2003) Expression and purification of glycine N-methyltransferases in *Escherichia coli*. *Protein Expression and Purification* **28**: 280-286.

Luo, M.Z., Orsi, R., Patrucco, E., Pancaldi, S. and Cella, R. (1997) Multiple transcription start sites of the carrot dihydrofolate reductase thymidylate synthase gene, and sub-cellular localization of the bifunctional protein. *Plant molecular Biology* **33**: 709-722.

Makarov, A.D. and Stakhov, L.F. (1972) Chemical nature and characteristics of phosphodoxin, a physiological co-factor of photophosphorylation. *Biofizika* **17**: 396-400.

Mason, J.B. and Selhub, J. (1988) Folate-binding protein and the absorption of folic acid in the small intestine of the suckling rat. *American Journal of Clinical Nutrition* **48**: 620-625.

McCullough, J.L., Chabner, B.A and Bertino, J.R. (1971) Purification and properties of carboxypeptidase G_1. *The Journal of Biological Chemistry* **246**: 7207-7213.

McGuire, J.J., Hsiesh, P., Coward, J.K. and Bertino, J.R. (1980) Enzymatic synthesis of folylpolyglutamates. Characterization of the reaction and its products. *The Journal of Biological Chemistry* **255**: 5776-5788.

McIntosh,S.R. and Henry, R.J. (2008) Genes of folate biosynthesis in wheat. *Journal of Cereal Science* **48**: 632-638.

McIntosh, S.R., Brushett, D. and Henry, R.J. (2008) GTP cyclohydrolase I expression and folate accumulation in the developing wheat seed. *Journal of Cereal Science* **48**: 503-512.

McKillop, D.J., Pentieva, K., Daly, D., McPartlin, J.M., Hughes, J., Strain, J.J., Scott, J.M. and McNulty, H. (2002) The effect of different cooking methods on folate retention in various foods that are amongst the major contributors to folate intake in the UK diet. *British Journal of Nutrition* **88**: 681-688.

McNulty, H., McPartlin, J.M., Weir, D.G. and Scott, J.M. (1995) Folate catabolism is related to growth rate in weanling rats. *Journal of Nutrition* **125**: 99-103.

McNulty, H. and Scott, J.M. (2008) Intake and status of folate and relates B-vitamin: considerations and challenges in achieving optimal status. *British Journal of Nutrition* **99**: S48-S54.

Min, H., Shane, B. and Stokstad, E.L.R (1988) Identification of 10-formyltetrahydrofolate dehydrogenase-hydrolase as a major folate binding protein in liver cytosol. *Biochimica et Biophysica Acta* **967**: 348-353.

Moffatt, B.A. and Weretilnyk, E.A. (2001) Sustaining S-adenosyl-L-methionine-dependent methyltransferase activity in plant cells. *Physiologia Plantarum* **113**: 435-442.

Moran, R.G., Werkheiser, W.C. and Zerkrzewski, S.F. (1976) Folate metabolism in mammalian cells in culture. I Partial characterization of the folate derivatives present in L1210 mouse leukemia cells. *Journal of Biological Chemistry* **251**: 3569-3579.

Mouillon, J.M., Ravanel, S., Douce, R. and Rébeillé, F. (2002) Folate synthesis in higher-plant mitochondria: coupling between the dihydropterin pyrophosphokinase and the dihydropteroate synthase activities. *Biochemistry Journal* **363**: 313–319.

Myszka, D.G. (1997) Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Current Opinion in Biotechnology* **8**: 50-57.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T. and Kimura, T. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of Bioscience and Bioengineering* **104**: 34-41.

Neuburger, M., Rébeillé, F., Jourdain, A., Nakamura, S. and Douce, R. (1996) Mitochondria are a major site for folate and thymidylate synthesis in plants. *The Journal of Biological Chemistry* **271**: 9466-9472.

Noiriel, A., Naponelli, V., Bozzo, G., Gregory III, J.F. and Hanson, A.D. (2007) Folate salvage in plants: pterin aldehyde reduction is mediated by multiple non-specific aldehyde reductases. *The Plant Journal* **51**: 378-389.

Nygren, L., Sternesjö, A. and Björck, L. (2003) Determination of folate-binding proteins from milk by optical biosensor analysis. *International Dairy Journal* **13**: 283-290.

Nygren-Babol, L., Sternesjo, A., Jagerstad, M. and Bjorck, L. (2005) Affinity and rate constants for interactions of bovine folate-binding protein and folate derivatives determined by optical biosensor technology. Effect of Stereoselectivity. *Journal of Agricultural and Food Chemistry* **53**: 5473-5478.

Ogawa, H. and Fujioka, M. (1982) Purification and properties of glycine N-methyltransferase from rat liver. *Journal of Biological Chemistry* **257**: 3447-3452.

Ogawa, H., Gomi, T., Takata, Y., Date, T. and Fujioka, M. (1997) Recombinant expression of rat glycine N-methyltransferase and evidence for contribution of N-terminal acetylation to co-operative binding of S-adenosylmethionine. *Biochemistry Journal* **327**: 407-412.

Ogawa, H., Gomi, T., Takusagawa, F. and Fujioka, M. (1998) Structure, function and physiological role of glycine N-methyltransferase. *International Journal of Biochemisty and Cell Biology* **30**: 13-26.

Ogawa, H., Konishi, K., Takata, Y., Nakashima, H. and Fujioka, M. (1987) Rat glycine methyltransferase. Complete amino acid sequence deduced from a cDNA clone and characterization of the genomic DNA. *European Journal of Biochemistry* **168**: 141-151.

Okita, T.W., Hwang, Y.S., Hnilo, J., Kim, W.T., Aryan, A.P., Larson, R. and Krishnan, H.B. (1989) Structure and expression of the rice glutelin multigene family. *Journal of Biological Chemistry* **264**: 12573-12581.

Oliver, D.J., Neuburger, M., Bourguignon, J., and Douce, R. (1990) Interaction between the component enzymes of the glycine decarboxylase multienzyme complex. *Plant Physiology* **94**: 833-839.

Oliver, D.J. (1994) The glycine decarboxylase complex from plant mitochondria. *Annual Review of Plant Physiology and Plant Molecular Biology* **45**: 323-337.

Oliver, D.J. and Raman, R. (1995) Glycine decarboxylase: Protein chemistry and molecular biology of the major protein in leaf mitochondria. *Journal of Bioenergetics and Biomembranes* **27**: 407-414.

Orsomando, G., Diaz de la Garza, R., Green, B.J., Peng, M., Rea, P.A., Ryan, T.J., Gregory, J.F. III and Hanson, A.D. (2005) Plant γ-glutamyl hydrolases and folate polyglutamates: characterization, compartmentation and co-occurrence in vacuoles. *The Journal of Biological Chemistry* **280**: 28877–28884.

Orsomando, G., Bozzo, G.G., Diaz de la Gaiza, R., Basset, G.J., Quinlivan, E.P., Naponelli, V., Rébeillé, F., Ravanel. S., Gregory J.F. III and Hanson, A.D. (2006) Evidence for folate-salvage reactions in plants. *The Plant Journal* **46**: 426-435.

Osborne, C.B., Lowe, K.E. and Shane, B. (1993) Regulation of folate and one-carbon metabolism in mammalian cells. I. Folate metabolism in Chinese hamster ovary cells expressing *Escherichia coli* or human folylpoly-γ-glutamate synthethase activity. *Journal of Biological Chemistry* **268**: 21657-21664.

Pattanayek, R., Newcomer, M.E. and Wagner, C. (1998) Crystal structure of apo-glycine N-methyltransferase (GNMT). *Protein Science* **7**: 1326-1331.

Perry, J. and Chanarin, I. (1970) Intestinal absorption of reduced folate compounds in man. *Britain Journal of Haematology* **18**: 329-339.

Picciano, M.F., West, S.G., Ruch, A.L., Kris-Etherton, P.M., Zhao, G., Johnston, K.E., Maddox, D.H., Fishell, V.K., Dirienzo, D.B. and Tamura, T. (2004) Effect of cow milk on food folate bioavailability in young women. *American Journal of Clinical Nutrition* **80**: 1565-1569.

Prinz-Langenhe, R. and Pietrzik, K. (1998) Bioavailability of folate from spinach. *Diet and Health News* **1**: 7-8.

Puupponen-Pimiä, R., Häkkinen, S., Aarni, M., Suortti, T., Lampi, A.M. and Eurola, M. (2003) Blanching and long-term freezing affect various bioactive compounds of vegetables in different ways. *Journal of the Science of Food and Agriculture* **83**: 1389-1402.

Ravanel, S., Cherest, H., Jabrin, S., Grunwald, D., Surdin-Kerjan, Y., Douce, R. and Rébeillé, F. (2001) Tetrahydrofolate biosynthesis in plants: molecular and functional characterization of dihydrofolate synthetase and three isoforms of folylpolyglutamate synthetase in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*. **98**: 15360–15365.

Ravanel, S., Block, M.A., Rippert, P., Jabrin, S., Curien, G., Rébeillé, F. and Douce, R. (2004) Methionine metabolism in plants: chloroplasts are autonomous for de novo methionine synthesis and can import S-adenosylmethionine from the cytosol. *Journal of Biological Chemistry* **279**: 22548-22557.

Rébeillé, F., Neuburger, M. and Douce, R. (1994) Interaction between glycine decarboxylase, serine hydroxymethyltransferase and tetrahydrofolate polyglutamates in pea leaf mitochondria. *Biochemical Journal* **302**: 223-228.

Rébeillé, F., Macherel, D., Mouillon, J.M., Garin, J. and Douce, R. (1997) Folate biosynthesis in higher plants: purification and molecular cloning of a bifunctional 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase/7,8-dihydropteroate synthase localized in mitochondria. *EMBO* **3**: 947-957.

Rébeillé, F., Ravanel, S., Jabrin, S., Douce, R., Storozhenko, S. and Van Der Straeten, D. (2006) Folates in plants: biosynthesis, distribution, and enhancement. *Physiologia Plantarum* **126**: 330-342.

Roje, S., Chan, S.Y., Kaplan, F., Raymond, R.K., Horne, D.W., Appling, D.R. and Hanson, A.D. (2002) Metabolic engineering in yeast demonstrates that S-adenosylmethionine controls flux through the methylenetetrahydrofolate reductase reaction in vivo. *Journal of Biological Chemistry* **277**: 4056-4061.

Rowling, M.J. and Schalinske, K.L. (2001) Retinoid compounds activate and induce hepatic glycine N-methyltransferase in rats. *Journal of Nutrition* **131**: 1914-1917.

Sadasivan, E., Cedeno, M.M. and Rothenberg, S.P. (1994) Characterization of the gene encoding a folate-binding protein expressed in human placenta. *Journal of Biological Chemistry* **269**: 4725-4735.

Salter, D.N. and Mowlem, A. (1983) Neonatal role of milk folate-binding protein: studies on the course of digestion of goat's milk folate binder in the 6-d-old kid. *British Journal of Nutrition* **50**: 589-596.

Salter, D.N., Scott, K.J., Slade, H. and Andrews, P. (1981) The preparation and properties of folate-binding protein from cow's milk. *Biochemistry Journal* **193**: 469-476.

Saraste, M., Sibbald, P. and Wittinghofer, A. (1990) The P-loop- a common motif in ATP- and GTP-binding proteins. *Trends in Biochemical Sciences* **15**: 430-434.

