Theses and Dissertations

12-2010

Applications of Site-Specific Recombination Systems in Transgene Expression and Marker Gene Removal

Mehmet Aydin Akbudak
University of Arkansas

**APPLICATIONS OF SITE-SPECIFIC RECOMBINATION SYSTEMS
IN TRANSGENE EXPRESSION AND MARKER GENE REMOVAL**

APPLICATIONS OF SITE-SPECIFIC RECOMBINATION SYSTEMS
IN TRANSGENE EXPRESSION AND MARKER GENE REMOVAL

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Cell & Molecular Biology

By

Mehmet Aydın Akbudak
Selcuk University
Bachelor of Science in Crop Science, 1999
Selcuk University
Master of Science in Crop Science, 2002

December 2010
University of Arkansas

ABSTRACT

Site Specific Recombination systems, such as FLP–*FRT* and Cre–*lox,* have been successfully used for site-specific gene integration and marker-gene deletion in plant systems. They are very useful tools in the integration of single-copy full-length transgene cassettes into the genome because the transgene integration via conventional methods often generate multi-copy locus. Such complex locus containing direct and inverted repeats of full-length and truncated copies of the transgene cassette generate aberrant RNA resulting in gene silencing. Therefore, for stable gene expression, a single copy transgene locus is preferred. However, even single copy locus sometimes succumbs to gene silencing. Although the mechanism is not very well understood, it is thought that transgene expression above a threshold level triggers gene silencing. Therefore, it is important to study the effect of transgene copy number on gene expression, and to control the locus structure and integrate full-length copies. In the present study, Cre–*lox* site-specific recombination system was used for integration of 1 – 3 C of green fluorescent protein (GFP) or β-glucuronidase (GUS) genes into a pre-determined integration locus in rice cells. Expression analyses revealed a clear 2 – 4 times increase in GFP and GUS productions correlated with transgene copy numbers (1 - 3C).

As a next step towards the practical implementation of this technology, a molecular strategy was developed for generating marker-free site-specific gene integration. This strategy relies on Cre–*lox*-mediated gene integration followed by FLP–*FRT*-mediated marker excision. The molecular strategy is designed to generate an integration locus consisting of strategically-placed *FRT* sites to remove marker genes.

In the original strategy, an inducible FLP–*FRT* system was included to control the marker excision step. This strategy was tested on two integration lines resulting in poor to undetectable excision of marker genes. In the subsequent modifications involving re-transformation of the integration lines with the improved version of FLP gene, called FLPe, marker excision was detected in the retransformed lines. The excision footprint was detected by PCR and Southern analysis in most of the lines, and excision efficiency determined in the selected two lines by real time PCR as 75 and 100%.

This dissertation is approved for
Recommendation to the
Graduate Council

Dissertation Director:

_____

Dr. Vibha Srivastava

Dissertation Committee:

_____

Dr. James McD. Stewart

_____

Dr. J. Brad Murphy

_____

Dr. Ralph L. Henry

_____

Dr. Cynthia L. Sagers

DISSERTATION DUPLICATION RELEASE

        I hereby authorize the University of Arkansas Libraries to duplicate this dissertation when needed for research and/or scholarship.

Agreed   _____

        Mehmet Aydın Akbudak

ACKNOWLEDGEMENTS

DEDICATION

To my dear grandmother, Vesile Huner,

who loved me and my sister more than anything in her life…

TABLE of CONTENTS

LIST OF PAPER(S)

Chapter II has been published in Molecular Biotechnology.

Akbudak MA, More AB, Nandy S and Srivastava V (2010) Dosage-Dependent Gene Expression from Direct Repeat Locus in Rice Developed by Site-Specific Gene Integration. *Molecular Biotechnology* **45**:15-23.

CHAPTER I

Introduction

Human beings have depended on seas, pastures, and agricultural areas as food sources throughout history. According to the statistics, however, the oceans and pastures have become less productive in the last 50 years because of over exploitation (FAO, 2009). Fishing has increased five-fold since the middle of the last century, and reached to the limits of its sustainable level today. Meat production increased three-fold between 1961 and 2000, but more recently has been decreasing due to overgrazing (FAO, 2009). Although food production has increased globally thanks to the development of high yielding varieties and the use of fertilizers, this improvement is highly variable between countries and regions (Brown *et al.,* 1999).

The world population almost completely relies on three major cereals, rice, wheat and corn as the primary energy source. To increase the cereal production by opening more area to agriculture is not an option as of today as nearly all arable land is currently in production. For example, cereal harvested area, which was 647 million ha in 1961, increased to 726 million ha in 1981 (FAO, 2009). Since 1981, due to the erosion and the use of the arable areas for non-agricultural purposes, the cereal production area decreased down to 660 million ha at the beginning of this century (FAO, 2009). Also, because the world population is growing rapidly and in parallel with this rapid growth, it is expected

that the cereal harvested per capita, which was 0.22 and 0.11 ha in 1961 and 2000 respectively, will decrease, and will become 0.07 ha in 2050 (Gilland, 2002).

Like soil, water is also a crucial input in agriculture. The growing world population, pollution and global warming make water resources much more limited every day. Projections show that the food production will be dramatically affected by the lack of water in the near future (Seckler *et al.*, 1999; Oki and Kanae, 2006). Chances of finding new water resources are so small that development of drought-tolerant and water efficient (producing more dry-material using unit water) plant varieties seems the only solution for a more sustainable agriculture.

Thanks to the advances in plant breeding in the last century, the yield in the world's agricultural areas has increased 3-fold since 1961. (FAO, 2009). However, this increase has not solved the problem of hunger on earth. In this yield increase that occurred especially in the last 50 years, two critically important stages, called as "revolution", stand out. The first is called "the green revolution," the development of high-yield varieties through use of fertilizers and improvement in agronomical techniques. The second one is called "the gene revolution," the use of genetic engineering methods in plant breeding, and has a strong contribution in the yield increase in the last 15 years.

Varieties that have high yield, high harvest index and profitability have been developed by conventional breeding for over 60 years. This genetic improvement, called "the green revolution" resulted in the development of dwarf rice and wheat varieties, which produce more grain per unit harvested area (Sinclair, 1998). However, "the green

2

revolution" has almost reached its biological borders, and obtaining yield increase via conventional breeding is extremely limited today (Brown *et al.,* 1999).

Therefore, the research effort has been focused on plant biotechnology studies in the last 25 years. Plant biotechnology offers genetic variability, which cannot be achieved by conventional plant breeding methods. Furthermore, it enhances agronomical and other economically important traits of plants at the molecular level. Two of the major applications of plant biotechnology in agriculture, introduction of *Bt* toxin (conferring insect resistance), and glyphosate (an herbicide) resistance genes into major crops, such as corn, soybean and cotton, ushered a new era in plant breeding. For example, the use of the *bt* toxin gene conferred 14% of yield increase in cotton (Reviewed by De Maagd *et al.,* 1999). Today, over 90% of the soybean grown in the United States is herbicide tolerant transgenic soybean. Similarly, over 60 % of the cotton and corn grown in the US in 2009 were either herbicide tolerant or *bt* gene added (USDA, 2009). Studies focused on identification of novel genes using functional genomics approach, and transfer of these genes between species have been growing every day. These studies have given very promising results in fighting with yield-limiting factors such as drought. In one of those studies, Nelson *et al.* (2007) demonstrated that corn varieties transformed by ZmNF-YB2 gene, identified in *Arabidopsis*, showed 50% of yield increase under drought conditions.

Plant genetic engineering has been used not only in agriculture but also in pharmacology because the plant cell can be transformed to produce human proteins and vaccines in vast amounts. The proteins and vaccines produced by plant cells are also cheap and safe as compared to those produced in bacteria and animal cells. Today, there

are many vaccines and proteins available that are synthesized in plant cells, and tremendous effort to introduce new ones to the market are underway.

The research on *Agrobacterium*-mediated gene delivery in 1981 is considered as a milestone in plant transformation and genetic engineering. It made possible identification of many genes from various organisms by expressing them in plants. The *Agrobacterium*-mediated gene transfer technique was followed in the last 30 years by several other transformation techniques, such as particle bombardment, polyethylene glycol (PEG)-mediated fusion of protoplast, electroporation, microinjection etc. Using these techniques, numerous plant species were transformed and released for commercial production.

Because the introduced DNA cannot be integrated into a specific target site in the higher plant genome random integration of the transgene by the host recombination machinery usually generates complex multi-copy insertions (Kohli *et al.,* 1999; Pawlowski and Somers, 1998; Macbool and Christou, 1999). The multi-copy insertions often consist of truncated fragments along with the complete copies of the transgene in direct and inverse orientations (Frame *et al.*; 2002; Grant *et al.*, 2004). Production of aberrant transcript from complex locus then leads to gene silencing; therefore, complex loci are not the desirable outcome in the transformation process (Meyer and Saedler, 1996).

Gene silencing has been reported by many researchers, and it simply can be described as the suppression or down regulation of the transgene. Most of the gene silencing occurs in the first generation; however, an active complex locus may undergo gene silencing in subsequent generations (De Carvalho *et. al.* 1992; Matzke *et. al.* 1994).

Although the gene silencing mechanism has not been completely illuminated, following factors have been found to be associated with gene silencing: a) Introducing the transgene into an unfavorable location in the genome (Matzke and Matzke, 1998), b) Presence of multicopy transgenes in the genome [HDGS: homology-dependent gene silencing] (Ye and Signer, 1996; De wilde *et. al*., 2000), c) Over-expression of the transgene mostly due to strong promoters (Napoli *et al.*, 1990; Van der Krol *et al.*, 1990; Que *et al.,* 1997) and d) Integration of vector backbone into the genome (Kohli *et al.*, 1999; Kononov *et al.*, 1997) have been identified as the primary factors causing gene silencing (Artelt *et al.*, 1991).

Although silencing of a single copy of transgene has also been reported (Elmayan and Vaucheret, 1996), HDGS is frequently observed in the presence of multiple homologous copies of a transgene (Kooter *et al.,* 1999). Complex integration patterns due to multi-copy transgene integrations lead to the formation of dsRNA, which triggers gene silencing in the cell (Hobbs *et al.*, 1990; Assaad *et al.*, 1993; Matzke *et al.*, 1994). It has been reported that transgene expression is inversely correlated with transgene copy number in case of complex integration (Hobbs *et al.,* 1990). Therefore, plants carrying single copy transgene are isolated in plant transformation. Such plants express the transgene at relatively high levels, and transmit stable expression to their progenies in subsequent generations.

The increase in the release of commercialized transgenic plants raises discussions and concerns regarding whether transgenic plants are safe for environment and human consumption. One of the concerns about transgenic plants is related to the presence of DNA fragments that are essential for transformation but unnecessary following the

identification of transgenic plants. Of these fragments, selectable marker genes, especially the antibiotic resistance genes raise serious concerns. Marker genes, which are transferred with the trait gene, are necessary to distinguish the transformed from the untransformed cells in the tissue culture. However, marker genes remain in the genome after the transformation, and continue to be expressed at high levels where they are almost always driven by strong constitutive promoters such as maize ubiquitin promoter or Cauliflower Mosaic Virus (CaMV) 35S promoter. They serve no useful function after identification of transgenic clones, thus potentially adding to environmental pollution and metabolic burden to plant systems (Khan and Maliga, 1999). Horizontal gene transfer of marker genes from transgenic plants into bacteria (via homologous recombination) could potentially occur, which further raises concerns about the presence of marker genes in transgenic plants (Pontiroli *et al.,* 2009). Although, this type of horizontal gene transfer does not take place in a statistically significant level (reviewed by Nielsen *et al.,* 1998), antibiotic resistance gene is considered as a tremendous risk, should bacteria acquire resistance it would make it difficult to treat diseases with those antibiotics. Therefore, strategies to remove marker genes from transgenic plants before their release to the environment are recommended by regulatory agencies (EFB, 2001).

Hence, obtaining marker-free transgenic plants has become a priority of transgenic plant production among the academic community and companies that on this area. There are several approaches developed by different research groups, and they will be discussed in detail in Chapter III.

Transformation strategies allowing site-specific modifications in the genome are important for (1) precise gene integration, which is a pre-requisite for stabilizing gene

expression, (2) marker gene removal. The two most popular tools used for these applications are, site-specific recombination, and zinc-finger nucleases. Advantages, drawbacks and technical considerations of each of these strategies are discussed below.

Site-specific recombination systems

The site-specific recombination (SSR) systems play a vital role in many biological systems by inserting, excising, and inverting DNA segments. Many SSR systems have been described in prokaryotes and lower eukaryotes. Five recombination systems predominantly used in genetic engineering are Cre–*lox*, FLP–*FRT*, R–*RS*, φC31 and Gin–*gix*. They catalyze reactions at unique DNA sites, resulting in exchange/ deletion/ inversion of DNA fragments depending on the orientation/ placement of the recombination sites. SSR reactions are reciprocal, and the relative orientation of recombination sites determines the outcome of the reaction (reviewed by Lyznik *et al.,* 2003; Gilbertson 2003; Ow 2002). If a DNA fragment is flanked by two recombination sites in direct orientation; in the presence of recombinase, the intervening DNA fragment is excised. If the DNA is flanked by oppositely oriented recombination sites, the intervening DNA is inverted (Fig. 1). SSR systems have been utilized for genetic engineering applications such as marker gene deletion (Dale and Ow 1991; Zhang *et al.,* 2003) and transgene integration (Albert *et al.,* 1995; Vergunst *et al.,* 1998) into the plant genome.