Schein, C.H. (1989) Production of soluble recombinant proteins in bacteria. *BioTechnology* **7**: 1141-1148.

Schirch, L.V. and Peterson, D. (1980) Purification and properties of mitochondrial serine hydroxymethyltransferase. *Journal of Biological Chemistry* **255**: 7801-7806.

Schneider, E. and Ryan, T. (2006) Gamma-glutamyl hydrolase and drug resistance. *Clinica Chimica Acta* **374**: 25-32.

Schoedon, G., Redweik, U., Frank, G., Cotton, R. and Blau, N. (1992) Allosteric characteristics of GTP cyclohydrolase I from *Escherichia coli*. *European Journal of Biochemistry* **210**: 561-568.

Scott, J., Rébeillé, F. and Fletcher, J. (2000) Folic acid and folates: the feasibility for nutritional enhancement in plant foods. *Journal of the Science of Food and Agriculture* **80**: 795-824.

Seshadri, S., Beiser, A., Selhub, J., Jacques, R.F., Rosenberg, I.H., D Agostino, R.B., Wilson, P.W.F., Wolf, P.A. (2002) Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. *The New England Journal of Medicine* **346**: 476-483.

Shane, B. (1989) Folylpolyglutamate synthesis and role in the regulation of one-carbon metabolism. *Vitamins and Hormones* **45**: 263-335.

Shaw, M.K. and Ingraham, J.L. (1967) Synthesis of macromolecules by *Escherichia coli* near the minimal temperature for growth. *Journal of Bacteriology* **94**: 157-164.

She, Q.B., Nagao, I., Hayakawa, T. and Tsuge, H. (1994) A simple HPLC method for the determination of S-adenosylmethionine and S-adenosylhomocysteine in rat tissues. The effect of vitamin B6 deficiency on these concentrations in rat liver. *Biochemical and Biophysical Research Communications* **205**: 1748-1754.

Shin, Y.S., Kim, E.S., Watson, J.E. and Stokstad, E.L.R. (1975) Studies of folic acid compounds in nature. IV. Folic acid compounds in soybeans and cow milk. *Canadian Journal of Biochemistry* **53**: 338-343.

Slock, J., Stahly, D.P., Han, C., Six, E.W. and Crawford, I.P. (1990) An apparent *Bacillus subtilis* folic acid biosynthesis operon containing *pab*, an amphibolic *trpG* gene, a third gene required for synthesis of *para*-aminobenzoic acid, and the dihydropteroate synthase gene. *Journal of Bacteriology* **172**: 7211-7226.

Somerville, C.R. (2001) An early *Arabidopsis* demonstration resolving a few issues concerning photorespiration. *Plant Physiology* **125**: 20–24.

Sørensen, H.P. and Mortensen, K.K. (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microbial Cell Factories* **4**: 1-8.

Staben, C. and Rabinowitz, J.C. (1984) Formation of formylmethionyl-tRNA and initiation of protein synthesis. *Folates and Pterins*. New York: Wiley.

Steinberg, S.E. (1984) Mechanisms of folate homeostasis. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **246**: G319-G324.

Steinberg, S.E., Campbell, C.L. and Hillman, R.S. (1979) Kinetics of the normal folate enterohepatic cycle. *Journal of Clinical Investigation* **64**: 83-88.

Stea, T.H., Johansson, M., Jägerstad, M. and Frølich, W. (2006) Retention of folates in cooked, stored and reheated peas, broccoli and potatoes for use in modern large-scale service system. *Food Chemistry* **101**: 1095-1107.

Steinmetz, K.L., Pogribny, I.P., James, S.J. and Pitot, H.C. (1998) Hypomethylation of the rat glutathione S-transferase pi (GSTP) promoter region isolated from methyl-deficient livers and GSTP-positive liver neoplasms. *Carcinogenesis* **19**: 1487-1494.

Stenberg, E., Persson, B., Roos, H., Urbaniczky, C. (1991) Quantitative determination of surface concentration of protein with surface plasmon resonance by using radiolabeled proteins. *Journal of Colloid and Interface Science* **143**: 513-526.

Strong, W.B., Tendler, S.J., Seither, R.L., Goldman, I.D. and Schirch, V. (1990) Purification and properties of serine hydroxymethyltransferase and C1-tetrahydrofolate synthase from L1210 cells. *Journal of Biological Chemistry* **265**: 12149-12155.

Stone, D.H. (1987) The declining prevalence of anencephalus and spina bifina: It's nature, causes and implications. *Developmental Medicine and Child Neurology* **29**: 541-546.

Storozhenko, S., Ravanel, S., Zhang, G.F., Rébeillé, F., Lambert, W. and Van Der Straeten, D. (2005) Folate enhancement in staple crops by metabolic engineering. *Trends in Food Science and Technology* **16**, 271–281.

Storozhenko, S., Brouwer, V.D., Volckaert, M., Navarrete, O., Blancquaert, D., Zhang, G.F., Lambert, W. and Van Der Straeten, D. (2007) Folate fortification of rice by metabolic engineering. *Nature Biotechnology* **25**, 1277–1279.

Stover, P. and Schirch, V. (1992) The metabolic role of leucovorin. *Trends in Biochemical Sciences* **18**: 102-106.

Stryer, L. (1995) Metabolic energy: generation and storage. *Biochemistry*. New York: W.H. Freeman and Company.

Suh, J.R., Oppenheim, E.W., Girgis, S. and Stover, P.J. (2000) Purification and properties of a folate-catabolizing enzyme. *Journal of Biological Chemistry* **275**: 35646-35655.

Suh, J.R., Herbig, A.K. and Stover, P. (2001) New perspectives on folate catabolism. *Annual Review of Nutrition* **21**: 255-282.

Sun, X., Bognar, A.L., Baker, E.N. and Smith, C.A. (1998) Structural homologies with ATP- and folate-binding enzymes in the crystal structure of folylpolyglutamate synthethase. *Proc Natl Acad Sci USA* **95**: 6647-6652.

Sun, X., Cross, J.A., Bognar, A.L., Baker, E.N. and Smith, C.A. (2001) Folate-binding triggers the activation of folylpolyglutamate synthethase. *Journal of Molecular Biology* **310**: 1067-1078.

Svendsen, I., Hanson, S.I., Holm, J. and Lyngbye, J. (1982) Amino acid sequence homology between human and bovine low molecular weight folate binding protein isolated from milk. *Calsberg Research Communications* **47**: 371-376.

Svendsen, I., Hansen, S.I., Holm, J. and Lyngbye, J. (1984) The complete amino acid sequence of the folate-binding protein from cow's milk. *Carlsberg Research Communications* **49**: 123-131.

Tanaka, H., Masuta, C., Uehara, K., Kataoka, J., Koiwai, A. and Noma, M. (1997) Morphological changes and hypomethylation of DNA in transgenic tobacco expressing antisense RNA of the S-adenosyl-l-homocysteine hydrolase gene. *Plant Molecular Biology* **35**: 981-986.

Tamura, T. and Strokstad, E.L.R. (1973) The availability of food folate in man. *British Journal of Haematology* **25**: 513-532.

Triglia, T. and Cowman, A.F. (1994) Primary structure and expression of the dihydropteroate synthetase gene of Plasmodium falciparum. *Proc Natl Acad Sci* **91**: 7149-7153.

Turner, S.R., Hellens, R., Ireland, R., Ellis, N. and Rawsthorne, S. (1993) The organisation and expression of the genes encoding the mitochondrial glycine decarboxylase complex and serine hydroxymethyltransferase in pea (*Pisum sativum*). *Molecular General Genetics* **236**: 402-408.

Usuda, K., Wada, Y., Ishimaru, Y., Kobayashi, T., Takahashi, M., Nakanishi, H., Nagato, Y., Mori, S. and Nishizawa, N.K. (2009) Genetically engineered rice containing larger amounts of nicotinamine to enhance the antihypertensive effect. *Plant Biotechnology Journal* **7**: 87-95.

van der Put, N.M.J., van Straaten, H.W.M., Trijbels, F.J.M. and Blom, H.J. (2001) Folate, homocysteine and neural tube defects: an overview. *Experimental Biology and Medicine* **226**: 243-270.

Vasconcelos, M., Datta, K., Oliva, N., Khalekuzzaman, M., Torrizo, L., Krishnan, S., Oliveira, M., Goto, F. and Datta, S.K. (2003) Enhanced iron and zinc accumulation in transgenic rice with the ferritin gene. *Plant Science* **164**: 371-378.

Vasina, J.A. and Baneyx, F. (1997) Expression of aggregation prone recombinant proteins at low temperature: a comparative study of the *Escherichia coli* cspA and tac promoter systems. *Protein Expression and Purification* **9**: 211-218.

Verma, R., Boleti, E. and George, A.J.T. (1998) Antibody engineering: comparison of bacterial, yeast, insect and mammalian expression systems. *Journal of Immunological methods* **216**: 165-181.

Verwei, M., Arkbage, K., Mocking, H., Havenaar, R. and Groten, J. (2004) The FBP binding characteristics during gastric passage are different for folic acid and 5-CH_3H_4 folate as studied in a dymanic in vitro gastrointestinal model. *Journal of Nutrition* **134**: 31-37.

Viswanathan, V.K., Green, J.M. and Nichols, B.P. (1995) Kinetic characterization of 4-amino 4-deoxychorismate synthase from *Escherichia coli*. *Journal of Bacteriology* **177**: 5918-5923.

Wagner, C., Briggs, W.T. and Cook, R.J. (1985) Inhibition of glycine N-methyltransferase activity by folate derivatives: implication for regulation of methyl group metabolism. *Biochemical and Biophysical Research Communications* **127**: 746-752.

Wang, Y., Nimec, Z., Ryan, T.J., Dias, J.A. and Galivan, J. (1993) The properties of the secreted gamma-glutamyl hydrolases from H35 hepatoma cells. *Biochemica et Biophysica Acta* **1164**: 227-235.

Ware, D.H., Jaiswal, P., Ni, J., Yap, I.V., Pan, X., Clark, K.Y., Teytelman, L., Schmidt, S.C., Zhao, W., Chang, K., Cartinhour, S., Stein, L.D. and McCouch, S.R. (2002) Gramene, a tool for grass genomics. *Plant Physiology* **130**: 1606-1613.

Waxman, S. and Schreiber, C. (1973) Characteristics of folic acid binding protein in folate-deficient serum. *Blood* **42**: 291-301.

Wei, M.M., Bailey, L.B., Toth, J.P. and Gregory, J.F. (1996) bioavailability for humans of deuterium-labeled monoglutamyl and polyglutamyl folates is affected by selected foods. *Journal of Nutrition* **126**: 3100-3108.

Welch, R.M. and Graham, R.D. (2005) Agriculture: the real nexus for enhancing bioavailable micronutrients in food crops. *Journal of Trace Elements in Medicine and Biology* **18**: 299-307.

Wingler, A., Lea, P.J., Leegood, RC. (1997) Control of photosynthesis in barley plants with reduced activities of glycine decarboxylase. *Planta* **202**: 171-178.

Wittwer, A.J. and Wagner, C. (1981) Identification of the folate binding proteins of rat liver mitochondria as dimethylglycine dehydrogenase and sarcosine dehydrogenase. Purification and folate binding characteristics. *Journal of Biological Chemistry* **256**: 4102-4108.

Yao, R., Schneider, E., Ryan, T.J. and Galivan, J. (1996a) Human gamma-glutamyl hydrolase: cloning and characterization of the enzyme expressed in vitro. *Proc Natl Acad Sci USA* **93**: 10134-10138.

Yao, R., Nimec, Z., Ryan, T.J. and Galivan, J. (1996b) Identification, cloning, and sequencing of a cDNA coding for rat gamma-glutamyl hydrolase. *Journal of Biological Chemistry* **271**: 8525-8528.

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P. Beyer, P. And Potrykus, I. (2000) Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287**: 303-305.

Yeo, E.J., Briggs, W.T. and Wagner, C. (1999) Inhibition of glycine N-methyltransferase by 5-methyltetrahydrofolatepentaglutamate. *Journal of Biological Chemistry* **274**: 37559-37564.

Yeo, E.J. and Wagner, C. (1992) Purification and properties of pancreatic glycine N-methyltransferase. *Journal of Biological Chemistry* **267**: 24669-24674.

Yoneyama, T., Wilson, L. and Hatakeyama, K. (2001) GTP cyclohydrolase I feedback regulatory protein-dependent and –independent inhibitors of GTP cyclohydrolase I. *Archives of Biochemistry and Biophysics* **388**: 67-73.

Zamierowski, M.M. and Wagner, C. (1977) Effect of folacin deficiency on folacin binding proteins in the rat. *Journal of Nutrition* **107**: 1937-1945.

Zelikson, R. and Luzzati, M. (1977) Mitochondrial and cytoplasmic distribution in Saccharmoyces cerevisiae of enzymes involved in folate-coenzyme-mediated one-carbon-group transfer. A genetic and biochemical study of the enzyme deficiencies in mutants tmp3 and ade3. *European Journal of Biochemistry* **79**: 285-292.

Zhong, R., Morrison, W.H. III., Negral, J. and Ye, Z.H. (1998) Dual methylation pathways in lignin biosynthesis. *Plant Cell* **10**: 2033-2046.

Zhu, T., Budworth, P., Chen, W., Provart, N., Chang, H.S., Guimil, S., Su, W., Estes, B., Zou, G. and Wang, X. (2003) Transcriptional control of nutrient partitioning during rice grain filling. *Plant Biotechnology Journal* **1**: 59-70.

APPENDICES

APPENDIX 1 PRIMER LISTS

Primers used for amplifying FBPs and FPGS genes

Genes	Primer names	Ta (°C)	Type	Primer sequence (5'→3')	Product size (bp)
cFBP	cFBP_f1	62	forward	AGGCTTCGCCCAGGCCCCC	n/a
	cFBP_r1		reverse	GGTCCTATCAGATCCCCTGGG	
	cFBP_f2	58	forward	GTACAAAAAAGCAGGCTTCGCCCA	n/a
	cFBP_r2		reverse	CAAGAAAGCTGGGGTCCTATCA	
	cFBP_f3	58	forward	GGGGACAAGTTTGTACAAAAAAGCAG	669
	cFBP_r3		reverse	GGGGACCACTTTGTACAAGAAAGC	
GNMT	GNMT_f1	60	forward	AGGCTTCAGGATGGTGGACAG	n/a
	GNMT_r1		reverse	GGTCCTACCAGGCTCAGCCTG	
	GNMT_f2	55	forward	GTACAAAAAAGCAGGCTTCAGGATGG	n/a
	GNMT_r2		reverse	CAAGAAAGCTGGGGTCCTACCAG	
	GNMT_f3	58	forward	GGGGACAAGTTTGTACAAAAAAGC	882
	GNMT_r3		reverse	GGGGACCACTTTGTACAAGAAAGC	
Glutelin-1 promoter	Glu_for	55	forward	CCATGGTTCTGTAGTACAGAC	851
	Glu_rev		reverse	GTCGACTGCCATGTTGTTGTA	
FPGS03g	FPGS03_f	60	forward	GTCGACATGCCTCACCGCCC	1608
	FPGS03_r		reverse	GCGGCCGCGTATCATTTCTTGACC	
FPGS10g	FPGS10_f	58	forward	GGTACCCGATCCGGCAAAAC	2057
	FPGS10_r		reverse	GGCTCGAGATGGCATCAGTGACA	

Oligonucleotide probes for amplifying a part of folate-related genes in RT-qPCR

Genes	Ta (°C)	Type	Primer sequence (5'⟶3')	Product size (bp)
LOC_Os04g56710	55	forward	GCGGAGATGGTGAAGTGATT	225
		reverse	GACCCAGAACTGCAATCGTT	
LOC_Os09g38759	55	forward	TGCTCTCATTTGCACGCACT	202
		reverse	CCTGACCTATCCGTGCTCAA	
LOC_Os08g44210	58	forward	GCGACAAGCTGATCCTGCGG	261
		reverse	GGGAACTTGAGCAGGGTTGAG	
LOC_Os06g48620	55	forward	TAGCACGTGGCAGAACACCA	222
		reverse	CCGACATTTTGGTTCCACGA	
LOC_Os05g15530	58	forward	CAAGGGGAACACCATGACCA	259
		reverse	GCCGATCCAAGTGCGAATCC	
LOC_Os02g17330	55	forward	GAAAATGGTGGATTCACCGGC	329
		reverse	ATTGGGCCTTCTTTGCCTGA	
LOC_Os07g42632	55	forward	TCCCCGGCTTATGTGACTGA	230
		reverse	TCATGGATGCGTTCATGTGG	
LOC_Os12g42870	58	forward	GATCATCAGCGCCAGAATGC	262
		reverse	AAGGCCCTCTTGGTTGGACA	
LOC_Os11g29390	58	forward	GGCGAAGAAGGTGATCTAGGTC	256
		reverse	CTGGCACGATAGCTCCCCGC	
LOC_Os12g26060	55	forward	TGCATGCTGACTACGTGGGA	347
		reverse	CGCTTCAACATGCGTACGGTAA	
LOC_Os03g02030	55	forward	ATACCACTGAACCGAACTGACC	249
		reverse	CTGAGACTGCACCATGGAAGA	
LOC_Os10g35940	55	forward	GCTACATAAGAAGCATTCTGTTACC	240
		reverse	ATGTAGCTGATCCAAAGGGCC	
LOC_Os05g44130	55	forward	CCTGATGAAAATGGCGAGGTC	206
		reverse	TCTGCGGGAGGTCTGTTCTG	
Rice actin	55	forward	TATCCCTCGTCTCGACCTTG	276
		reverse	AAGGAAGGCTGGAAGAGGAC	
Rice actin	58	forward	ATCCCTCGTCTCGACCTTGC	244
		reverse	GGAAACGCTCAGCACCAATG	

Primers used for screening T-DNA insertion lines of *FPGS* 03g02030 and *FPGS* Os10g35940 genes

Genes	Primer names	Ta (°C)	Plant identify	Primer sequence (5'→3')	Band size (bp)
FPGS 03g02030	3gE1_for	62	Wild type	CATCACCAACCCACACCCACACCTA	991
	3gE4_rev			GTGAACAGCCCTGTGCTGAAACC	
	3gE1_for	62	Inserted line	CATCACCAACCCACACCCACACCTA	667
	RB_rev			CCACAGTTTTCGCGATCCAGACTG	
	3gE1_for	58	mRNA expression	CATCACCAACCCACACCCACACCTA	1016
	3gE11_rev			CCACGCAGTCCAAGATGTTGATCC	
FPGS Os10g35940	10gE3_for	58	Wild type	CGGTTGAAGGTGGTTCATGTC	1629
	10gE7_rev			CCTCAAGCACAGCTACATCC	
	10gE3_for	55	Inserted line	CGGTTGAAGGTGGTTCATGTC	1183
	LB_rev			CGGACACTGATAGTTTCGGATC	
	10gE3_for	58	mRNA expression	CGGTTGAAGGTGGTTCATGTC	453
	10gE7_rev			CCTCAAGCACAGCTACATCC	
hygromycin	hpt_for	55	Inserted line	ACATTGTTGGAGCCGAAA	368
	hpt_rev			GAATTCAGCGAGAGCCT	

Primers used for screening transgenic lines of cFBP, GNMT, FPGS 03g02030 and FPGS Os10g35940 genes

Genes	Primer names	Ta (°C)	Type	Primer sequence (5'→3')	band size (bp
cFBP	riceglu_for	60	forward	CCGCATCCATAAATCGCCCC	700
	cFBPrt_rev		reverse	GGTCGAACCACATCTGAATGCAG	
GNMT	riceglu_for	60	forward	CCGCATCCATAAATCGCCCC	920
	GNMTrt_rev		reverse	CACCCAGGACGCTGTGCTG	
FPGS03g02030	riceglu_for	58	forward	CCGCATCCATAAATCGCCCC	475
	3gE4_rev		reverse	GTGAACAGCCCTGTGCTGAAACC	
FPGS10g35940	riceglu_for	58	forward	CCGCATCCATAAATCGCCCC	1050
	10gE7_rev		reverse	CCTCAAGCACAGCTACATCC	

Primers used for generating probes for the southern analysis of T-DNA insertion lines of *FPGS* 03g02030 and *FPGS* Os10g35940 genes

Genes	Primer names	Type	Primer sequence (5'→3')	Probe size (bp)
Hygromycin probe	hpt_for	forward	ACTATCGGCGAGTACTTCTAC	966
	hpt_rev	reverse	ACCGCGACGTCTGTCGAGA	
FPGS03g02030	03g_for	forward	CAAACGGAAACCACACAACA	777
	03g_rev	reverse	GACAGACAAACCGAGGAGGA	
FPGS10g35940	10g_for	forward	GTCTTCAATCTTCAATTGCT	563
	10g_rev	reverse	CTCGACTTTGGTTATTGATT	

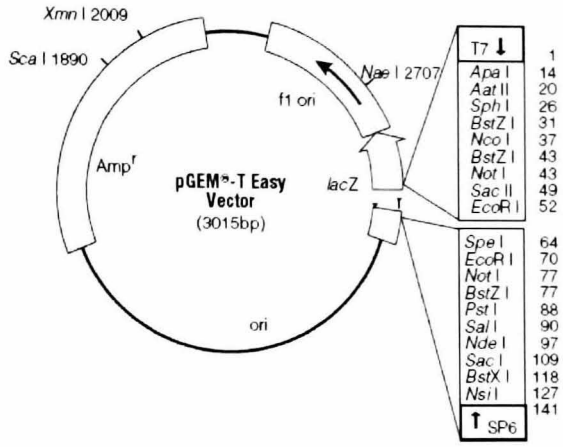
The primer sequences were obtained from Delphine Mieulet (French Rice Functional Genomics Centre, Montpellier).