Among these SSR systems, the Cre–*lox* of bacteriophage P1 and FLP–*FRT* from 2 μm plasmid of yeast, *Saccharomyces cerevisiae*, are the well characterized (Sauer, 1987; Schwartz and Sadowski, 1990). Unlike transposons, Cre–*lox* and FLP–*FRT* are conservative systems; therefore, there is no nucleotide gain or loss during the reactions. The recombination sites (*loxP, FRT*) are comprised of two 13 bp recombinase binding sites and one 8-bp spacer (Fig. 2). *LoxP* and *FRT* are conserved in evolution, and they have a sequence similarity over 50%. The spacer sequence, which is flanked by binding sites, determines the directionality of the recombination site. Since reactions catalyzed by SSR systems such as Cre–*lox* and FLP–*FRT* are reversible, the integration locus is unstable. This is not a desirable feature in genetic engineering; therefore, mutant recombination sites have been developed that are competent in only forward reactions (Albert *et al.,* 1995; Schlake and Bode, 1994). Such sites may contain mutations in the flanking left (e.g., *lox*71, *lox*75) and the right arm (e.g., *lox*66, *lox*76). A recombination between two mutant recognition sites results in the generation of a wild type and a double mutant recombination (*Dm*) site. This newly generated double mutant site is impaired in recombination, making the reaction quite irreversible. A second type of mutant recombination site consists of spacer mutated recombination sites, which recombine with an identical mutant generating a reversible reaction. Spacer mutated recombination site cannot recombine with alternative recombination sites. Hence, spacer mutated recombination sites are useful when the two recombination reactions should occur in a single system.
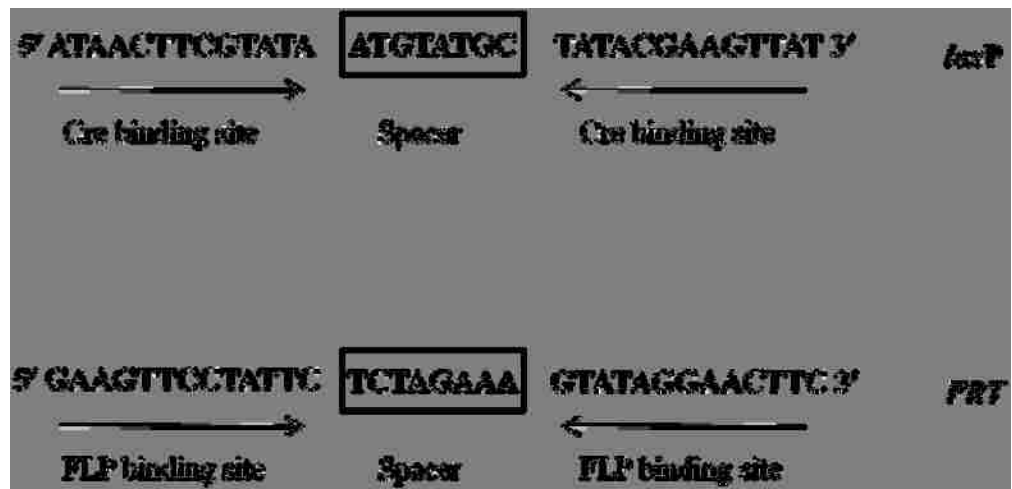
Fig 2. Nucleotide sequences of *lox*P and *FRT* recombination sites.

Cre–*lox* SSR System

Cre–*lox* system is composed of the Cre protein and the 34-bp recombination site, called *lox* (locus of x-over). Cre–*lox* is a highly efficient SSR system in plants, mammalian, and other organisms (Buchholz *et al.* 1996, Raymond and Soriano, 2007). Its functionality in yeast was first shown by Sauer (1987) over two decades ago. To this date, it has been used in the transformation of many plant and animal systems. The Cre gene contains 1032 bp, which codes for 343 aminoacids. Its molecular weight is 38.5 kDa and it is derived from *Enterobacteria* phage P1 (Accession No: AB542060). Cre is a member of the tyrosine recombinase family. The wild type *lox* site is called *lox*P and consists of an 8-bp spacer region flanked by two 13-bp inverted repeats (Fig. 2). Cre protein binds to these repeats, and creates a nick in the spacer sequence to initiate recombination (Chen and Rice, 2003). The orientation of the spacer region, which is unidirectional, determines the type of the recombination reaction. A reaction between two directly oriented *lox* sites results in the excision of the *lox*-flanked DNA as an circular molecule, while a reaction between two oppositely oriented *lox* sites results in the inversion of the *lox*-flanked fragment in the locus. While inversion is freely reversed; deletion is not reversed in practice as the integration of the excised molecule is not kinetically favored. Using these features of Cre–*lox* system, two essential applications in genetic modification of plants have been achieved: site-specific transgene integration (Albert *et al.,* 1995; Day *et al.,* 2000; Srivastava *et al.,* 2004), and marker-gene deletion (Dale and Ow, 1991; Zuo *et al.,* 2001; Hoa *et al.,* 2002; Zhang *et al.,* 2003; Wang *et al.,* 2005).

FLP-*FRT* SSR System

FLP–*FRT* is generally considered the second best site-specific recombination system. It is not as efficient as Cre–*lox*, because Cre has 82-fold higher affinity for its target, *loxP*, than FLP for its target, *FRT* (Ringrose *et al.,* 1998). SSR activity of Cre protein was found to be even higher than the improved versions of FLP gene, FLPe and FLPo (Raymond and Soriano, 2007). However, functionality of FLP–*FRT* has been shown in maize, rice, tobacco (Lloyd and Davis, 1994) and *Arabidopsis* (Kilby *et al.,* 1995). Recently, Li *et al.* (2009b) showed site-specific integration of transgenes (*hpt, yfp, cfp*) via FLP–*FRT* mediated DNA exchange cassette in soybean. The FLP gene contains 1272 bp, which codes for 423 amino acids. Its molecular weight is 46 kDa, and it is derived from 2 μm plasmid of *Saccharomyces cerevisiae*, a species of budding yeast. Like Cre, FLP is also a member of the tyrosine recombinases family. FLP protein functions by binding 34-bp *FRT* (<u>FLP</u> <u>R</u>ecognition <u>T</u>arget) sites. The type of the recombination reaction is determined by the orientation of spacer regions of *FRT* sites as in *lox* sites (Fig. 1).

Zinc-finger Nucleases

Zinc-finger nucleases (ZFNs), which combine a *Fok*I non-specific cleavage domain with polymeric zinc-finger domains, are becoming an attractive tool for site-specific engineering of many organisms and cells (Wu *et al.,* 2007). ZFNs introduce site-specific double-strand breaks (DSB) into the genome, which stimulate homologous recombination

(HR) (Puchta *et al.*, 1993; Wright et al., 2005; Remy *et al.*, 2010). As a cellular response, cells repair DSBs using non-homologous end joining (NHEJ) or HR to maintain their genomic integrity. HR machinery uses the homologous DNA sequence of the undamaged pair of the chromosome as a template. However, if an extra-chromosomal modified DNA fragment is introduced into the cells with the ZFNs, this fragment may serve as a donor DNA for sequence replacement. ZFN based techniques provide many advantages in reverse genetic studies by allowing gene-targeting, as well as site-specific integration of big DNA molecules into the genome. They have been used for the knock-in (Townsend *et al.,* 2009; Hockemeyer *et al.,* 2009; Bozas *et al.,* 2009) and/or knock-out of genes (Carroll *et al.,* 2008; Shukla *et al.,* 2009) in cultured cells of a variety of organisms including maize, *Arabidopsis* and tobacco. ZFN-based targeted mutagenesis has also been achieved in *Arabidopsis* (Lloyd *et al.,* 2005). Several ZFNs show low affinity and specificity for the target sites, and end up binding to secondary sites, and leading to off-target cleavage. Therefore, sustained expression of ZFNs has been found to be toxic to cells. The off-target activity resulting in the unpredictable nucleotide gains and losses in the genome is considered as the main drawback of ZFNs.

Use of Site- specific Recombination Systems in Plant Transformation

Use of SSR systems in plant transformation brings effective solutions to the problems associated with complex transgene integrations (i.e., gene silencing in a single generation or successive generations). Different studies have demonstrated the efficacy of SSR systems such as Cre–*lox*, FLP–*FRT* and R–*RS* in precisely integrating foreign gene

cassette (reviewed by Srivastava and Gidoni, 2010). Some of these studies have also evaluated the expression of GUS gene from site-specific integration locus (Day *et al.*, 2000; Srivastava *et al.*, 2004; Chawla *et al.*, 2006; Nanto *et al.*, 2009). All of these studies found stable predictable expression of GUS in SSI lines through multiple generations. Day *et al.* (2000) reported that 50% of tobacco SSI lines were silenced for GUS expression; however, this silencing was most likely associated with the use of viral promoter. Accordingly, when plant promoter was used for GUS expression, GUS silencing was undetectable in rice SSI lines (Srivastava *et al.*, 2004), whereas the 35S promoter driven GUS gene was occasionally silenced in SSI lines. Inheritance of stable GUS expression from SSI locus in rice was shown by Chawla *et al.* (2006). They also showed that GUS expression doubled in the homozygous progeny of each of the 11 SSI lines examined, a phenomenon not commonly observed in transgenic plants produced by conventional methods. Most recently, Nanto *et al.* (2009) showed the expression stability of GUS gene integrated site-specifically using R-*RS* recombination system.

All of the transformation studies above, focused on the integration of a single transcription unit or a single transgene. However, several agronomic and medically important traits are conferred by more than one gene. Therefore, multigene transformation technology is extremely important for future genetic engineering (reviewed by Daniell and Dhingra, 2002; Halpin *et al.*, 2001). Introduction of multiple gene cassettes and expression of these genes was studied by De Majnik *et al.* (1997). While integration of multiple genes can be obtained by random transformation approach, expression of each gene at optimum levels is not obtained. The reason of gene silencing again lies in formation of complex locus consisting of multiple units (full-length or

truncated). Since SSR-mediated integration consists of precise full-length integration, this method would be suitable for integrating multigene cassette and ensuring optimum expression of each gene unit.

On the other hand, presence of marker genes in transgenic plants is still an obstacle in the public acceptance of transgenic product. Most of the SSR-mediated gene integration approaches utilize marker genes, which cannot be removed unless a removal strategy is incorporated.

The present study attempted to develop solutions for challenges related to (a) multigene engineering, (b) presence of marker genes in site-specific integration locus. The objectives of the study are:

1. Study the expression of multiple gene units integrated into rice genome by Cre-*lox* mediated site-specific integration.
2. Test a design for marker gene removal from site-specific integration locus.

CHAPTER II

Abstract

In the standard plant transformation practice, transgene copy number is often inversely correlated with transgene expression. As the integration locus generated by standard methods is mostly complex, consisting of both full-length and partial copies arranged in direct or inverted repeat configurations, it is difficult to parse the effect of copy number and locus structure. To clearly study the effect of transgene copy number on gene expression, it is important to control the locus structure and integrate full-length copies. In the present study, the effect of transgene copy number on transgene expression in plant cells was determined using rice callus as a model. To generate full-length integrations, Cre-*lox*-mediated site-specific gene integration method was used. Transgenic rice lines consisting of 1 – 3 copies of β-glucuronidase or green fluorescent protein genes were developed. Site-specific integration lines were characterized and subjected to expression analysis. Lines containing 2 or 3 copies of either reporter genes displayed 2 - 4 times higher expression compared to the single-copy lines. Therefore, dosage-dependent transgene expression can be obtained by integrating full-length copies, and site-specific gene integration approach can serve as an efficient tool for generating precise multi-copy integrations.

Keywords: Gene dosage, GUS and GFP expression, Site-specific recombination, Cre-*lox*, Rice transformation.

Introduction

Transgene expression in plant cells is subject to complex regulatory mechanisms. The pattern of integration is one of the major factors influencing gene expression [1–3]. In wild-type background, where gene silencing process is active, an inverse correlation between gene copy number and expression is commonly observed [4–7]; whereas a positive correlation of the same is observed in gene silencing mutant backgrounds [6, 8]. Various factors responsible for initiating gene silencing have been identified with each leading to the formation of a dsRNA [9]. DsRNA can either be directly transcribed from an inverted repeat locus or originate from the secondary processing of over-expressed transcripts [10, 11]. Hence, complex integration patterns consisting of rearranged copies often succumb to gene silencing [12–15]. Therefore, single-copy transgenic plants are generally considered good candidates for long-term propagation as they are more likely to produce stable transgene expression through subsequent generations [15, 16].

We hypothesized that precise full-length integration of multiple transcription-units will produce dosage-dependent transcript level resulting in higher protein production. As each unit in this locus would produce full-length transcripts, it should avoid succumbing to silencing induced by direct transcription of aberrant RNA. However, it may still be vulnerable to silencing as a result of gene over-expression. A number of studies have alluded to a gene silencing pathway consisting of degradation of mRNA expressed above a putative threshold level [11, 17]. We applied Cre-*lox* mediated site-specific gene integration approach to generate precise integration locus consisting of 1, 2 or 3 copies (C) of transgenes, using separate constructs, in a wild-type background to

17

study the effect of gene dosage on expression level. We found that increase in the copy number from 1 to 2 or 3 resulted in the increase in expression to 2 - 4 times. This expression was consistently maintained in callus through extended period of growth in media containing proper selection. While dosage dependent expression has been reliably produced in gene silencing mutants [6, 8], the same in wild-type plants is rarely observed [7]. The data presented here shows that proper gene expression from each unit of the direct repeat locus can occur in wild-type background, if full-length units are integrated. Potential applications of this technology are discussed.

Materials & methods

<u>**Vector Construction**</u>

The plasmid pVS55 [18] was digested with *Hind*III to remove the GUS gene cassette, and ligated with a *Spe*I linker to generate pAM10, which served as the backbone of all integration vectors described in this work. Two separate plasmids containing either GUS or GFP (G) in a cassette consisting of *Xba*I-*35S*pro-*G*-*nos3'*-*Spe*I-*Xb*aI were developed. These *Xba*I cassettes were cloned into the *Spe*I site of pAM10 to develop GUS (pAM11) or GFP (pAA4) vectors. These vectors contain 1C of the transgene and a unique *Spe*I site for subsequent addition of transgene copies. In this manner, 2C (GUS: pAM12, GFP: pAA5) and 3C vectors were developed (GUS: pAM13, GFP: pAA6).

Rice Transformation

Rice tissue culture media and protocols were essentially as described by Hiei *et al*. [19]. The rice line T5 (Taipei-309) that contains a target site as depicted in Figure 1a was used for all rice transformations. T5 has been described earlier [18]. Plasmids pAM11, pAM12, pAM13, pAA4, pAA5 and pAA6 were separately coated on 1 μm gold particles for bombardment of scutellar callus developed from mature seeds of a rice line containing a Cre-*lox* target site [20]. Particle bombardment was performed using a PDS 1000 (Bio-Rad Inc.) gene gun. The bombarded callus was selected in the presence of 100 mg/L geneticin (Gibco BRL). The selected lines were maintained on geneticin containing medium and sub-cultured every month.