Primers used for RT-qPCR analysis for transgenic lines of cFBP, GNMT, FPGS 03g02030 and FPGS Os10g35940 genes

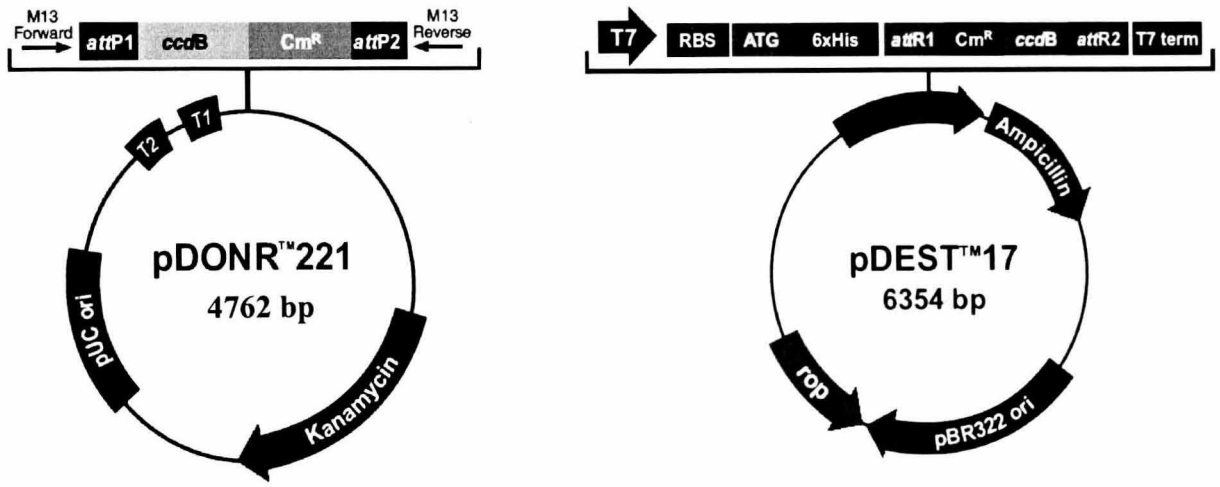
Genes	Ta (°C)	Type	Primer sequence (5'→3')	band size (bp)
cFBP	60	forward	CCTGCAAGAGCAACTGGCAC	195
		reverse	GGTCGAACCACATCTGAATGCAG	
GNMT	60	forward	GACGTCAGTGCTGACAGTAAAC	198
		reverse	CACCCAGGACGCTGTGCTG	
FPGS03g02030	55	forward	ATACCACTGAACCGAACTGACC	249
		reverse	CTGAGACTGCACCATGGAAGA	
FPGS10g35940	55	forward	GCTACATAAGAAGCATTCTGTTACC	240
		reverse	ATGTAGCTGATCCAAAGGGCC	

APPENDIX 2 CLONING VECTOR MAPS

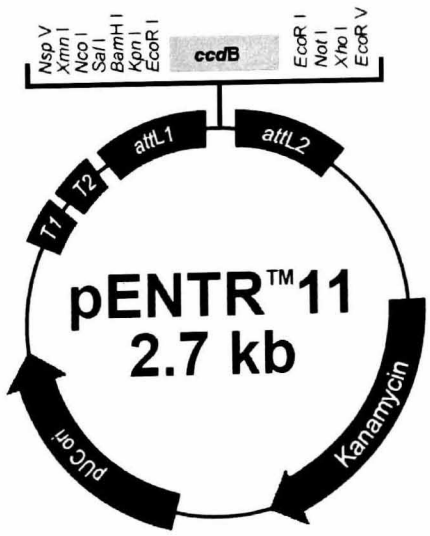
pGEM-T easy vector restriction map (from www.Promega.com)



pDONR221 and pDEST17 Gateway vector maps (from www.invitrogen.com)



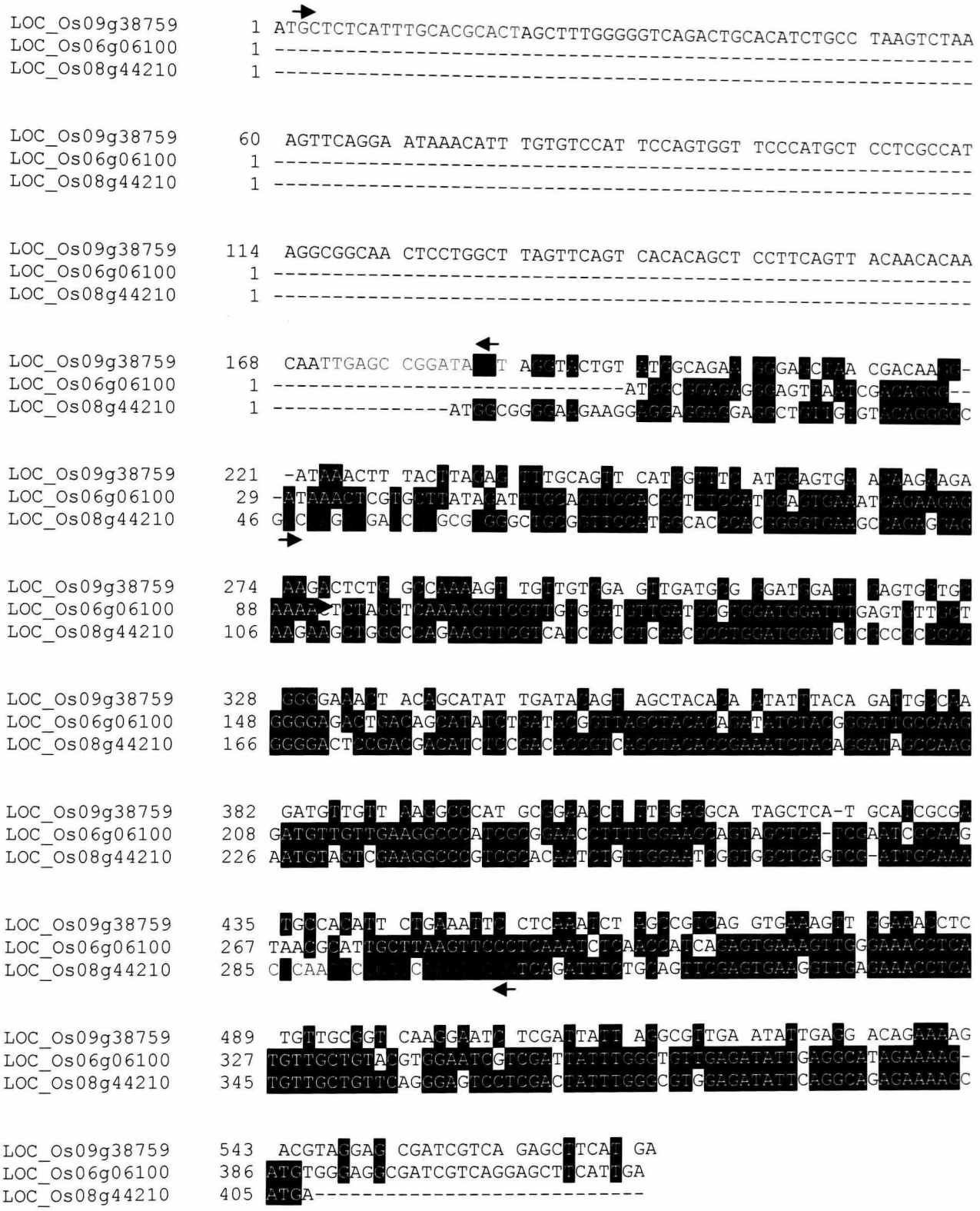
pENTR11 vector map (from www.invitrogen.com)



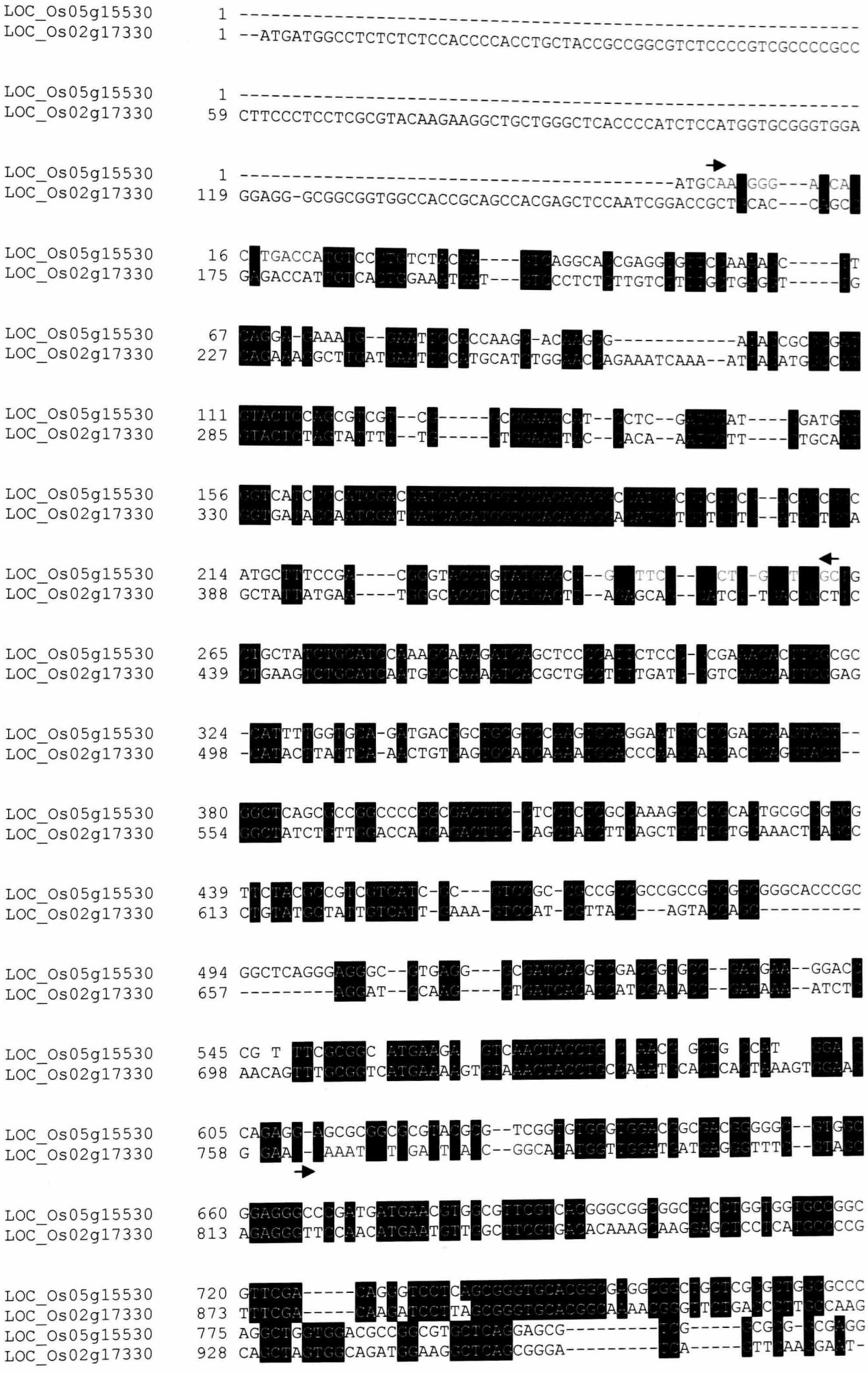
APPENDIX 3 NUCLEOTIDE ALIGNMENT OF *DHNA*, *ADC LYASE*, *DHFR* and *FPGS* GENES

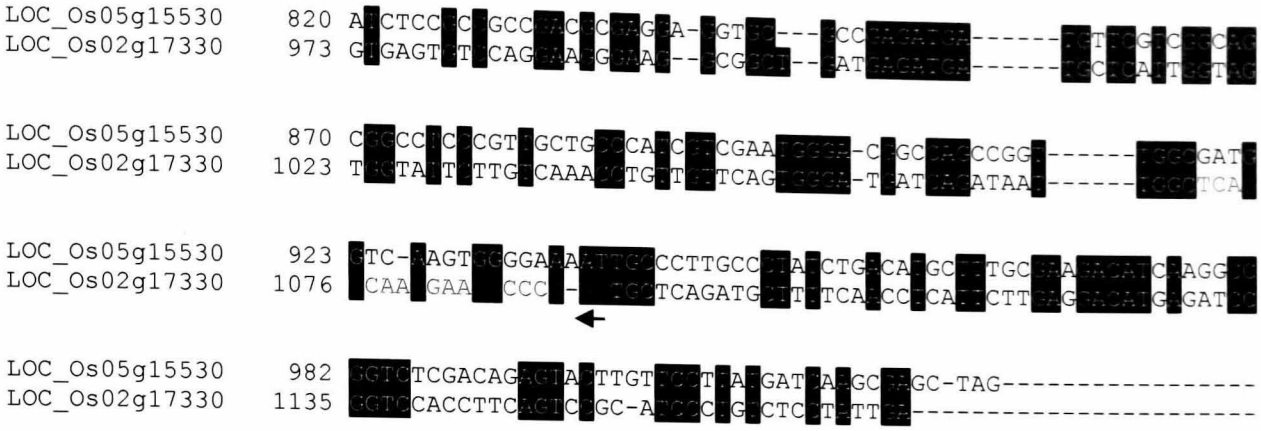
Modified from http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html; January 2009. The black highlighted letters represent the same bases between genes; the red letters represent the forward primer; the pink letters represent the reverse primer; the head arrows indicate the direction of designed primers.