Molecular Analysis

Polymerase Chain Reaction (PCR) was carried out on the genomic DNA with primers *a* (5′ TCTACTTCTGTTCATGTTTGTG 3′), *b* (5′ CTCGATGCGATGTTTCGCTT 3′), *c* (5′ GATTAGAGTCCCGCAATTAT 3′) and *d* (5′ CTAATCGCCATCTTCCAGCA 3′) using Taq Polymerase (Promega Inc.) following manufacturer's recommendations. For Southern hybridizations, ~5 μg of genomic DNA was digested with appropriate restriction enzyme, fractionated on 0.8% agarose gel, blotted on nylon membrane and hybridized with ^{32}P labeled DNA probes using standard protocols.
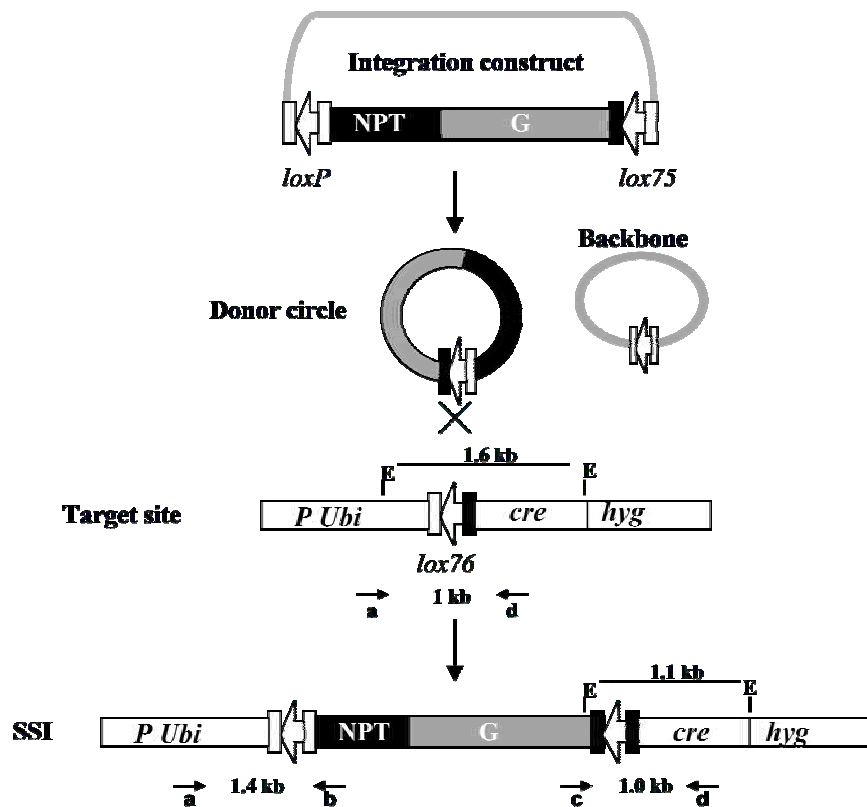
Figure 1a. Molecular strategy of site
-specific integration. (a) Rice target lines containing a single copy of the target site
(*lox76*) and Cre activity are bombarded with the integration construct, which
undergoes *loxP* X *lox75* recombination to separate from its backbone and generate a
donor circle consisting of genes and *lox75*. Integration of the donor circle via *lox75* X
lox76 recombination into the target site generates a precise and selectable site-specific
integration (SSI) locus.
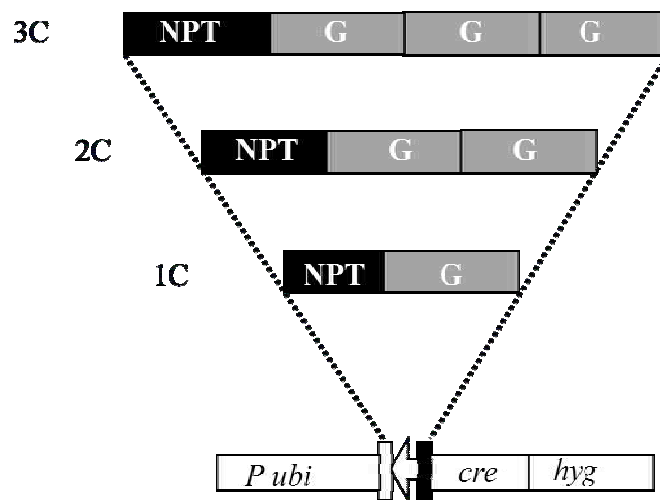
Figure 1b. Molecular strategy of site-specific integration. (b) Allelic series of GUS or GFP (*G*) genes containing 1 – 3 copies (C) of each gene was generated by the site-specific gene integration approach. The target construct consists of a *cre* gene driven by maize ubiquitin-1 promoter (*P Ubi*) and a *lox76* site embedded in the leader sequence of the *cre* gene. In addition, the target construct contains a hygromycin selectable marker (hyg: 35S-hpt-nos). The integration construct consists of a promoterless neomycin phosphotransferase gene (*NPT*) and GUS or GFP gene (*G*) driven by *35S* promoter. Each gene contains nopaline synthase 3' transcriptional termination signal (not shown). The construct is flanked by *loxP* and *lox75* sites in a pBluescript SK backbone. PCR primer (*a-b*, *c-d*, *a-d*) sites and the expected sizes of the amplified fragments are indicated. *Eco*RI (E) sites in *cre* gene and the expected

Expression Analysis

The β-glucuronidase (GUS) activity was detected by incubating callus in GUS stain containing 1 mM X-Gluc (Gold Biotechnologies, St. Louis, Mo.) as described by Jefferson [21]. The GUS activity was measured using the FluorAce β-glucuronidase reporter kit (Bio-Rad Inc., Hercules, CA). Total protein in plant extracts was measured using the DC protein assay kit (Bio-Rad Inc., Hercules, CA). A standard curve prepared with the dilution series of 4-Methylumbelliferone (4-MU) as recommended by the supplier was used to calculate GUS activity. A unit of GUS activity is defined as nmol 4-MU per minute per mg soluble protein.

GFP activity was detected by fluorescence microscope (Nikon Diaphot 300) and photographed by Spot 2 camera (Diagnostic Instruments, Inc) using the software Spot v 4.0.9 (Diagnostic Instruments, Inc). For GFP measurements, callus was ground in extraction buffer (10 mM Tris-EDTA, pH 8.0) at 4°C, centrifuged at 13,000 rpm for 20 min to collect the supernatant. Protein concentrations of the extracts were determined using the DC protein assay kit (Bio-Rad Inc., Hercules, CA). GFP quantification was done using VersaFluor fluorometer (Bio-Rad Inc.) fitted with a 490 ± 5 nm excitation filter and a 510 ± 5 nm emission filter. A dilution series ($0.1 - 1$ mg/ml) of purified rGFP-S65T protein (Clontech Inc., San Diego, CA) was made in the extraction buffer to generate the standard curve. A unit of GFP is expressed as mg GFP per 100 mg soluble protein.

Results

Molecular Strategy

The strategy of Cre-*lox* mediated site-specific gene integration has been described before [18, 20], and depicted in Figure 1a. Briefly, an integration construct containing a promoterless neomycin phosphotransferase II gene (*NPT*) and a gene-of-interest (*G*) between a *loxP* and a mutant *lox75* [22] is introduced into target cell line. The target site consists of a mutant *lox76* [22] present between the promoter and the coding sequence of *cre* gene. Cre activity in target cells facilitates separation of the vector backbone from the construct followed by integration of the construct (donor circle) into the target site via a *lox75* X *lox76* recombination. The resulting integration locus contains a single-copy of the construct consisting of *NPT* and *G*, and expresses *NPT* gene by trapping the promoter of *cre* gene. In the present work, a previously described integration construct, pVS55 [18], was modified to incorporate 1 - 3 copies of either β-glucuronidase (*GUS*) or green fluorescent protein (*GFP*) genes, each driven by a CaMV 35S promoter. Site-specific integration lines containing 1 - 3 copies of each transgene (Figure 1b) were developed using a rice target line previously described [20]. The target locus contains a single-copy of target construct (Figure 1a), and the site-specific integration (*SSI*) locus contains defined junctions (*a-b, c-d*) and restriction map (Figures 1a, 2a).
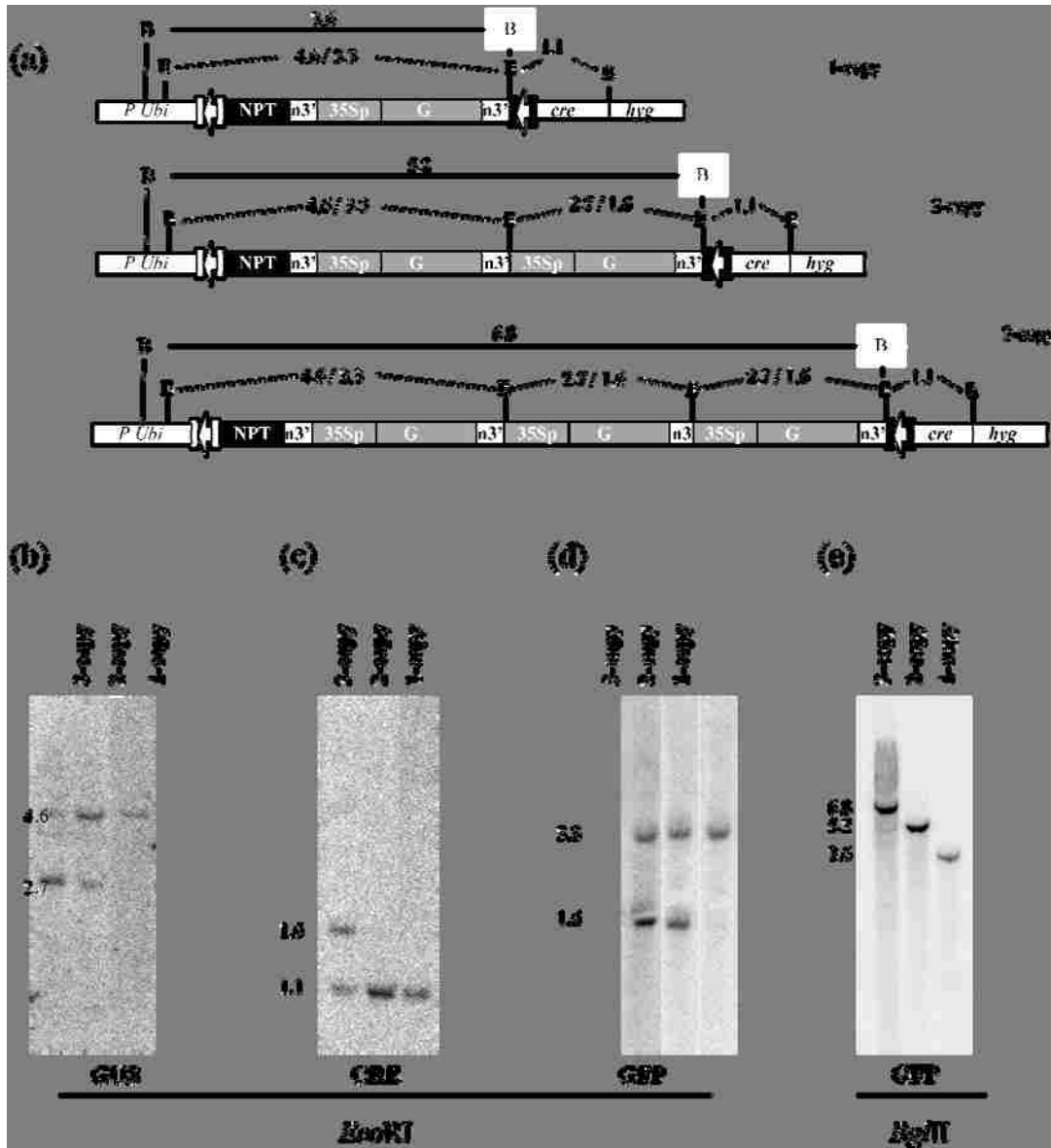
Figure 2. Southern analysis of transgenic lines. (a) *Eco*RI and *Bgl*II maps of the predicted site-specific integration locus containing 1, 2 or 3 copies of GUS or GFP gene. Fragment sizes for GUS (4.6 and 2.7 kb) and GFP (3.3 and 1.6 kb) integrations are shown. (b - e) Southern hybridization of *Eco*RI or *Bgl*II digested genomic DNA of the representative GUS and GFP lines. DNA probes are indicated below each blot, and band sizes are given kb.

24

Characterization of Transgenic Callus Lines

Two different experiments involving bombardment of T5 callus with each of the three constructs (1C, 2C and 3C) of *GUS* or *GFP* genes generated a number of geneticin[TM] resistant lines (Table 1). Bombardments of GUS 1C, 2C, and 3C plasmids generated 16, 12 and 5 geneticin resistant lines, respectively; whereas bombardment of GFP 1C, 2C and 3C plasmids generated 5, 3 and 11 geneticin resistant lines, respectively (Table 1). Each line was analyzed by PCR using primers *a-b* and *c-d* to determine the presence of precise SSI junctions; in addition, primer pair *a-d* was used to determine the presence of target site (Figure 1a). Subsequently, Southern hybridizations on *Eco*RI digested genomic DNA were carried out to confirm the presence of SSI structure, and detect any random integration (Figure 2a-d). The latter may be present in addition to the SSI as 'illegitimate' recombinations can also occur. In GUS experiments, 9 out of 16 one-copy lines, 7 out of 12 two-copy lines, and 2 of 5 three-copy lines contained the predicted integration junctions as revealed by PCR and Southern analysis (Figure 2a, b; Table 1). The remaining lines either displayed a truncated GUS fragment and/or complex integrations (data not shown). Similarly in the GFP experiment, 3 out of 5 one-copy, 3 two-copy, and 5 out of 11 three-copy lines contained precise full-length integrations (Figure 2a, d; Table 1). In addition, Southern analysis revealed the absence of 'illegitimate integrations' in these lines. To confirm the presence of all copies, Southern hybridization was done on *Bgl*II digested genomic DNA of the selected GFP lines (Figure 2a, e). This analysis confirmed the presence of direct repeat locus consisting of 2 and 3 copies (Figure 2e). The relative copy number in these SSI lines was determined using real-time quantitative PCR.

Table 1. Characterization of transgenic lines

Exp. no.	Copy no./ gene	No. of lines[1]	PCR			Southern hybridization	
			a - b	*c - d*	*a - d*	Precise[2]	Comp-lex[3]
1	1/GUS	16	16	16	6	9	7
	2/GUS	12	12	12	5	7	5
	3/GUS	5	4	4	3	2	3
2	1/GFP	5	5	5	0	3	2
	2/GFP	3	3	3	0	3	-
	3/GFP	11	11	11	1	5	6

[1]Geneticin resistant lines,
[2]Presence of expected fragments as shown in Figure 2b-e;
[3]Detection of random and complex integration bands.

Using 50 ng of genomic DNA to amplify transgene fragment in 10 different lines, a linear trend between the estimated copy number and cycle threshold (C_T value) was observed (data not shown). This analysis suggests that the SSI lines contain the expected copy number.

The attempts to regenerate these callus lines failed; therefore, most lines were maintained for up to 2 years by transferring to fresh selection plates regularly. To assess the potential problem of chimerism, defined as contamination of 'untransformed' target cells, three different approaches were used: (1) GFP lines was regularly observed under fluorescent microscope, (2) Southern hybridization with *cre* probe was assessed for the relative intensity of integration and target locus i.e. 1.1 kb and 1.6 kb bands, respectively, and (3) PCR with primers *a-d* was done on all lines (Figure 1a). Fluorescence microscopy suggested that less than 3-mo-old callus of GFP lines 177 and 107 were probably chimeric, while the other GFP lines displayed uniform expression. Accordingly, Southern hybridization of *Eco*RI digested genomic DNA of line 107 detected higher intensity of the target locus band (1.6 kb) compared to the integration locus band (1.1 kb). The remaining GFP and GUS lines either showed equal intensity of the two bands suggesting the presence of hemizygous integration in a homozygous target locus or presence of only the 1.1 kb integration locus band (originating from integration into a hemizygous target locus) (Figure 2c). Finally, PCR analysis was done using primers *a-d* to detect target locus. SSI lines derived from hemizygous locus are not expected to amplify any fragment in this PCR reaction; however a positive amplification was obtained with GUS lines 1-4 and 2-5, indicating chimerism (Table 2). In summary, the analyses above indicated that two GUS lines and two GFP lines were initially

27

Table 2. GUS expression analysis

GUS line	Copy #	PCR ($a - d$)	Chimeric	GUS activity[1] (nmole/min/mg)
1-2	1	-	No	0
1-4	1	+	Yes	3.33 ± 0.57
1-5	1	-	No	2.00 ± 0.00
1-6	1	-	No	3.67 ± 0.57
1-7	1	-	No	2.50 ± 0.87
1-8	1	-	No	4.67 ± 1.15
1-10	1	-	No	6.67 ± 0.57
1-11	1	+	No	4.67 ± 2.08
1-12	1	+	No	4.00 ± 1.00
2-1	2	-	No	12.67 ± 1.15
2-2[*]	2	-	No	-
2-5	2	+	Yes	0
2-6	2	-	No	17.67 ± 1.52
2-7	2	+	No	10.50 ± 3.53
2-8	2	+	No	6.00 ± 1.15
2-9[*]	2	+	No	-
3-1	3	+	No	12.67 ± 2.08
3-4	3	+	No	28.00 ± 6.24

[1]Average values with standard deviation (sdom; n=3).
[*]Lines were lost after initial characterization.

chimeric; however, continued selection on geneticin over a period of 2 years mitigated this problem for 107. Line 177, on the other hand, was analyzed only up to 3-mo-old.

Transgene Expression Analysis

Precise site-specific integration lines were subjected to GUS or GFP expression analysis, as appropriate (Table 2, 3). GUS activity was assessed using histochemical staining and MUG assay, while GFP analysis was done using fluorescence microscope and spectrofluorometer. Two GUS lines, 1-2 and 2-5, did not show any GUS activity upon histochemical staining; and therefore were excluded from further analysis. The remaining 1C GUS lines expressed within a range of 2.00 – 6.67 units (Table 2). Of the five 2C GUS lines, lines 2-2 and 2-9 were lost after initial characterization, and the remaining four expressed GUS activity within a range of 6.00 to 17.67 units. The two 3C GUS lines, 3-1 and 3-4, displayed GUS activity at 12.67 units and 28 units, respectively (Table 2). The expression variation within a group (1C, 2C or 3C) of lines was up to 3X, which is within the range reported earlier for isogenic SSI lines and presumably conferred by somaclonal variations [20]. As a result of this variation, the three groups of SSI lines display overlapping ranges of expression rather than a clear jump (Figure 3). However, all of 2C and 3C SSI lines display a significant increase in expression when compared to complex multi-copy lines generated by random integration (see Figure 3: complex lines), indicating a positive effect of gene dosage.

GFP lines displayed a clear positive effect of gene dosage. The 1C lines produced 0.130 - 0.149 units of GFP; while, 2C lines generated 0.320 - 0.450 units of GFP. The

Table 3. GFP expression analysis

GFP line	Copy #	PCR $(a - d)$	Chimeric	GFP amount[1]
177	1	-	Yes[2]	0.130 ± 0.04
178	1	-	No	0.147 ± 0.03
179	1	-	No	0.149 ± 0.02
180	2	-	No	0.450 ± 0.03
181	2	-	No	0.330 ± 0.01
182	2	-	No	0.320 ± 0.02
107	3	+	Yes[2]	0.655 ± 0.03
154	3	+	No	0.670 ± 0.01
157	3	+	No	0.693 ± 0.04
160	3	-	No	0.745 ± 0.20
162	3	-	No	0.725 ± 0.10

[1]Average values with standard deviation (sdom; n=3).
[2]Based on fluorescence microscopy
GFP titer is given as mg/100 mg soluble protein.

Figure 3. Effect of the GUS gene dosage on expression in rice callus lines.
Error bars depict standard deviation of the mean (sdom; n=3).
*Average expression for complex lines is derived from 5 different lines containing random integration of multiple copies.
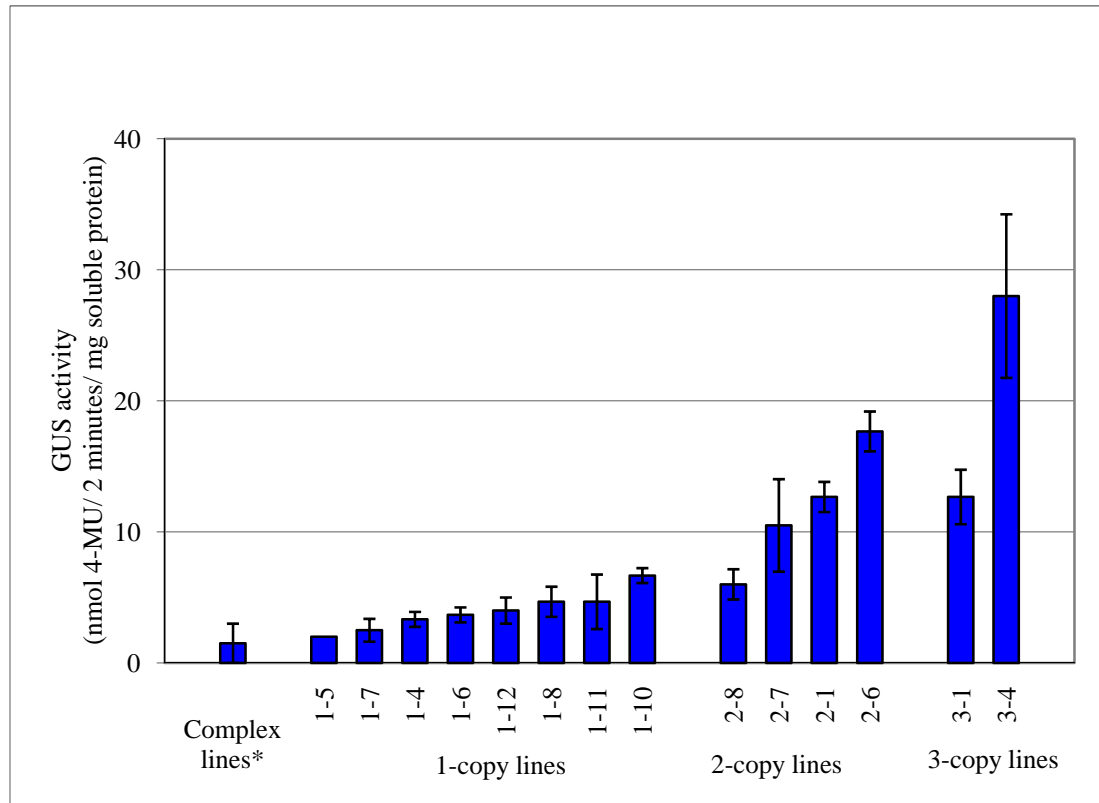
Figure 4. Effect of the GFP gene dosage on expression in rice callus lines.
Error bars depict standard deviation of the mean (sdom; n=3).
*Average expression for complex lines is derived from 8 different lines containing random integration of multiple copies.
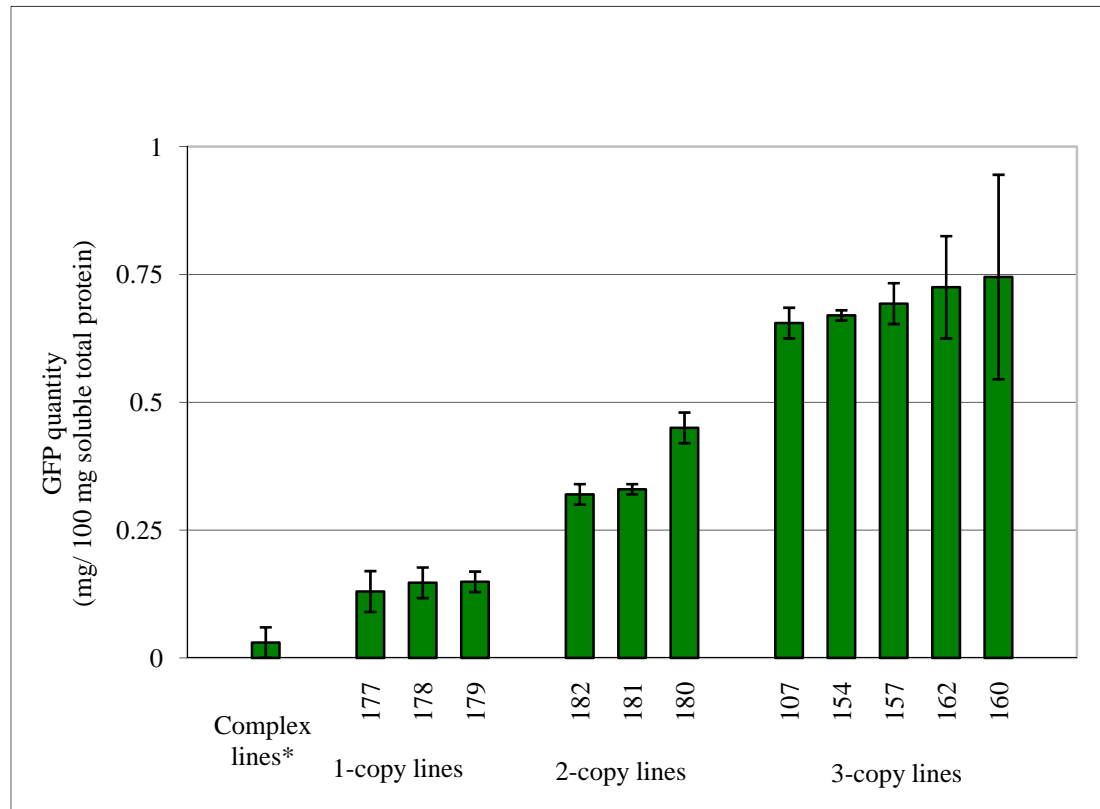
five 3C lines expressed GFP between 0.655 - 0.745 units (Table 3; Figure 4). The quantitative data correlated with the visual GFP expression analysis under fluorescence microscope (Figure 5). As seen with complex GUS lines, complex lines of GFP also contained much lower activity than SSI lines in spite of containing multiple GFP copies (Figure 4).

Next, correlation of gene copy number with transcript dosage was studied using Real-Time quantitative PCR (RT-qPCR). Using 0.5 ng of total RNA, C_T values for amplification of GUS, GFP or phytoene desaturase (PDS; GenBank no. AF049356) fragments were determined, where PDS served as internal control (Supplemental Tables 1, 2). While PDS transcript dosage remained more or less constant in different lines, GUS or GFP transcript levels increased with the increase of gene copy number (Supplemental Figures 1, 2). In summary, both GUS and GFP SSI lines displayed a positive correlation between gene copy number and expression levels (Figures 3, 4). It can be further concluded that transgene expression from SSI locus was stable throughout the growth phase and through sub-cultures on media containing geneticin (100 mg/l). Expression analysis showed that over a period of 2 years, transgene activity was consistent in all of the lines (data not shown). Finally, no detectable gene silencing was found in these lines as a result of extended growth on tissue culture media. Gene silencing was detected in a few lines soon after their isolation; however, all active lines continued to express *GUS* or *GFP* gene at consistent levels through sub-cultures.
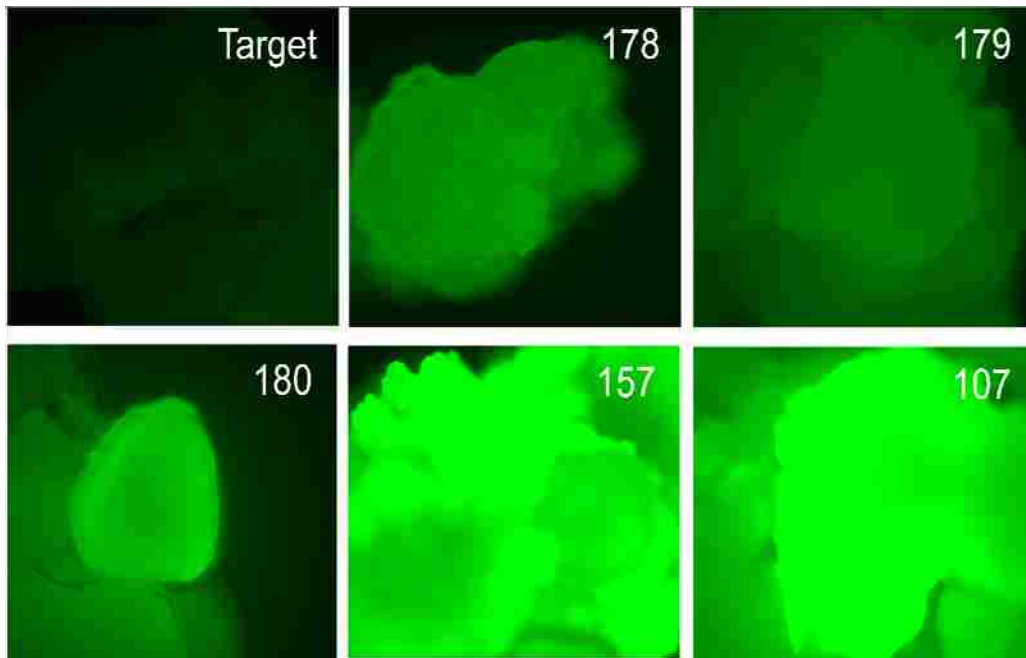
Figure 5. GFP expression on 1-yr-old callus of one-copy (178, 179), two-copy (180), and three-copy (157, 107) SSI lines along with target line callus under fluorescence microscope at 4X magnification. All SSI lines were exposed for 500 msec, and target line for 3 sec to capture the image. The variation in intensity seen in the images is due to multiple layers of focus.