DHNA cDNA

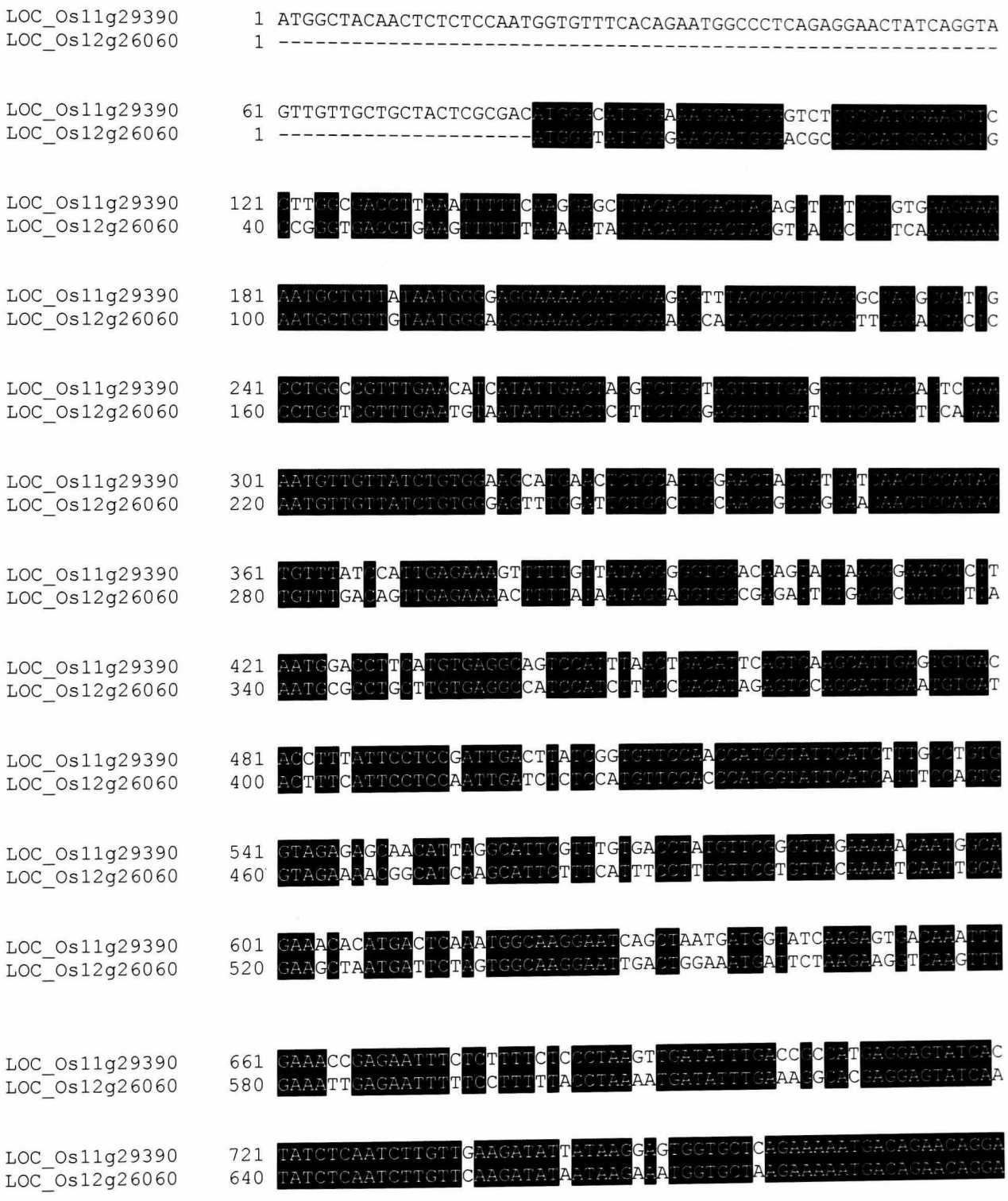


ADC lyase cDNA



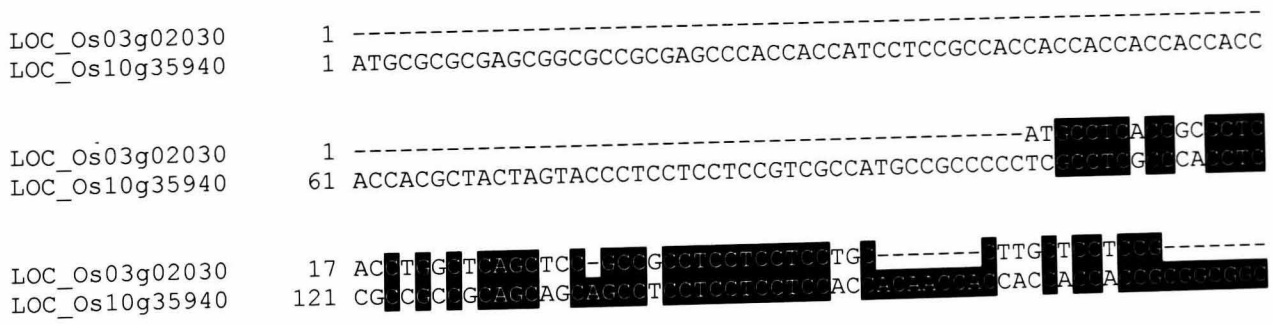

LOC_Os05g15530 820
LOC_Os02g17330 973
LOC_Os05g15530 870
LOC_Os02g17330 1023
LOC_Os05g15530 923
LOC_Os02g17330 1076
LOC_Os05g15530 982
LOC_Os02g17330 1135

DHFR cDNA

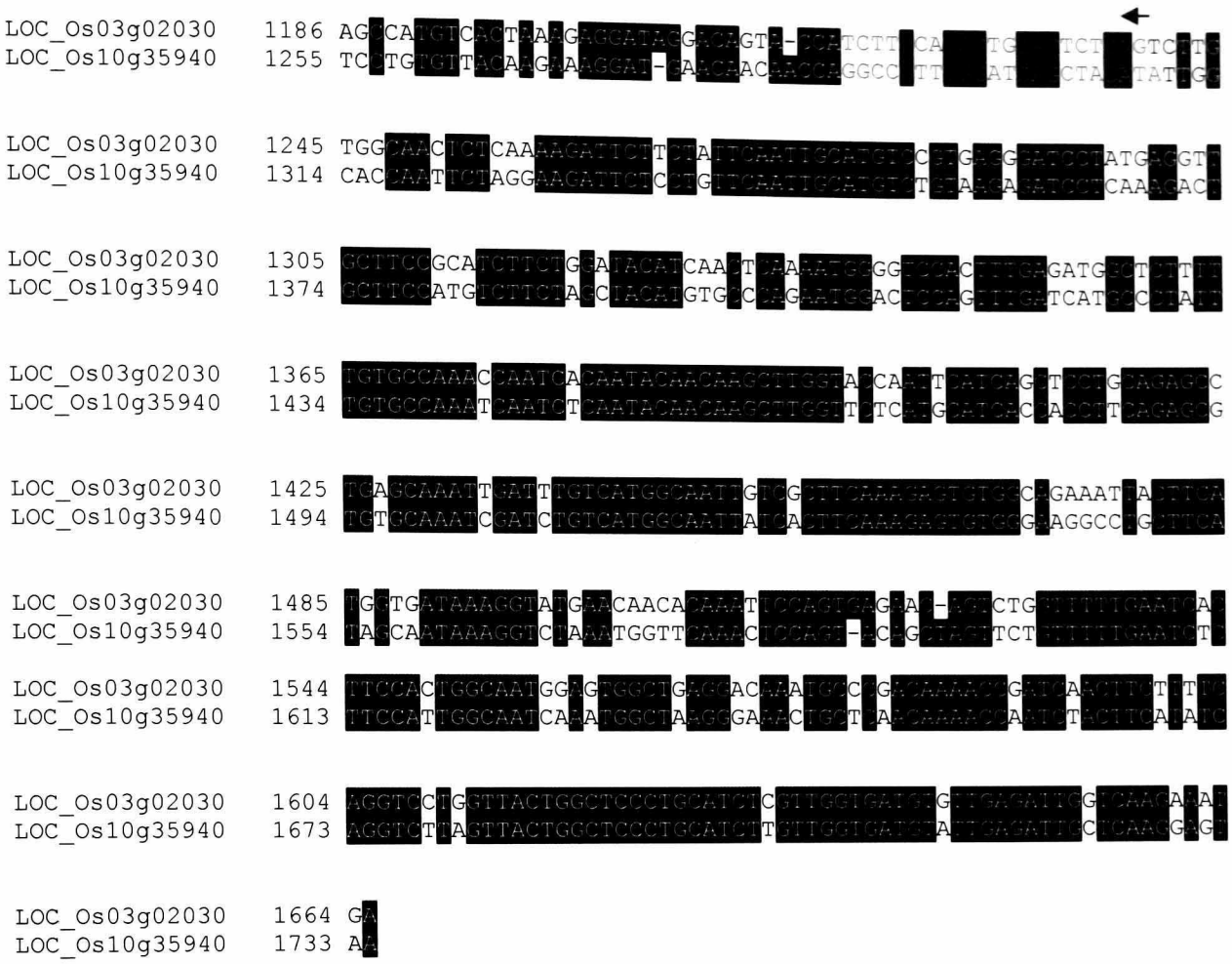

LOC_Os11g29390 1 ATGGCTACAACTCTCTCCAATGGTGTTTCACAGAATGGCCCTCAGAGGAACTATCAGGTA
LOC_Os12g26060 1
LOC_Os11g29390 61
LOC_Os12g26060 1
LOC_Os11g29390 121
LOC_Os12g26060 40
LOC_Os11g29390 181
LOC_Os12g26060 100
LOC_Os11g29390 241
LOC_Os12g26060 160
LOC_Os11g29390 301
LOC_Os12g26060 220
LOC_Os11g29390 361
LOC_Os12g26060 280
LOC_Os11g29390 421
LOC_Os12g26060 340
LOC_Os11g29390 481
LOC_Os12g26060 400
LOC_Os11g29390 541
LOC_Os12g26060 460
LOC_Os11g29390 601
LOC_Os12g26060 520
LOC_Os11g29390 661
LOC_Os12g26060 580
LOC_Os11g29390 721
LOC_Os12g26060 640