Incidence of Silencing

A previous study on Cre-*lox*-mediated site-specific gene integration in three different rice target locus demonstrated that each line containing a precise integration of the *GUS* gene, driven by maize ubiquitin-1 promoter, expressed GUS activity within 2-3 fold variation [20]. In over 100 lines analyzed, no incidence of gene silencing was observed in cell cultures or in regenerated plants. However, one of the twelve 1C GUS lines and one 2C GUS line generated in the present study completely lacked GUS activity. While base mutations in the *GUS* gene cannot be ruled out, Southern analysis revealed that these lines contained a locus that was structurally identical to that of the expressed lines. Therefore, these lines were considered to be silenced (Table 2). Since the silenced line (1-2) was isogenic to the expressed lines (e.g. 1-4), a stochastic epigenetic process was suspected in establishing gene silencing. Elmayan and Vaucheret [23] reported silencing of single-copy transgenes controlled by 35S promoter. Day *et al*. [24] reported correlation of silencing of 35S-GUS gene with hypermethylation of transgene sequence in isogenic lines. Therefore, methylation pattern of *GUS* gene was assessed in this line and compared with the active isogenic line 1-4 by digesting genomic DNA with *Hpa*II and probing with GUS and 35S promoter fragments. This analysis showed that *GUS* gene in both silenced and expressed lines is unmethylated in *Hpa*II sites (data not shown). Although out of the scope of the present work, a detailed analysis of the silenced locus using bisulfite sequencing may display methylation marks in the silenced lines.

Discussion

The main purpose of this work was to assess the utility of site-specific gene integration technology for expressing transcription units from an integration locus. Earlier, we reported that site-specific gene integration approach is highly reliable for developing a single-copy transgene locus, which is suitable for (a) minimizing expression variability between transgenic lines, (b) ensuring consistent expression through generations [16, 20]. In addition, each transgenic line generated by Cre-*lox* mediated site-specific integration displayed characteristic allelic gene dosage effect i.e. the homozygous T2 progeny of each line displayed 2-fold higher expression compared to their hemizygous siblings [16].

Here, we sought to determine whether a positive correlation of gene dosage with expression level could be obtained from a locus containing 1 – 3 full-length copies side by side. This information would be useful for developing two important applications: (1) boosting gene expression; (2) expressing different genes of a metabolic pathway from a single locus. While other means of boosting expression are available, this approach provides unique advantages. For example, if the trait conferred by a gene could be enhanced by enhancing gene activity, two or more full-length copies could be integrated to achieve higher expression. As expression level of a native gene or an ortholog would be limited by its promoter strength, expression could be boosted by increasing transcription units. However, to avoid gene silencing, it is important to integrate precise full-length copies, which is efficiently achieved by site-specific integration approach. Alternately, a chimeric gene consisting of strong foreign (such as 35S promoter) or native (species specific) promoter could be developed to obtain higher gene activity. However,

use of foreign promoters may not be desirable as future transgenic plants are likely to contain intragenic structures such as native promoters [25], and duplication of the native promoter such as ubiquitin or actin gene promoters may induce ectopic DNA-DNA interactions, leading to gene silencing [26, 27]. A potential problem with increasing transcription units is that the specificity of a native promoter may change due to the effect of regulatory elements in the gene stack. Similarly, the promoter activity of the introduced genes may be deregulated by proximal regulatory elements.

The site-specific gene integration approach described here will also be useful for expressing metabolic pathways in plants. While metabolic pathways involve coordinated expression of multiple genes, deficiency of a few genes is usually responsible for the absence of a given metabolite. Thus, engineering of metabolic pathways quite often requires expression of 1 - 3 different genes in plants [28–31].

In the present work, we used Cre-*lox* mediated site-specific gene integration method to incorporate 1 - 3 copies of *GUS* and *GFP* genes in rice genome. Protein analysis on these lines revealed that 2 - 3 copy locus tends to express at significantly higher levels than one-copy locus (Figures 3, 4). The 2C GUS and GFP lines mostly generated 2-fold higher expression than their respective 1C lines. The 3C GUS and GFP lines displayed 2-4-fold higher expression than their respective 1C lines (Figures 3, 4). Although the production of SSI lines was limited as a result of low transformation efficiency, analysis of the available lines suggested that if not all, at least some 2C and 3C lines will display an increase in expression that is proportional to the gene dosage. Most importantly this work demonstrates that stable expression of multiple transcription units can be obtained with the site-specific gene integration approach. In principle, any

multi-copy (full-length) locus should produce higher expression; however, frequent occurrence of truncation and rearrangement produced by random integration approach leads to gene silencing instead [2, 14, 32]. *Agrobacterium*-mediated transformation generally produces full-length T-DNA integrations. In one study, such multi-copy T-DNA loci were found to display a positive gene dosage effect on transgene expression [7]. However, several other studies have reported gene silencing of single-copy and multi-copy T-DNA locus [1, 23, 33]. As T-DNA integration is based on 'illegitimate' recombination [34, 35], Cre-*lox* mediated site-specific recombination would be a better approach to generate a tandem-array-locus. A potential drawback of the site-specific integration approach is that frequency of precise integration with a large construct, consisting of repeat structures, could be low. What would be the effect of a similar size construct without the repeats? This question needs to be addressed as it is highly relevant for expressing metabolic pathways in plants. Secondly, presence of repeats within a locus may induce homologous recombination [36]; however, the 2C and 3C loci developed in the present study were stable after 2 year of continuous growth in geneticin-containing media (Figure 2e). Additionally, this technology could also serve as a platform for production of pharmaceutical proteins in plant cell cultures. However, cell cultures maintained for long period of time may accumulate harmful somaclonal variations. Therefore for protein production, development of fresh culture from frozen stocks may be a better approach than relying on long-term cell culture.

In conclusion, this work demonstrated that (a) a construct carrying 2 - 3 tandem copies can be reliably integrated by the site-specific integration approach, (b) direct

repeat structures remain stable through constant selection, and (c) precisely integrated gene copies confer additive effect on gene expression

Acknowledgements

References

1. De Wilde, C., Houdt, H. v., Buck, S. d., et al. (2000). Plants as bioreactors for protein production: Avoiding the problem of transgene silencing. *Plant Molecular Biology. 43*, 347-359.

2. Kohli, A., Twyman, R. M., Abranches, R., et al. (2003). Transgene integration, organization and interaction in plants. *Plant Molecular Biology. 52*, 247-258.

3. Matzke, A. J. & Matzke, M. A. (1998). Position effects and epigenetic silencing of plant transgenes. *Current Opinion in Plant Biology. 1*, 142-148.

4. De Buck, S., Windels, P., De Loose, M., et al. (2004). Single-copy T-DNAs integrated at different positions in theArabidopsisgenome display uniform and comparable ß-glucuronidase accumulation levels. *Cellular & Molecular Life Sciences. 61*, 2632-2645.

5. Hobbs, S. L. A., Kpodar, P., & DeLong, C. M. O. (1990). The effect of T-DNA copy number, position and methylation on reporter gene expression in tobacco transformants. *Plant Molecular Biology. 15*, 851-864.

6. Luo, Z. H., & Chen, Z. X. (2007). Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in arabidopsis. *Plant Cell. 19*, 943-958.

7. Schubert, D., Lechtenberg, B., Forsbach, A., et al. (2004). Silencing in arabidopsis T-DNA transformants: The predominant role of a gene-specific RNA sensing mechanism versus position effects. *Plant Cell. 16*, 2561-2572.

8. Butaye, K. M. J., Goderis, I. J. W. M., Wouters, P. F. J., et al. (2004). Stable high-level transgene expression in arabidopsis thaliana using gene silencing mutants and matrix attachment regions. *Plant Journal. 39*, 440-449.

9. Herr, A. J., & Baulcombe, D. C. (2004). RNA silencing pathways in plants. *Cold Spring Harbor Symposia on Quantitative Biology. 69*, 363-370.

10. Tomita, R., Hamada, T., Horiguchi, G., et al. (2004). Transgene overexpression with cognate small interfering RNA in tobacco. *FEBS letters. 573*, 117-120.

11. Vaucheret, H., Béclin, C., Elmayan, T., et al. (1998). Transgene-induced gene silencing in plants. *Plant Journal. 16*, 651-659.

12. Kohli, A., Leech, M., Vain, P., et al. (1998). Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hot spots. *Proceedings of the National Academy of Sciences of the United States of America. 95*, 7203-7208.

13. Pawlowski, W. P., & Somers, D. A. (1998). Transgenic DNA integrated into the oat genome is frequently interspersed by host DNA. *Proceedings of the National Academy of Sciences of the United States of America. 95*, 12106-12110.

14. Kumpatla, S. P., & Hall, T. C. (1998). Recurrent onset of epigenetic silencing in rice harboring a multi-copy transgene. *Plant Journal. 14*, 129-135.

15. Meng, L., Ziv, M., & Lemaux, P. G. (2006). Nature of stress and transgene locus influences transgene expression stability in barley. *Plant Molecular Biology. 62*, 15-28.

16. Chawla, R., Ariza-Nieto, M., Wilson, A. J., et al. (2006). Transgene expression produced by biolistic-mediated, site-specific gene integration is consistently inherited by the subsequent generations. *Plant Biotechnology Journal. 4*, 209-218.

17. Lechtenberg, B., Schubert, D., Forsbach, A., et al. (2003). Neither inverted repeat T-DNA configurations nor arrangements of tandemly repeated transgenes are sufficient to trigger transgene silencing. *Plant Journal. 34*, 507-517.

18. Srivastava, V., & Ow, D. W. (2002). Biolistic mediated site-specific integration in rice. *Molecular Breeding. 8*, 345-350.

19. Hiei, Y., Ohta, S., Komari, T., et al. (1994). Efficient transformation of rice (oryza sativa L.) mediated by agrobacterium and sequence analysis of the boundaries of the T-DNA. *Plant Journal. 6*, 271-282.

20. Srivastava, V., Ariza-Nieto, M., & Wilson, A. J. (2004). Cre-mediated site-specific gene integration for consistent transgene expression in rice. *Plant Biotechnology Journal. 2*, 169-179.

21. Jefferson, R. A. (1987). Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Molecular Biology Reporter. 5*, 387–405.

22. Albert, H., Dale, E. C., Lee, E., et al. (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant Journal. 7*, 649-659.

23. Elmayan, T., & Vaucheret, H. (1996). Expression of single copies of a strongly expressed 35S transgene can be silenced post-transcriptionally. *Plant Journal. 9*, 787-797.

24. Day, C. D., Lee, E., Kobayashi, J., et al. (2000). Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. *Genes & Development. 14*, 2869-2880.

25. Rommens, C. M., Haring, M. A., Swords, K., et al. (2007). The intragenic approach as a new extension to traditional plant breeding. *Trends in Plant Science. 12*, 397-403.

26. Voinnet, O., Vain, P., Angell, S., et al. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. *Cell. 95*, 177-187.

27. Watanabe, K., Pecinka, A., Meister, A., et al. (2005). DNA hypomethylation reduces homologous pairing of inserted tandem repeat arrays in somatic nuclei of arabidopsis thaliana. *Plant Journal. 44*, 531-540.

28. Grotewold, E. (2008). Transcription factors for predictive plant metabolic engineering: Are we there yet? *Current Opinion in Biotechnology. 19*, 138-144.

29. Nawrath, C., Poirier, Y., & Somerville, C. (1994). Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of arabidopsis thaliana results in high levels of polymer accumulation. *Proceedings of the National Academy of Sciences of the United States of America. 91*, 12760-12764.

30. Paine, J. A., Shipton, C. A., Sunandha Chaggar, et al. (2005). Improving the nutritional value of golden rice through increased pro-vitamin A content. *Nature Biotechnology. 23*, 482-487.

31. Zang, Y. X., Lim, M. H., Park, B. S., et al. (2008). Metabolic engineering of indole glucosinolates in chinese cabbage plants by expression of arabidopsis CYP79B2, CYP79B3, and CYP83B1. *Molecules and Cells. 25*, 231-241.

32. Svitashev, S. K., Pawlowski, W. P., Makarevitch, I., et al. (2002). Complex transgene locus structures implicate multiple mechanisms for plant transgene rearrangement. *Plant Journal. 32*, 433-445.

33. Iyer, L. M., Kumpatla, S. P., Chandrasekharan, M. B., et al. (2000). Transgene silencing in monocots. *Plant Molecular Biology. 43*, 323-346.

34. Gheysen, G., Villarroel, R., & Montagu, M. v. (1991). Illegitimate recombination in plants: A model for T-DNA integration. *Genes & Development. 5*, 287-297.

35. Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., et al. (1991). T-DNA integration: A mode of illegitimate recombination in plants. *EMBO Journal. 10*, 697-704.

36. Assaad, F. F., & Signer, E. R. (1992). Somatic and germinal recombination of a direct repeat in arabidopsis. *Genetics. 132*, 553-566.

CHAPTER III

Abstract

Site Specific Recombination systems, such as FLP–*FRT* and Cre–*lox,* function efficiently in plant cells and carry out recombination between the introduced *FRT* and *lox* sites, respectively. This has led to the development of two major applications: marker-gene deletion and site-specific gene integration. Both recombination systems have been successfully used to delete marker genes. However, only Cre–*lox* has so far been successfully used in transgene integration into a previously introduced *lox* site in the plant genome. There are several advantages of site-specific gene integration over random integration. Most significantly, site-specific integration locus expresses transgene consistently through successive generations. As a next step towards the practical implementation of this technology, a molecular strategy was developed for generating marker-free site-specific gene integration. This strategy relies on Cre–*lox*-mediated gene integration followed by FLP–*FRT*-mediated marker excision. The molecular strategy is designed to generate an integration locus consisting of strategically-placed *FRT* sites to remove marker genes. In the original strategy, an inducible FLP–*FRT* system was included to control the marker excision step. This strategy was tested on two integration

lines resulting in poor to undetectable excision of marker genes. In the subsequent modifications involving re-transformation of the integration lines with the improved version of FLP gene, called FLPe, marker excision was detected in the retransformed lines. The excision footprint was detected by PCR and Southern analysis in most of the lines, and excision efficiency determined in the selected two lines by real time PCR as 75 and 100%.

Introduction

Thanks to many advances in genetic engineering in the last 10 years, plant transformation has become routine in many laboratories today. During transformation, however, only a few cells receive the foreign DNA, even fewer integrate it into their genome, and the rest remain untransformed due to the absence of highly-efficient plant transformation systems. Therefore, selection of transformed cells from the mass of untransformed ones becomes necessary. Selection is a very time-consuming and expensive procedure, which necessitates using marker genes during transformation process. These marker genes generally confer resistance to an antibiotic or herbicide that inhibit the growth of untransformed cells, and allows growth of the transformed cells in the culture. For this purpose, antibiotic resistance genes such as neomycin phosphotransferase II (*nptII*), hygromycin phosphotransferase (*hpt*), and herbicide resistance genes such as phosphinothricin acetyl transferase (*bar*) and glyphosate resistance (*gox*) have been extensively used in transgenic plant production.

Identification of transgenic plants without the use of the selectable markers has also been proposed by several research groups. De Vetten *et al.* (2003) reported

46

transformation of potato without the use of any selection markers using *Agrobacterium* strains AGL0 and LBA4404. In five independent experiments, De Vetten *et al.* (2003) inoculated ~8000 stem explants of potato variety Karnico with the two *Agrobacterium* strains containing a marker-free vector to obtain the transgenic lines. Using PCR detection method, they analyzed leaf and stem tissues of regenerated 5017 shoots and found that the average transformation efficiency was 4.5% for AGL0 and <0.2 % for LBA4404. They also conducted a similar experiment on embryogenic callus of cassava, which is another vegatatively propagated crop, and obtained similar transformation efficiency, 4.7%. Although the transformation rate is in an acceptable range, the approach was only tested on two plant species that regenerate many shoots per explant. To reach the same transformation and regeneration efficiency for other plant species might require inoculation of a large number of explants, which is impractical. Even though marker genes have no effect on transformation efficiency, as mentioned before, they are very useful to distinguish transformed cells from untransformed ones, which is a labor intensive process. As seen in this research, De Vetten *et al.* had to grow and analyze 4791 untransformed shoots along with 226 transformed shoots. Since no further analysis was conducted on the regenerated plants in this research, chimeras, which may not transmit the gene into the next generation, could not be ruled out. Finally, screening >5000 candidates with PCR is an extremely expensive approach for routine transformation projects.

Ahmad *et al.* (2008) conducted a study to develop marker-free potato plants transformed by superoxide dismutase (SOD) and ascorbate peroxidase (APX) genes. In seven independent experiments, they inoculated 500 auxiliary buds with *Agrobacterium*

harboring a vector containing SOD and APX gene cassettes. Using PCR, they screened 556 regenerated shoots, and identified 12 plants which were transformed in both transgenes, for an average transformation efficiency of 2.2%. Although the successful transcription of the transgenes was confirmed by RT-PCR in all 12 plants two selected lines revealed by Southern analysis to possess multi-copy insertions of the T-DNA. This is undesirable in transgenic plants because gene silencing associated with multi-copy loci may occur not only in the first generation but also in successive generations. In another study, Li *et al.* (2009a) transformed tobacco leaves using three different binary vectors harboring the GUS gene, which allowed them to produce shoots in the absence of selective compounds. They conducted GUS histochemical assays on leaves, shoots and seedlings, and calculated the transformation efficiency as 35.1, 3.1 and 2.2% for the GUS^+ leaves, GUS^+ shoots, and GUS^+ T_1 seedlings, respectively. They reported that 91% of the GUS^+ T_0 plants generated by the most effective binary vector were also PCR positive, and one-third of the GUS staining and PCR positive shoots were chimeric or escapes.

The low transformation efficiency in transgenic plant production, therefore, necessitates using marker genes, mostly antibiotic resistance genes. Further, some type of screening is necessary such as GUS staining or PCR. However, GUS gene would not be desirable in transgenic plants, and PCR screening would be too expensive. Marker genes are inexpensive approach to solve these problems; however, they are not desirable in transgenic plants due to several concerns discussed below:

a) *Health-related concerns:* Transfer of antibiotic resistance genes from transgenic food into the bacteria in human or animal gut is considered a potential risk, because it would make it difficult to treat diseases with those antibiotics. As reviewed by Nielsen *et al.* (1998), plant genes can be transferred into other organisms via horizontal gene transfer: Smalla and co-workers isolated more than 5000 kanamycin resistant bacteria from soil samples of fields where transgenic sugar beet containing *nptII* was grown. Further, Pontiroli *et al.* (2009) showed that the *aadA* gene (conferring resistance to spectinomycin and streptomycin) transferred from transplastomic tobacco into bacteria. They inoculated intact and ground leaves of wild-type and transplastomic tobacco plants with *Acinetobacter baylyi* harboring two types of plasmids containing chloroplastic gene sequences. After 15 days of the co-cultivation at 28°C, they isolated the bacteria from the decaying and intact plant tissues. The spectinomycin resistant colonies were analyzed with PCR and fluorescent microscope, showing *aadA* gene completely and partially transferred to the bacteria from the transplastomic plants by homologous recombination. Additionally, some scientists are still cautious about the medical implications of consuming transgenic plants, mainly due to the presence of marker genes, although there is no report showing plant DNA can be transferred, integrated and expressed in mammalian cells. On the other hand, the transfer of the antibiotic genes into bacteria is enough to raise medical concerns.

b) *Metabolic burden:* Plastid transformation, which offers some advantages (high expression, maternal inheritance and multi-gene engineering) over nuclear

transformation, has been used in the production of transgenic plants. However, expression of tens of thousands of marker gene copies in a transplastomic line is a heavy metabolic burden for the plant system. It has been reported that the marker protein may form as much as 10% of the total soluble protein in a cell in some cases (Maliga, 2002). Although chloroplasts have their own DNA, it is also possible that transplastomic DNA can be transferred into nuclear DNA, based on the endosymbiotic theory. This may result in vertical and horizontal transfer of marker genes to the other organisms due to unforeseen escapes from the choloroplast DNA to the nuclear DNA.

c)   *Gene stacking:* Currently, most transgenic plants are modified in the single-gene traits. Introduction of polygenic traits and engineering biochemical pathways require multi-gene engineering or gene stacking. However, introduction of multiple genes in a single transformation step leads to rearrangements and truncations of the introduced genes, which makes one by one transformation necessary. For each transformation step, a marker gene would be needed; however, the number of marker genes available for plant transformation is limited. Reviewed by Halpin (2005), this is one of the challenges in gene stacking, which requires marker gene deletion.

d)   *Agronomic and environmental concerns:* Vertical transfer of herbicide resistance genes from transgenic plants to their weedy relatives is considered a major risk necessitating marker gene removal from transgenic plants. For example, gene

flow from cultivated rice to wild *Oryza* species and red rice was reported by many groups (Song *et al.*, 2003; Gealy *et al.*, 2003; Chen *et al.*, 2004; Messeguer *et al.*, 2004). Besides, as reviewed by Timmons *et al.* (1996) and Daniels and Sheail (1999), undesired elements, such as marker genes and plasmid backbones introduced during the transformation cause genetic pollution.

Because of all these concerns, many regulatory organizations, such as USDA, FAO, EU (European Union), either recommend or require the removal of marker genes from transgenic plants (EFB, 2001). To this end, several approaches were proposed to eliminate marker genes in transgenic plants. Among these, three approaches are prominent:

a)   *Co-transformation of trait and marker genes*: In this technique, the marker gene and the trait gene are transferred into the plant cells on independent constructs. Integration of each construct into separate genetic location would allow segregation of the marker gene and the trait gene in the progeny. (Komari *et al.* 1996; Matthews *et al.* 2001; Miller *et al.* 2002; Huang *et al.* 2004; Park *et al.* 2004). Due to various factors, however, the co-transformation and co-integration rates can vary dramatically, which is considered to be the limiting factor for this technique. For example, Zhao *et al.* (2007) co-transformed a selectable marker gene, *bar*, with the non-selected *cecropin*B gene cassettes into rice via particle bombardment, and let transgenic plants segregate in progenies for *bar* and *cecropin*B genes. They found that the co-transformation efficiency was relatively

low, 50~60%. Besides, the production frequency of marker-free plants was 6% in one of the lines, while no marker-free plants were obtained from the other two co-transformed plant lines. Adopting the same approach, Matthews *et al.* (2001) conducted another study that involves the introduction into barley of a plasmid carrying the marker and the trait genes on two separate T-DNAs via *Agrobacterium* transformation. Similarly, they reported a low frequency for the isolation of marker-free lines in the next generation (i.e., 16% of all transformed plants).

b)   *Transposon-based marker removal method*: This technique is based on co-transformation of the trait gene with a marker gene which is inserted into a transposon. Following transposition, the marker gene leaves its original integration locus, allowing generation of marker-free transgenic plants. Variable rates of transformation efficiency, genomic instability and imprecise excision are classified as the drawbacks of this technique. Cotsaftis *et al.* (2002) showed the use of transposon-mediated retransposition of transgene to generate marker-free rice plants. They transformed scutellar calli of rice with a T-DNA harboring ubi-*cry1B* gene (trait gene) cassette flanked by minimal terminal inverted repeats of *Ds* followed by an *AcTpase* gene driven by a constitutive promoter (35S). Sixty-eight independent rice transformants were generated in this study, and excision and reinsertion of *Ds-cry1B* occurred at 37% and 25% respectively in the T_0 generation. They analyzed five independent transformants harboring 2 – 4 reinserted *Ds-cry1B* copies in T_1 progeny which revealed 0.2 to 1.4 new

transpositions per plant. Further, the segregation of the germinally inherited *Ds-cry1B* element resulted in five marker-free, T-DNA free, high *cry1B* expressing lines out of seven actively transposing lines.

c)   *Site-specific recombination (SSR)-mediated marker gene deletion*: This method is considered as the most promising method for generating marker-free transgenic plants (Luo *et al.* 2007, Hohn *et al.* 2001). The simplest strategy to remove marker genes using SSR is to introduce a construct consisting of a marker gene flanked by recombination sites. With such a strategy, recombinase may be provided into the cell by retransformation or by crossing T_0 plants with recombinase-expressing plants. Recombinase-mediated excision of marker gene results in the formation of a marker-free transgenic locus. Recombinase gene can be segregated in subsequent generations. This approach has been further streamlined by the use of conditional/ inducible promoters. In this approach, both marker gene and the recombinase gene are flanked by the recombination sites. Induction of recombinase gene results in the excision of both genes and formation of a marker-free transgenic locus. Inducible promoters such as heat (Wang *et al.*, 2005) and chemical inducible (Zuo *et al.*, 2001) have been successfully used for this purpose.

Site-specific recombination systems are also used for the precise, site-specific integration (SSI) of foreign genes into a pre-determined genomic locus (as described in Chapter II). SSI approach produces higher number of transgenic lines that express the

gene at high levels through subsequent generations, when compared to the conventional transformation approach (Srivastava *et al.*, 2004; Chawla *et al.*, 2006). The mechanism of SSI is very simple. It starts with the introduction of a DNA fragment into cells, having a specific target site. Expression of recombinase gene in the cells drives the integration of the incoming DNA into the target locus, generating a precise SSI locus. The mechanism of SSR systems is introduced in detail Chapter I (Fig. 1) and II (Fig. 1a). Precise integration or gene targeting into a pre-determined locus is especially important in preventing unpredictable transgene expression and recurrent gene silencing through successive generation, frequently seen when transgenes are randomly integrated. Transgene expression variation in successive generations is not acceptable in commercialized varieties; however, the random integration mechanism often generates complex insertions, which are prone to gene silencing. SSR systems, which catalyze precise recombination reactions in the plant genome without any nucleotide gain or loss, are very useful in obtaining stable expression in transgenic lines

The use of site-specific recombination systems has been tested for precise integration of DNA fragments into a pre-determined locus by many groups, and has been reported as efficient in generating transgenic plants expressing the transgene at a predictable level (Albert *et al.*, 1995; Srivastava and Ow, 2002; Srivastava *et al.*, 2004).

Similarly, site-specific recombination systems enable marker gene deletion as discussed above. The present project aims to combine these two applications of site-specific recombination systems into a single transformation technology. A strategy for this technology was proposed by Srivastava and Ow (2004) (Fig. 3). In this strategy, two SSR systems are used. First, a donor plasmid is integrated by using the first SSR system

into target cells, and then, undesirable DNA elements are removed by the use of the second SSR system from the SSI locus, leaving a marker-free SSI locus behind. Thereby, two separate applications of SSR technology are combined to generate marker-free site-specific gene integration system. In the present research, we tested the strategy using Cre–*lox* system for SSI and FLP–*FRT* for DNA excision. The detailed strategy is described below.

Molecular Strategy for Marker-free Site-specific Gene Integration

This strategy is based on the use of Cre–*lox* and FLP–*FRT* systems together. Cre–*lox* system is dedicated for the site-specific integration step. Mutant *lox* sites, *lox*75 and *lox*76 are intended to stabilize the integration structure. Heat-inducible FLP–*FRT* system, present in the target locus, is dedicated for the marker deletion step. The target lines are developed by transforming rice with pAK7 construct (Fig. 1). Single-copy pAK7 transformants are then bombarded with pRP4 (Fig. 1), and the bombarded callus is selected on Geneticin for isolating SSI events. In the SSI locus, unneeded DNA is flanked with *FRT* sites, which can be removed by inducing *FLP* activity at 42°C. The resulting marker-free locus will contain the gene-of-interest (GUS) flanked by *FRT* sites.

Our objectives were:

1- To test the efficiency of soybean heat-shock promoter, HSP17.5E, for inducing FLP-*FRT* recombinations

2- To generate SSI lines, and study FLP-mediated excision upon heat treatment.