LOC_Os11g29390 781
LOC_Os12g26060 700
LOC_Os11g29390 841
LOC_Os12g26060 760
LOC_Os11g29390 901
LOC_Os12g26060 820
LOC_Os11g29390 961
LOC_Os12g26060 880
LOC_Os11g29390 1021
LOC_Os12g26060 940
LOC_Os11g29390 1081
LOC_Os12g26060 1000
LOC_Os11g29390 1141
LOC_Os12g26060 1060
LOC_Os11g29390 1201
LOC_Os12g26060 1120
LOC_Os11g29390 1261
LOC_Os12g26060 1180
LOC_Os11g29390 1321
LOC_Os12g26060 1240
LOC_Os11g29390 1381
LOC_Os12g26060 1300
LOC_Os11g29390 1441
LOC_Os12g26060 1360
LOC_Os11g29390 1501
LOC_Os12g26060 1420
LOC_Os11g29390 1561
LOC_Os12g26060 1480

FPGS cDNA


LOC_Os03g02030 1
LOC_Os10g35940 1 ATGCGCGCGAGCGGCGCCGCGAGCCCACCACCATCCTCCGCCACCACCACCACCACCACC
LOC_Os03g02030 1
LOC_Os10g35940 61 ACCACGCTACTAGTACCCTCCTCCTCCGTCGCCATGCCGCCCCCTC
LOC_Os03g02030 17
LOC_Os10g35940 121


LOC_Os03g02030 62
LOC_Os10g35940 181
LOC_Os03g02030 111
LOC_Os10g35940 241
LOC_Os03g02030 166
LOC_Os10g35940 301
LOC_Os03g02030 226
LOC_Os10g35940 361
LOC_Os03g02030 286
LOC_Os10g35940 421
LOC_Os03g02030 346
LOC_Os10g35940 481
LOC_Os03g02030 406
LOC_Os10g35940 541
LOC_Os03g02030 466
LOC_Os10g35940 601
LOC_Os03g02030 526
LOC_Os10g35940 661
LOC_Os03g02030 586
LOC_Os10g35940 721
LOC_Os03g02030 646
LOC_Os10g35940 738
LOC_Os03g02030 706
LOC_Os10g35940 784
LOC_Os03g02030 766
LOC_Os10g35940 844
LOC_Os03g02030 826
LOC_Os10g35940 904
LOC_Os03g02030 886
LOC_Os10g35940 964
LOC_Os03g02030 946
LOC_Os10g35940 1024
LOC_Os03g02030 1006
LOC_Os10g35940 1084
LOC_Os03g02030 1066
LOC_Os10g35940 1144
LOC_Os03g02030 1126
LOC_Os10g35940 1195


LOC_Os03g02030 1186
LOC_Os10g35940 1255
LOC_Os03g02030 1245
LOC_Os10g35940 1314
LOC_Os03g02030 1305
LOC_Os10g35940 1374
LOC_Os03g02030 1365
LOC_Os10g35940 1434
LOC_Os03g02030 1425
LOC_Os10g35940 1494
LOC_Os03g02030 1485
LOC_Os10g35940 1554
LOC_Os03g02030 1544
LOC_Os10g35940 1613
LOC_Os03g02030 1604
LOC_Os10g35940 1673
LOC_Os03g02030 1664
LOC_Os10g35940 1733

APPENDIX 4 PROBE SEQUENCES FOR THE SOUTHERN BLOT ANALYSIS

FPGS gene sequences were modified from Gramene website; April 2009. Hygromycin sequence (AB289770) was obtained from ncbi website, April 2009. The black highlighted letters represent exons of genes; the non-highlighted letters represent introns of genes; the grey highlighted letters represent the probe sequence used in the southern blot analysis.

Hygromycin resistance gene

CTATTCCTTTGCCCTCGGACGAGTGCTGGGGCGTCGGTTTCCACTATCGGCGAGTAC
TTCTACACAGCCATCGGTCCAGACGGCCGCGCTTCTGCGGGCGATTTGTGTACGCCC
GACAGTCCCGGCTCCGGATCGGACGATTGCGTCGCATCGACCCTGCGCCCAAGCTGC
ATCATCGAAATTGCCGTCAACCAAGCTCTGATAGAGTTGGTCAAGACCAATGCGGAG
CATATACGCCCGGAGCCGCGGCGATCCTGCAAGCTCCGGATGCCTCCGCTCGAAGTA
GCGCGTCTGCTGCTCCATACAAGCCAACCACGGCCTCCAGAAGAAGATGTTGGCGAC
CTCGTATTGGGAATCCCCGAACATCGCCTCGCTCCAGTCAATGACCGCTGTTATGCG
GCCATTGTCCGTCAGGACATT
GTTGGAGCCGAAATCCGCGTGCACGAGGTGCCGGACTTCGGGGCAGTCCTCGGCCCA
AAGCATCAGCTCATCGAGAGCCTGCGCGACGGACGCACTGACGGTGTCGTCCATCAC
AGTTTGCCAGTGATACACATGGGGATCAGCAATCGCGCATATGAAATCACGCCATGT
AGTGTATTGACCGATTCCTTGCGGTCCGAATGGGCCGAACCCGCTCGTCTGGCTAAG
ATCGGCCGCAGCGATCGCATCCATGGCCTCCGCGACCGGCTGCAGAACAGCGGGCAG
TTCGGTTTCAGGCAGGTCTTGCAACGTGACACCCTGTGCACGGCGGGAGATGCAATA
GGTCAGGCTCTCGCTGAATTCCCCAATGTCAAGCACTTCCGGAATCGGGAGCGCGGC
CGATGCAAAGTGCCGATAAAC
ATAACGATCTTTGTAGAAACCATCGGCGCAGCTATTTACCCGCAGGACATATCCACG
CCCTCCTACATCGAAGCTGAAAGCACGAGATTCTTCGCCCTCCGAGAGCTGCATCAG
GTCGGAGACGCTGTCGAACTTTTCGATCAGAAACTTCTCGACAGACGTCGCGGTGAG
TTCAGGCTTTTTCAT

FPGS Os03g02030 gene (5′UTR to 3′UTR)

ACTCTTTGACCTCGATTTTTTTTTTCTCTCAAGCATCAATGGTGTTTCCCCATCTGAAGT
TTTTTCAGTTATCTGAACATCTGTTAGTGATAGAAACTTATACCTTTTTCCTCAACGATT
CCACAATCACATTTGCCATATAATTGATAATAAATCAGGGCATGACATGTCACATATCCT
TTAACCGATCGTACTATACCAGTATACCTGAAACAAAACAGCACACATAACATTCCAACT
CTAATCTGAACTTCCCCGCAAAAAGAAAAACTCTAATCTGAACTTGAGCATTCTAAGACG
TGGTTTCACCGTCTTGCCCAAAATCTAGTTTTGTCAAGGCATCAACGCCACCACAACACG
GTGCTCGCTCCGGTTGACATGGCCGCCGCCGCCGTGTCGCCATCCGCCGGAACCGGTGGT
GGTGGTGGTGCCGCTGCTCCAAGCCGCGACCCTCCAATTTCCCGTTATAAGCAGAGGCAT
GACAGAAACCTGAGAACCCGCCTCCTAGAGGAGAAAGCGTATATTCGGTAACACTTTGCA
AAAGTGCGAGGGCGACCGCGAATCCGCGATAAAGAATCCGGGGGGACCCACTTGTCAGAG
GAAGGGTACAGAGACAGACAAACCGAGGAGGAGGAGGAGGGCCCACCAGCCACCACGGCG
GCGGCGGCGCCATCACCAACCCACACCCACACCTACTACTACGACGACACAAGTAGTGGT
CGTCGAATCCGCGGCGGCGCTCGCCTGAGCTTGCGCTGCACCTGCACGCGGCATATGCCT
CACCGCCCTCACCTGGCTCAGCTCCGCCGCCTCCTCCTCCTGCCTTGCTCCTCCGTCCCC
GTCCCCGTCCCCGCCCCGCCGCCTCTCCTCCCCTCGCGCCATGGCCTCCTCCTCCGTCTCC
GCCGCCGCCGCCGCCGCGCAGGCAGGTAAGCAACGCCACGCCCTGATTTTGTGCTGCTGC
TGCTGCTTCGATTGATGGGGTTGGCTGGTGGTGGATACTGGATTGAGTTGATTCGTGGGG
CGTAGGTGGTGCTGTGGCGGCAGCGGAGTACGAGGACGTGATGGGGCGGCTCTCCTCGCT
CATCACGCAGAAGGTGCGGGCGCACAGCGGCAACCGGGGAAACCAGTGGGACCTCATGGC