3- To test the combined use of Cre–*lox* and FLP–*FRT*, for plant transformation

55

a)

pRP4
8.3 kb
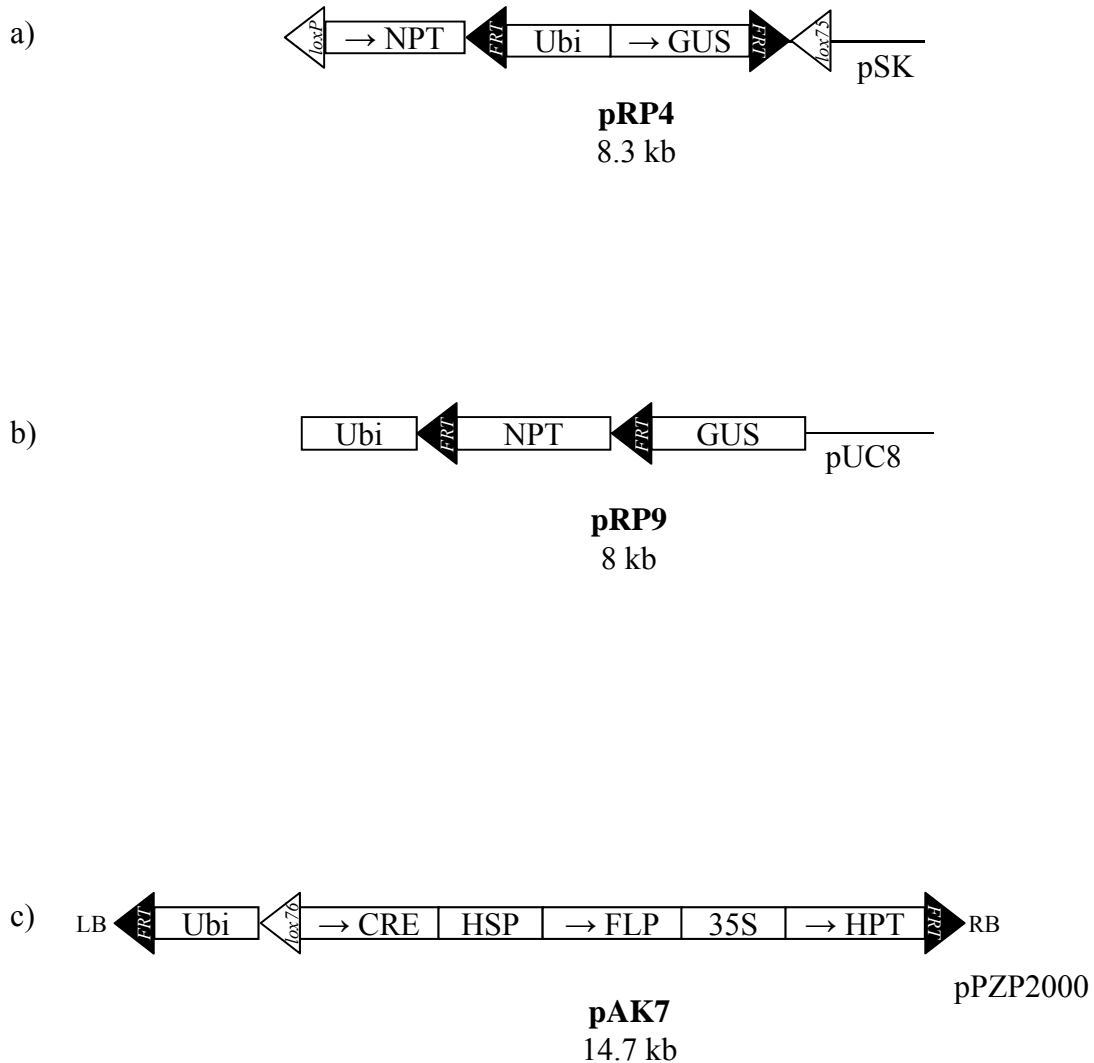
b)

pRP9
8 kb

c)

pAK7
14.7 kb

Fig 1. Diagrammatic illustration of recombination constructs. *Ubi*: maize ubiquitin promoter; *Npt*: neomycin phosphotransferase II gene; *GUS*: β- glucuronidase gene; *FLP*: FLP recombinase gene; *FLPe*: enhanced FLP recombinase gene; *CRE*: Cre recombinase gene; *35S*: cauliflower mosaic virus 35S promoter; *hpt*: hygromycin phosphotransferase gene; *hpt*: hygromycin phosphotransferase; *HSP*: soybean heat-shock promoter (HSP17.5E); *LB*: left border; and *RB*: right border. Each ORF is followed by *nos3′* (nopaline synthase transcription terminator) (not shown). Arrowheads represent recombination sites: FRT or lox as indicated. Arrows show direction of transcription unit.

56

The following experimental steps were carried out:

a)   Develop target cell lines: Target vector, pAK7, was introduced into rice cells by *Agrobacterium*-mediated transformation, and the resulting transformants were analyzed for gene copy number and recombinase activity.

b)   Develop site-specific integration lines: Donor construct (pRP4) was introduced into target cell lines by particle bombardment, and the resulting lines were analyzed for SSI locus structure.

c)   Study FLP-mediated excision in SSI lines and analyze the resulting locus.

Materials and Methods

<u>**Plasmids**</u>

Four plasmids, pAA7, pAA8, pAA9 and pAA10 were constructed for this project (Fig. 2, and Appx. Fig. 1 – 3). pVS55, pAK7, pRP4, pRP9, pUbi-FLP and pUbi-Bar were already available in our lab when the project was initiated (Fig. 1 – 3 and Appx. Fig. 2). PG35-FLPo and PG35-FLPe were kindly provided by Dr. James Thompson (Appx Fig. 1). pRP9 was used during co-bombardment to detect FLP, FLPo and FLPe activities *in vivo*.
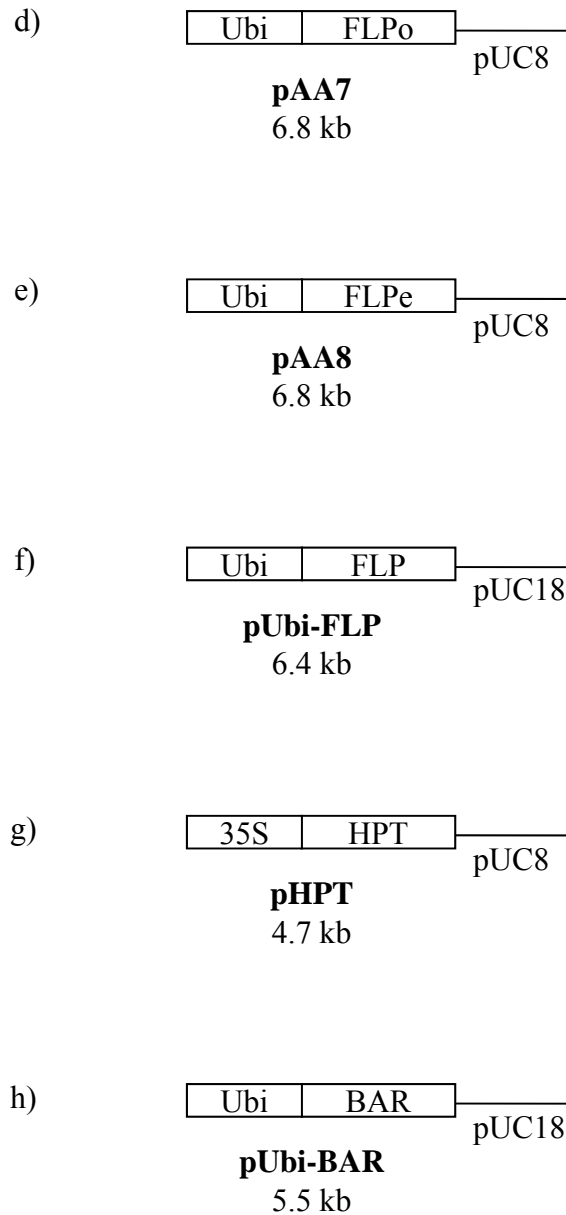
Fig 2. Diagrammatic illustration of plasmids, (d) pAA7, (e) pAA8, (f) pUbi-FLP, (g) pHPT and (h) pUbi-BAR. *Ubi*: maize ubiquitin promoter; *Npt*: neomycin phosphotransferase II gene; *GUS*: β- glucuronidase gene; *FLP*: FLP recombinase gene; *FLPe*: enhanced FLP recombinase gene; *FLPo*: codon optimized FLP recombinase gene; *35S*: cauliflower mosaic virus 35S promoter; *hpt*: hygromycin phosphotransferase gene; *Bar*: bar gene. Each ORF is followed by *nos3′* (nopaline synthase transcription terminator) (not shown).
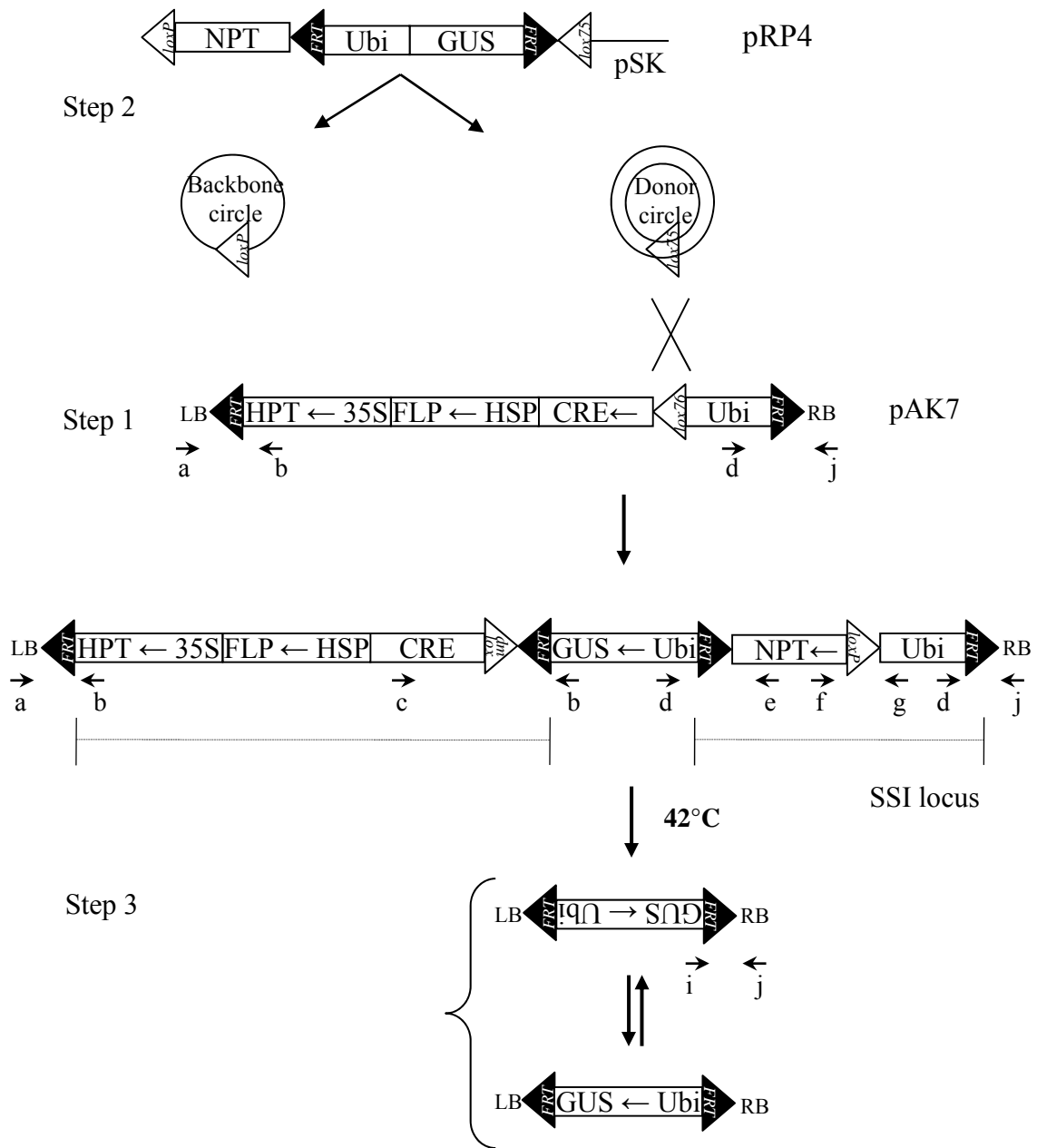
Fig 3. Marker-free site-specific gene integration strategy based on Cre-lox-mediated gene integration followed by FLP-FRT-mediated marker excision. Experimental steps (1-3) are shown that involve generation of target line with pAK7, transformation of target line with pRP4, and heat-treatment of site-specific integration (SSI) line to generate the marker-free locus in two possible orientations. Small arrows show primer (a-j) positions. Location and orientations of recombination sites (FRT and lox) are shown by arrowheads. *Ubi*: maize ubiquitin promoter; *Npt*: neomycin phosphotransferase II; *GUS*: β- glucuronidase; *35S*: cauliflower mosaic virus 35S promoter; *hpt*: hygromycin phosphotransferase; *HSP*: soybean heat-shock promoter (HSP17.5E); *LB*: left border; *RB*: right border; *Dm lox*: double mutant lox site. Each ORF is followed by *nos3'* (nopaline synthase transcription terminator) (not shown).

59

Construction of pAA7 and pAA8

The construction of the plasmids, pAA7 and pAA8 are illustrated in Appx. Fig. 1. The plasmid pRP7 was digested with *Sma*I and *Sac*I to remove the GUS gene. PG35-FLPo and PG35-FLPe were digested with *Asc*I and *Sac*I and then was end-filled by Klenow exo⁻ to obtain FLPe and FLPo genes. pRP7 backbone obtained from the *Sma*I and *Sac*I digestion was ligated with the FLPo and FLPe gene to yield pAA7 (6.8 kb) and pAA8 (6.8 kb) respectively.