GCATTACCTCCAGGTCCATTTCCTCCTCCACTCCCCACTTGTCAAAATAGTATATTAGGA
TGGCTCATGGCTGGTTGTTTGATTTGATCGGGTGGATTTTGTTGCTTGCCGCGGTTCGCG
CAGATTCTGGAGCTGGAGGAGCCGATCGCGCGGATGAAGGTGATTCACGTTGCCGGGACC
AAAGGAAAGGTACTAACTCACTGATGATCAGCTCTAGCTAGAAAGCTGGATGTTGTGTGG
TTTCCGTTTGGGTGAAGAGTGTTTAGAGAGTAGCGATGCAAGGTGCATAAATCACTAATG
GAACATGTCTACGCAGAATCTTAATGCAACCAATCTGCACATTAATGTTTACCTATGCAA
TAGTTTACTTCTGTTTGGTAATGAAACCAATTAATGTATAATAACTTATCTTCCTTATGA
CGTGGAGTTTGAACAACACATGGGCTTTATCTACAGGGTTCAACATGCACATTCACAGAG
GCAATCCTGCGATCATGTGGTTTCAGCACAGGGCTGTTCACCTCACCACATTTGATGGAT
GTCAGGGAGCGATTCCGGCTAAATGGGTAAAATTACTTTTCATTTTTCTCCTTTACTTAT
CGTTGACTCTAATGTTGAATTGCTAATGCTTCTCATTTCGATACTTTAGAATAAGCTTCT
GGACACACATGTTCTATACCATATCTTTTATGTTATTTGGATAACTATATATTCTGCATT
GTACTTAAAATTCCGGCAAATAATTGAATGGACTTGATTTTGCGTGCATCATATCCACCA
TGTCTTTTATGTTATTTGGATAACTATGTATTTTACTTGTTCTGATCTAGCTCCTAATTG
TGAGAGCATCAGGGTGGATATTTCCGAAGAAAAGTTTTTGAAATACTTCTGGTGGTGCTG
GAATAAGCTCAAGGTATATGTCTATCTCTATCTCTGTTGTTTCATATGATGGCACTCTGT
GCTGGGCGCTAGCACTGATGAAAAACAACGACAAAATAAATGATACCAGGAACCTGTGCA
TTTCCTCTTTGTGTTTTTCTGCTGCTTTCGTTATCAAAACTTCTAGATGATGATGTACAT
GGGTGTACTTTCAGGAGAAGACTGATGATGATATTCCCATGCCAACCTACTTCAGATTCC
TTGCATTGTTAGCATTCAAGATATTCTCAGCTGAACAGGTGAGATGTTTCTGGGTTGGCT
GCTAGGATGCCTCGCAACAACCATGCTTGTTGTAAAATGTAACATCTTCTCCCAATCTTA
ATAAAATTCGGCCGGGTCACTTCTTCGCTTCCTGCCTGGCAAAAAAGATATAACAGCAGA
CACTTCCTGCTTTTTAAATTTTTGTTTTACTGCTTGCATCTGTTTGTACACATAACTTGT
TCCTCCATTTTACTTGAATGACATTGTAACTTTATAGGGCTTTAATGTACTAAGGTAGTC
GCACCAAAGACTTATATCACATATTTCCTTTTTTGAGACTTGGAATGTAGTGGAACCATT
TATGATTTTCTGGGTAAACATTATTTGATGTTTTACTCCCTCTTTTCGTGGGTCAAACAG
GTTGATGTTGCTGTTCTCGAGGTTGGCCTTGGAGGGAAGTTTGATGCAACTAATGTGGTA
TGTTCTGTTCTGTTGGGTGCAAGTGTTAGTTTCTTTGTAATCGAAATCCAAGACAAACTT
GAACTATATGAAAATCTTGTTTGCCCTAGGCTGTTTGCTTCTTCTTTCTTTTCTTTGCGG
CCGTCCTTTCATCCCATACAAAAGGTGTGATCACAAAGCAAACTGTACTAGTAGTGTACT
AGATATAGTCATTCGGGTCTAACCATATGCTGATGTCTTCTATTTTAATCAGAGTTACAG
TTTAGCTTTTGTTTCTTCTGAACATTAATGAATTCATTCAGATCAAGTACTGCATCCTGA
CGACTCAAGGCATAAATAATTCAAAATAATAAATGCGTAACAGGTGTATAACTTTGCACA
TATGCAGGTTGAAGCACCTGTGGTTTGTGGAATAGCTTCCCTTGGATACGACCATATGGA
AATTCTTGGTTAGTCATGAAACTGACAGTTGATGCCCCTCTATTACTAGGAAAAGTAGTT
TCATGCTTTTGCCATGTGCCCTGTGAGCAGGAAACACACTTGGAGAAATTGCTGGTGAGA
AGGCAGGGATTTTCAAGGTCATTTTAACTACTACCCTAACAAAGTCATGCTACAAATTAG
GATGAACTGTTATATTATCACTTTTTTCACTTGGACTGTTGCCGATTAAAAGAGAATACG
TAGCCAACATTTCTTGATTTTGTTGATTTGCTTGTGGCGTCTGATATCAACAGAAGGGAG
TTCCGGCCTACACTGCTCCACAGCCAGAAGAGGCAATGATTGCTCTGAAACAAAGGGCTT
CAGAATTGGGTGTATGTTTTATTTTAAACTTGTCCCTTATGTTTTGTAATATTTGTAAAC
TTGATAGTTAATACAGCTGAAAGCATCTTCTAAAATAAGCTAGTCTGTGCCTTCTGAAGG
TTTCACTCCAAGTTGCACATCCTTTGGAACCCCATCAATTAAAGGATCAACATCTTGGAC
TGCGTGGTGAACACCAATATGTAAATGCTGGCCTTGCAGTTGCATTGGCTAGTACATGGC
TTGAGAAGCAAGGACATGTGGAAAGAATACCACTGAACCGAACTGTAAGTCTCAATCAAT
AATTGTTGCTTACAAAAGTTGGTATGATAATTTCTTTCTGCTCTTATGGAATTTAGGACC
CGTTATGATAATTTCTTTCTGCTTTTATGGAATTTAGGACCCGTTACCAGATCAGTTTAT
TAGTGGGCTGTCAAATGCCTCTTTGCAAGGCCGAGCGCAGATCATTACAGATTCACAAGT
AAACTCAGGAGAGGAAGATAAAGACTGTTCTTTGGTTTTCTATTTGGATGGGGCTCACAG
TCCTGAAAGTATGGAAATTTGTGCTAGATGGTTTAGCCATGTCACTAAAGAGGATAGGAC
AGTACCATCTTCCATGGTGCAGTCTCAGTCTTGTGGCAACTCTCAAAAGGCATGCCTCAC
AGTTTCTCTAGTTTCATCTTACTTGAATGGACTTGATTTTGCATGCATCATATCATTGAA
TCTCCTTTCCCCTTATATGAAAATTTACCTTTGGACCAATCACAGATTCTTCTATTCAAT
TGCATGTCCGTGAGGGATCCTATGAGGTTGCTTCCGCATCTTCTGGATACATCAACTCAA
AATGGTAGGTAAAAATTGCTTGAATGGAAAAATGAAGTCCATATTCACTCATCTATCTAA
TACTCCCACTTTATAATGTGTACTGGTTGTTTCCACTAACAACTAGAGTTCTGCACAAAC
TTATTTGTTCACTGTTGATTTCACTTTCAAGTTTAATTTGTTCCTGTTCTTGGTTGAATG
TTTGCATTTATTGTTGTTTCATTTTTCCCAGGGGTCCACTTTGAGATGGCTCTTTTTGTG
CCAAACCAATCACAATACAACAAGCTTGGTACCAATTCATCAGCTCCTGCAGAGCCTGAG

CAAATTGATTTGTCATGGCAATTGTCGCTTCAAAGAGTGTGGCAGAAATTACTTCATGGT
GATAAAGGTTCTTCTTAACTTGATTACCAATAGACATTCATGTTTCTATTGGTCATATGC
AGTTTTCTTGTTGAAACAAGTTACTTTTAGCTTGAAGATGATATCTGAATTTAGAAGCAA
ACGGTGCAGGTATGAACAACACAAATTCCAGTGAGAACAGTCTGGTTTTTGAATCACTTC
CACTGGCAATGGAGTGGCTGAGGACAAATGCCCGACAAAACCGATCAACTTCTTTTCAGG
TAGGTGCTTCCATCAAACTATAACCTATAACTTCAGAAATGTTTTCCTTTTTTAAATGCT
AGTGTTTCATCCTAATATGTATTCCTGCATTTATAACTAGAGTGCTGTGCTATAGATGGC
TACTTTATGTTCACCATCACTGTCCTGTAATCTGTGATATTTTTAGTTTTGGCATGTGAT
CATGTTTATTGAATTTCCTATTTTGCATAAATTCGATTTTTTGGAAACATCGTAACATGT
GTGTGCGCAAGCTACTTCCGCACTTGGCGCTGATCCATTTGTGTCTGATACACTGAATTA
GGTCCTGGTTACTGGCTCCCTGCATCTCGTTGGTGATGTGTTGAGATTGGTCAAGAAATG
ATACTTCTTCTTGGAACGATATACACTTAGAAGTTCAAATTGACGCTGAGCCAATTTCGC
GGTTTACGTTTGAGATCCCCTGACGGCAACCGGGCGGTTACAGGATACAACAATTTCAAT
TTCAGGTGTTTTGGATTCTATGCTGACCAGGCAGCGAGCATATCGCGTATACTCTTCTTG
TTATGTGTTTCATGGCTCCTTTTTGCCACTAATCCGTATTCTTTACCCAGAAAACTTCAA
TGCCCATTACAACAAGTAATAAATTGAATACATGAATAAGATGGTGGCAATCATGAATAG
GATACACTTTTCTAAGCTTCTGGTGAAGATGTTCGATTCATCCCTGAGGTTCACGTGCAT
TTAGTCAACTCTGATGAATGCCTAACCAAATACTGATACACATGATTGATGATTTCCATT
GATATCACAATGATAGGACACAACTGCGTTTTTCTTCGTTGCACCCGTGTGACATATCTG
AGAGGCCAAGACTGCTGTTTCCTACTAACTCTTAAAAAAAGATTAAGACTGCTCTTGTCC
TGAATGAAGATGCCATTAGCTCTACTGTTAAGGTTCTTAATCCCATTTGGCATGTTGTCC
ACATGAAGCCAAATGGTGAGGAAAGCACAATCCAACCCCTCCATGTCACAAAGTTGGCTG
GGTTTGTCTTATCAAGAGAGGAAAGCTCCTTGTCGATACAGATGGCCCTCTGCCGCTCTT
CTTGTCTGGGTCACAATAAAGTTTTAGCCGGCTATCTCAAAGGATTTAGTTTTTTTTTTC
CTCCTTTACTTGTTCTTAAAGGATTTAGTTGTTTAGCCTGCAGGCTATAAAGTGAACAAC
CAAAGACAATTTTCTAGTAAAGGTTTTCTTGTAGCTTCCAGCAGCTAGCTTAGCTCTTAG
CTCTTCCCTTCATCTTGGCTCTTCTTGTAGAAGTAGCAA

FPGS Os10g35940 gene (5′UTR to 3′UTR)

AGGTAAATGCACTACAACTACAAGCTCCCTTAAGATATGTTCTTAAATCTCGTGTGTAGA
TTTAAAATAAATTTCATCACAAGAGAGTCTTAAAAAACTCATGTTCGAATTTATTTTAAA
GAGAAAACGTATACTTAAGTTAGATTAGGTCCGAGCAATACTAAGGCTCTGTTCGAGAGT
ATGTGTTGCCAATCCTCATCTCTCTCATTTTTCATACGCACGTTTTTCAAACCGCTTAAC
GGTGTGTTTTTTAAAAAATAAATTCTATAGAAAAATTACTTTATAAAAATCATATTAATA
TATTTGTAAAATTTATTTTAGTTATTACTTAATTAATTATACGCAAATAAGCTGTTCCGT
TTTCCGTGGGGGATGGGAAGGTTCAGCCTAAACCAACCAGGCAACCAGGGCCATATTAAG
TTGAGCTAAGCGTCAGTATTTGAAGCCCAGAAGAAACCTAGCCCAGATTCCAAACCAAAC
CGGTTGCCTCGTAACCCACAAAGCCCATCATCAGCAAGCCCAACTCCGCGGTGGTGCCAC
GAGCGAGCAAAGCAAAACCCGACCATAAAACCGCAAAACGCCCCGAGCCACTCCGGATCA
CACGATCCGGCAAAACGAAAACGATGCGCGCGAGCGGCGCCGCGAGCCCACCACCATCCT
CCGCCACCACCACCACCACCACCACCACGCTACTAGTACCCTCCTCCTCCGTCGCCATGC
CGCCCCCTCGCCTCGCCCACCTCCGCCGCCGCAGCAGCAGCCTCCTCCTCCTCCACCACA
ACCACCACCACCACCGCGGCGGCGGCAGCGCCCCTCCCCACCCTCTCCTGCGCCCGCCGC
CGCCGCCGCAGTCCCACCCCTTCCAGCTCCTCACCCCGCGCGCCGCCATGGCCAGCGTCG
CGCAGCCAGGTACGTAACGTAAACCGAAGCTCGCGAGGGTTTCTCTCTGTCGTGTAGTTC
TGGGTGTGGTGATTGATTGGGATTTGGGATTGATTCGGTGGTGGTGGTGGTGGTGGTGGT
GGTGCAGGAGTCGCGGCGGGGTCGGCGGAGTACGAGGAGGTGCTGGGGTGCATCTCGTCG
CTGATCACGCAGAAGGTGCGGGCGGACACCGGCAACCGTGGGAACCAGTGGGAGCTCATG
GCCAAGTACCTGCAGGTGCGCTCCCGTGGGTGGAGCTCGACTTTGGTTATTGATTGCTCT
ATTGGTGGTTTCTGATTGGGATTGATTTTGGTTGGCGTTGCTAGATACTGGAGCTGGAGG
AGCCGATCGCGCGGTTGAAGGTGGTTCATGTCGCAGGGACGAAAGGGAAGGTGAGTACAC
TACAGCAGAGAGTGGCTGGAGGGATTCACATTGCTGTGTTAGAGTGATTTGAACTGATTT
GGACAAAAGGGAAGGTGTGCAGTTTCAGTAATGTATCGCATATTTTTCTGTAAATATGCG
CATCAATAGTAATAGCAAAACCAGTGGTCGGTATTAGCTGCATTGTTTGCATTTGGTTAC
CTGTGAAATCAAGTGGGATTGCTATGTGTGGTGCGTACAGTTTTCTCTAGAAACTCGCAT
GTTTGGTGGAGCTGTCCCTGCCTGAGAAATGAGGATTTCACCTAACTGACTTAATTGTTA
GGTTTAATTCTGTCTATTTAGGGCTACGTGTTTGTGTTTGCTAGTTGTTTGTATTGGCTG

TCACTGAGTGTAGTGATGCGGTGATGCCTGACCAATGTAAGCAATTGAAGATTGAAGACA
CCCTTTAGACTGATTTTGTGGTGCATATATATTGAATAGTATACCAAAGAGAATCGTATT
TCAATATATTCCAAGGAGAAGTCCCACTAAAAGCACTGAAATATCATTTTCCTGGTATCG
TTTAGGGTTCAACATGCACATTTGCCGAGTCAATTCTTCGATCATGTGGTTTCCGCACTG
GGTTGTTCACCTCACCACACTTGATGGATGTTCGTGAGAGATTTCGGCTTAACGGGTATT
AGTTCACTTCAATCTGCCGCATTGCGACCTTTATTTGTTCTAATTTTTTTTATATATCTT
TAGCTTGTTATAATTAGATTACCAGGGATTTGTAGTTTTGCATTAACAACTTTCTTTGAT
TATTGCCATGACTGAACTCTTGTGCTCTCTTAATTTCCTCTCTGGTGAGTTCCTAGGTTG
GACATTTCGGAAGAAAAGTTTATAAGGTATTTCTGGTGGTGCTGGAACAAACTGAAAGTA
AGGATCCCTCATTTATATGTTTCACGTGGTGAAAGTTATGGTTGTATCATACTGACTATG
GGATATTGTTGGATTTGCATCATTTTCTATAAGCTTTGAAACACTCACTGTGGTTAATCT
GTTGTGCCTTCCAGGATAAGACTGGTGGTGATATTCCAATGCCAGCCTATTTTAGATTCC
TTGCGCTGCTAGCATTCAAGATATTTTCTGATGAGCAGGTAACATTCAAATACTTTCTGT
CCTAAGCTTTTTCAATCCTGTAGTAAATCTAGTTAAGTTTATCCCAGCTTTCTATTTAAA
TGGATTTTGAAAACTTCACTAATACTAATCATCAAAAGAATGGATTTTCATGGACATTGT
CCCACTCAACATATACTGAATTCGTATATTTCTCAAAGAAAAGTCATATCATATTATGTT
AGATCTTTATTGAATCTAAATGGTTGCTAGTTCTATGATGTTTTCATCACCCTTCACTTG
TTATCTGCAGTTCATGTGATTAGTAGAATTATGAAATGATTAGTTGTCTGATAAGTTTAT
TTACAAGCATTATCATCTTTGAGTAGATAATCACTGTGCACTTCTTGAATTTAACAATTG
CACCCATGACTTCCCTAGCGATTTATTTAAATCCACTTTCTTTGTTTTTTGGTTTAAGGT
GGATGTAGCTGTGCTTGAGGTTGGTTTGGGAGGAAAATATGATGCAACGAACGTGGTGCG
TGACATTTTACTGTCAAATAATGAATGGCAGTATATTTATATGGCATAACTGCCAACCAA
AAATAAAGATTATATCATCTCATTTTGACCTTTTAGTATGAATGCTGAATGTTTGTGAAG
TCCACAAAATGTTACCATCTAGGTTCCGCTGTAAAACATGCAAATTTGTTTGCATTTTGT
GTTCAACTCATGTCCTGTTCCATTGTAGCCCACTGAGATTGCTGCCAATATCTTGGCTTG
AACAAAAAACTGAAGAGTATAATATATTTGAAAATTTTGCAGGTCAAAGCACCTGTAGTT
TGTGGAATATCTTCCCTTGGATATGATCACATGGAGATTCTTGGTTAGTTAAGAAACTGA
AGGAAGAAATTCCTCTCATACATAACAAACTCACGTCACACTATACGTGTTTGTTTTGGC
TGATGAGCAGGAAATACGCTAGGAGAAATTGCTGGGGAGAAAGCCGGAATATTAAAGGTT
CTGTACTATTTACTTTTCCACAATATTTATCAATAAATAAATATGTGTAACATGTTTTCT
TTGCTCTGATATCCTTCAGTTAACTACTTCCATAGCTAAATTTATATGTTGCTTAATCCC
TGTGTAGCCTCTTTCTAAATGCAGAAAGGAGTTCCAGCATATACAGTTCCACAACCAGAA
GAGGCAATGTCTGTACTGAAGCACAGAGCTTCTGAATTGGGTGTATGTTATTTCCTGTGC
TTCATTTTTCTTTTGGGGTTTCCACCATGCCTTGCACTCTAATGTATCCATGAACATTGT
TACAATCAGTATAGTACAGCTTTTTTTCATTGACACGAACAACTTAAATTCTTTAAAGGT
TCCTCTTCAAGTAGTCCAGCCTTTGGATCCACAACAATTGGATGATCAACCTCTTGGGCT
ACATGGTGAACACCAATACATGAATGCTGGCCTTGCAGTCGCATTGGTTAATACCTGGCT
TCAAAGGCAAGGGCATTTCAACATACTACATAAGAAGCATTCTGTAAGTACTCTAGCCCA
AACTGGCATGTAAAGAAAGTTAAGTTCAGTAATCAACCTAACCTGGAATGCAGGTTACCT
TGCCAGATCAATTCATCGAAGGGCTGTCAAGTGCTTGTTTGCAAGGACGAGCACAGATTG
TTCCAGATCCGGAAGTGCTGTCCAAAGACAGCAGTTCCTTAATTTTCTATTTGGATGGAG
CACATAGCCCAGAAAGTATGGAAATTTGTGCGAAGTGGTTTTCCTGTGTTACAAGAAAGG
ATGAACAACAACCAGGCCCTTTGGATCAGCTACATATTGGCACCAATTCTAGGAAGGTAT
GCCTTGCAGTTGCAGTGAGCCAGTGAGAGGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGGCTCATATCATTGTTCTTTCTTTCTTATGAGCACTCTTTGTTTGTGAT
CCTTGCAGATTCTCCTGTTCAATTGCATGTCTGTAAGAGATCCTCAAAGACTGCTTCCAT
GTCTTCTAGCTACATGTGCCCAGAATGGTATTCTAATCAAATTCATTTCCCTTGCATAGT
TTACCTTTTCTCACTATCCAAGATCATTACATTTCTATTGTGAATTGTATGAATGCTTTT
GTTTCTGAATTCTGATGAGCAATCATTTTCTCAGGACTCCAGTTTGATCATGCCCTATTT
GTGCCAAATCAATCTCAATACAACAAGCTTGGTTCTCATGCATCACCACCTTCAGAGCGT
GTGCAAATCGATCTGTCATGGCAATTATCACTTCAAAGAGTGTGGGAAGGCCTGCTTCAT
AGCAATAAAGGTTGCTCTAAACTGAATTGCAATAATTATTATTGTTCATTTATTGCCAGC
TTGTGTTTTCTTATTGAAACAAAGTAAGTTCACTGGGTTGAAGATGGTATATAAACTTAA
AAATTGGAACGGTGCAGGTCTAAATGGTTCAAACTCCAGTACAGCTAGTTCTGTTTTTGA
ATCTCTTCCATTGGCAATCAAATGGCTAAGGGAAACTGCTCAACAAAACCAATCTACTTC
ATATCAGGTGAAAACTTCTTATATACCTTCATTTTTGTTCTCCCAATTGTTTGTATCATG
TAGGACTAATTTTTCTTCCAACCATACCTAGCTACTGTACCATAGTCAGGACACAGGGAT
ATCACATAATTCATTGTCAGCTTAGCACATGTATATGTAGTTGTGTAGTGACCATCACTG
ATTCATCATTCTTGTCTCTTGTGTACCCTTCCATCCCTTCGATGTGCTTGCGTGTGTGCA

ACTGCTAGTGAATGCACAGTTCTTTTTTCTTTTTTGGTTCTGCTTTTCCATGATCTGGAA
TACCTAGTAATTGTTTATGTGCATGTGTTTATTTTATGTCCCTTTAATTATGTTATACCT
GTTGTTTTTTATGTTTTTTTGTCTCTAGCAGGCAGCAGCTACTTTCAAAATTGATGCTGA
TACATTTGCATTTGTACACTTTGACCAG**GTCTTAGTTACTGGCTCCCTGCATCTTGTTGG**
TGATGTATTGAGATTGCTCAAGGAGTAACTTCATCCAGAGTGATGCATAGTTCAGATTGA
AATGGCTGCTTGCAGTAGCACAAGCTGGCCACTACAACAATTTATGATTGGGATACAAAG
CGAAACCACCTGATTGGAACTTGGAAGAATTAAAGAATGAAAGTCGTCTTGATTCAGATA
CTAACCAACTGAAAACTTGTATTTCTCCAATTTTCGTGTTTCTGCCTTTCCAGGCAGTGT
GTGCCTTGTAGTCTACTGTTGTTCAGAAGAGTGCCACCCAATTGTGGAATGTACAAACAT
TGAATAATTATATGATGTCACTGATGCCATTTTTTAGACTTACCTGGATTGGATTCTTCC
CCCAAATTCCCATTCTATAAAGCAAAATGCTAATAGCTTATACATTTATATCCCCTGTTG
GAAGATAGAAACTGAAGGTGGCATATTGACATCGTTAGCATTTTCCTGCAAATACATGTT
ATTGTATTTTTGGGCAAAAGCAGACTGATACGGGTACTCTGTCTTTTTCAGAATACATTT
TTGGGGCTGAAGCCAGATCAAAAATAATTTGCTTGTAGCTTGCCTGTGGCCCTGAGCAAA
GGAAGATTTCACACTTTGGGTGAGACATACATCTTCCCCCCTGTTATTCCAGTTCGTATA
CAATAGTTACTCTGGGAAACTGTCTGTTCGGATCATACCATGAGCTAGTTTTTGTTCTTG
TGCCTTTTGTTCTTATGTAAGCTTATGTGTCATCTTTCTTTGTTTGTGAAGAACCATTCT
TTTTAAGGTCACATTTCTCTAAAAGGCCAATGACATGCATGCTGTTTAATTTTATCATGT
GTGCAAATGGTTATACATTTGTTTGTTTGTGTTATTCTCGGTGCACTAAGAATAGCTTGG
TCAACTTCTTCCTCAAAAGAGAAAAATGACCGGCTGATAACAATGGCATTTAAATAACTC
TTCTGTAAATGTCTTGGGCTATTGTTAGTCACTAACTATTATT