Construction of pAA9 and pAA10

The construction of the plasmids pAA9 and pAA10 are illustrated in Appx. Fig. 2 & 3. A 3.2 kb *Hind*III, *Alw*NI and *Xho*I fragment (NPT gene and pSK backbone) of plasmid pVS55; a 0.9 kb *Eco*RI and *Alw*NI fragment (a part of pSK backbone and *lox* sites) of plasmid pVS55 and a 4 kb *Hind*III and *Eco*RI fragment (Ubi promoter and FLPo gene) of PAA7 were used in tri-molecular ligation to yield 8.2 kb plasmid pAA9. Similarly, the same two fragments from plasmid pVS55 and a 4 kb *Hind*III and *Eco*RI fragment (Ubi promoter and FLPe gene) of PAA8 were used in tri-molecular ligation to yield pAA10 (8.2 kb).

Electroporation of *Agrobacterium*

Competent cells of *Agrobacterium* strain EHA105 were transformed with pAK7 to obtain *Agrobacterium* strain for transformation. One µl (0.5 ug) of pAK7 was mixed with 50 µl of EHA105 cells by pipeting, and kept on ice for 1-2 minutes. The whole content of this mixture was transferred into a pre-chilled genePulser cuvette (Biorad Inc, USA), and placed in Micropulser (Biorad Inc, USA), set for 2.5 kv pulse and 5 msec time constant. After a single pulse, one ml LB media (20 g/L, EMD Chemicals, Germany) was added into the cuvvette and incubated for one hour at 28°C. Subsequently, the contents of the cuvette were transferred to 1.5 ml eppendorf tube and centrifuged for two minutes at 14 000 rpm. One ml of the supernatant was then removed from the tube and the rest (100 – 200 µl) was used to suspend the bacterial pellet by pipeting. The suspension was spread on LB agar (35 g/L, Sigma-Aldrich Inc., MO, USA) plates containing Spectinomycin (100 mg/L). The plates were incubated at 28°C overnight. Next day, ~10 colonies were picked from the plates, and subjected to plasmid isolation. The plasmids isolated were digested with 2-4 different endonucleases to verify the vector in the *Agrobacterium* strain.

Agrobacterium-mediated transformation of rice

The standard rice transformation protocol was used to transform Nipponbare callus (Hiei *et al.* 1994). Twenty ml of LB media containing 2 µl of 0.1 M Asetosyringone (AS) was inoculated with the *Agrobacterium* suspension and shaken at 28°C for 1-2 days (until

turbid, OD-600: ~0.8). One ml from this culture was added into 20 ml AAM media containing 20 µl 0.1 M AS in a conical tube (Becton Dickinson, NJ, USA). The calli were immersed in this bacterial suspension for 10 – 15 min. Excess bacteria were removed from the tube and the calli were transferred onto a filter paper (9 cm in diameter, Whatman Grade No. 2 filter paper) sterilized in a petri dish. Subsequently, they were placed on 2N6D media plates containing 0.1 M AS (1 ml/L) and subjected to 2-4 days of co-cultivation at room temperature in dark. The calli were washed twice in autoclaved water, and then washed twice in autoclaved water containing Carbenicillin (500 mg/L) to remove *Agrobacterium*. The co-cultured calli were dried on filter paper and plated on 2N6D media plates supplemented with Carbenicillin (100 mg/L) and Hygromycin (50 mg/L). The ingredients of all tissue culture media used are given in Table 1. The plates were sealed with Parafilm (Pechiney Inc., WI, USA) and incubated at 28°C in dark, and monitored daily for *Agrobacterium* growth. If a portion of callus was covered with *Agrobacterium* growth, that piece was removed from plate and the washes were repeated. The disinfected calli are transferred onto a fresh plate with the uninfected ones. Proliferating hygromycin-resistant calli were re-plated onto the same fresh media after 2 – 4 weeks.

Tissue culture and particle bombardment-mediated rice transformation

Scutellar callus was induced by plating rice seeds (var. Nipponbare) on 2N6D media. Plates were incubated at 25°C under light for 3-4 weeks. Following callus formation, healthy calli were transferred to fresh 2N6D plates for further proliferation, and they were

Table 1. Tissue culture media

Components (per liter)	2N6D	BM	PR	R	MS/2	AAM
N6 basal salt mixture (g)	3.98	3.98	3.98	3.98	2.2	
Myo-inositol (g)	0.01	0.1	0.1	0.1		
Casaminoacids (g)	0.3	1	1	1		0.5
L-Proline (g)	5.75					
2.4-D (10mg/ml) (µl)	200	200				
Sucrose (g)	30	30	30	30	15	68.5
N6 Vitamin (1000x) (ml)	1	1	1	1		
Phytogel (g/500 ml bottle)	1.25	1	1	1	1	
Sorbitol (g)		63				
BAP (1mg/ml)			2	3		
NAA (1mg/ml)			1	0.5		
ABA (5 mg/ml)			1			
MS Vitamins (1000x) (ml)					2	1
pH	5.8		5.8	5.8	5.8	5.2
10x KCl (ml)						100
25x AA Salts (ml)						40
Glucose (g)						36
Glutamine (g)						0.876
Glycine (g)						0.0075
Aspartic Acid (g)						0.266
Arginine (g)						0.174

BM: Bombardment media; PR: Pre-Regeneration media; R: Regeneration media

maintained under the same conditions. Before bombardment, the callus was placed on bombardment media (Table 1) for at least 2 hours. PDS-1000/He system (BioRad) was used for particle bombardment. The plasmid was coated on 1 μm gold particles for bombardment of callus lines. Each transformation experiment consisted of 10 plates.

After bombardment, the calli were left on the bombardment media overnight followed by transfer to 2N6D media the next day. They were maintained on 2N6D media for a week before transferring onto the selection media, which consisted of 2N6D with the appropriate selection agent (100 mg/L Geneticin™ or 5 mg/L biolaphos). The resistant callus was transferred onto pre-regeneration media for 1 week and then regeneration media, which were placed in a growth chamber, maintained at 25°C and under. 23h light 1h dark lightening regime was applied in the growth chamber. The selected lines were maintained on Geneticin containing medium and were sub-cultured every month.

DNA extraction

The UEB method was used to extract DNA from callus tissues (Robinson and Parkin, 2008). Approximately 0.4 g callus was ground in 400 μl UEB buffer using 1.5 μl centrifuge tube and blue pestles. The suspension was extracted with phenol : chloroform and aqueous phase was precipitated with 95% isopropyl alcohol (IPA). Finally, the precipitated DNA was washed with 70% ethanol and dissolved in sterile ddH$_2$O.

Polymerase Chain Reaction (PCR)

Using the UEB method, DNA was extracted from the transformed lines to obtain the template for PCR. The PCR reaction mixture (25 µl) consisted of 5 µl 5x green GoTaq Flexi Buffer (Promega Inc.), 2 µl $MgCl_2$ (25 mM), 1 µl PCR nucleotide mix (0.2 mM of each dNTP), 1 µl forward and reverse primers (25 mM), 0.5 GoTaq DNA Polymerase (5u/µl) (Promega Inc.), 1 µl of template DNA (100 ng) and 13.5 µl sterile water. The PCR was conducted in MyCycler™ thermal cycler (BioRad Inc. USA). PCR reactions consisted of 40 cycles of 1 minute denaturation at 95ºC, 1 minute annealing at various temperatures — depending on melting temperatures (T_m) of primers sets —, 1 minute extension at 72ºC followed by a final extension step at 72 ºC for 15 minutes. The PCR products were maintained at 4ºC until they were fractionated in 0.8% agarose gel. Red™ gel imaging system (Alpha Innotech, USA) was used to obtain gel photographs. To determine the size of PCR products 1 kb ladder was run along with PCR products on the gel. Primer sequences used in this assay are given in Table 2.

Southern Blot Analysis

Approximately 10 µg of genomic DNA digested with appropriate restriction endonuclease was fractionated in 0.8% agarose gel and then transferred to a nylon membrane (Hybond-N+, Amersham Bioscience, UK). DNA on the nylon membrane was fixed using UV Stratalinker® 2400 (Stratagene). The membrane was transferred to a hybridization bottle (Hybond) and pre-hybridized at 65ºC in a hybridization oven for at

Table 2. PCR Primers

Code	Primer	Sequence
a	LB	5'-TTAATGTACTGAATTAACGCCG-3'
b	NOS 1	5'-GATTAGAGTCCCGCAATTAT-3'
c	CRE UAG	5′-CTAATCGCCATCTTCCAGCA-3′
d	RUBI 597	5'-AGGCTGGCATTATCTACTCG-3'
e	KAN F	5'-GCATCGCCTTCTATCGCCTT-3'
f	KAN R	5'-CTCGATGCGATGTTTCGCTT-3'
g	UBI	5'-TCTACTTCTGTTCATGTTTGTG-3′
h	RB 2	5'-GAAGGCGGGAAACGACAATCT-3'
i	GUS F3869	5'-CACCATCGTCGGCTACAG-3'
j	RB	5'-AAACGACAATCTGATCCAAG-3'
k	HYG F	5'-ACCGCGACGTCTGTCGAG-3'
m	HYG B	5'-CCAGTGATACACATGGGGATC-3'
n	CRE ATG	5'-ATGTCCAATTTACGTACCGT-3'
o	KAN 1	5'-GATGGA TTGCACGCAGGTTC-3'
p	KAN 2	5'-AAGGCGATAGAAGGCGATGC-3'
r	GUS EV	5'-CGGTCAGTGGCAGTGAAG-3'
s	GUS 3P	5'-AGTTCATGCCAGTCCAGCG-3'
t	FLP R	5'-TGCGTCTATTTATGTAGGATG-3'
u	FLP FII	5'-GCATCTGGGAGATCACTGAG-3'
v	FLPe F	5'-CGCGCCACCATGAGCCAATTT-3'
w	FLPe R	5'-ATGCGGGGTATCGTATGCTTCC-3'
y	GUS F3380	5'-TTCTGCGACGCTCACACCGAT-3'
z	CREREVATG	5'-ACGGTCAGTAAATTGGACAT-3'
a1	UBI1812	5'-TCTAACCTTGAGTACCTATCTATT-3'
b1	FLP RII	5'-CTCAGTGATCTCCCAGATGC-3'

least 2 hours using 10 ml of pre-hybridization solution (5x SSC, 5x Denhart's solution and 0.5% SDS) and 0.1 mg/ml of denatured herring sperm DNA. Following pre-hybridization process, DNA was hybridized with radio-labeled probes at 65ºC overnight. The radio-labeled probes were synthesized using radioactive dCTP [alpha32P] and Random Primed DNA labeling kit (Roche Inc. USA) as per the manufacturer's instructions. Next day, the membrane was washed with pre-warmed (at 65ºC) wash solutions I (2x SSC and 0.5% SDS), II (1x SSC and 0.25% SDS) and III (0.5x SSC and 0.125% SDS) for 15 minutes each at 65ºC in the Hybaid hybridization oven (Fisher Thermo Inc., USA). After the last wash, the membrane was wrapped in stretch film (Saran™) and then placed in a storage phosphor screen (Molecular Dynamics) overnight. The screen was developed using Storm 540 phosphoimager scanner (General Electric, USA).

Histochemical GUS Assay

A rapid technique for detection of GUS expression, GUS histochemical staining was used to determine the expression of GUS gene in transformed rice. For the detection of GUS activity, the calli were immersed in GUS stain for one hour at room temperature. GUS stain was composed of the following components: 100 ml of 1M sodium phosphate buffer (pH: 7), 5 ml of 0.1M $K_3Fe(CN)_6$, 5 ml of 0.1 of $K_4Fe(CN)_6$, 20 ml of 0.5 M Na_2EDTA (pH: 8), 10 ml of X-Gluc stock (100 mg X-Gluc dissolved in 1 ml N,N-dimethyl formide) in 860 ml of ddH_2O.

Real-time quantitative Polymerase Chain Reaction (RT-qPCR)

Real Time quantitative Polymerase Chain Reaction (RT-qPCR) was carried out on Mx3000P thermocycler (Stratagene). The qPCR reaction mixture for each well (25µl) consisted of 12.5 µl Brilliant II SYBR® Green QPCR Master Mix, 0.5 µl of forward and reverse primers (20 µM), 0.05 µl of ROX (Reference dye), 1 µl of template RNA free DNA (50 ng) and 10.45 µl nuclease-free PCR grade water (Fermentas Inc, Canada). Primer sequences used in this assay are given in Table 3. Following the RNAse treatment, the DNA quantity was determined by NanoDrop (Thermo Scientific, USA). After the mixture was added into qPCR plates, the plates were briefly centrifuged at 4000 rpm to eliminate bubbles in the wells.

PCR reactions consisted of 1 cycle of 10 minute activation at 95ºC, 40 cycles of 30 seconds denaturation at 95 ºC, 1 minute annealing at various temperatures — depending on melting temperatures (T_m) of primers sets —, 1 minute extension at 72ºC, followed by a 1 cycle of dissociation segment consisted of 1 minute denaturation at 95ºC, 30 seconds at 55ºC and 30 seconds at 95ºC for data acquisition. The fluorescence data for amplification was collected at the end of the annealing step (END), while it was collected during the "plateau" of 55ºC to 95ºC (ALL). The PCR products were fractionated in 0.8% agarose gel. Red™ gel imaging system was used to verify if DNA fragments were amplified at the expected sizes. To determine the size of PCR products 1 kb ladder was run along with PCR products on the gel.

Table 3. RT-qPCR primers

Code	Primer	Sequence
q1	CRE F2128	5'-AAATGCTTCTGTCCGTTTGC-3'
q2	CRE R2388	5'-ATTGCTGTCACTTGGTCGTG-3'
q3	GUS F2130	5'-TGATCAGCGTTGGTGGGAAAGC-3'
q4	GUS R2534	5'-GTGGTGTAGAGCATTACGCTGC-3'
q5	NPT F265	5'-AGGGACTGGCTGCTATTGG-3'
q6	NPT R418	5'-GATGTTTCGCTTGGTGGTC-3'

Young leaves and immature embryos were collected from 1.7D plants, which had been obtained from pRP9 transformation (explained in detail on page 89 and Fig.11), were used for FLP transient expression analysis. The leaves were wiped with 70% ethanol a couple of times under laminar hood and cut in approximately 1-cm length with a sterile scalpel before placing onto the bombardment media. Three leaves were bombarded on each bombardment plate. Immature seeds from a 1.7D plant were soaked in a solution containing 30% sodium hypoclorite (NaOCl) + 0.5% SDS (20%) for 15 min and washed 5 times in autoclaved water. Then the seeds were transferred onto a paper tissue autoclaved in a petri dish. Immature embryos were cut out from the seeds using sterile scalpel under a microscope in a laminar hood. Subsequently, the embryos were placed onto bombardment media; each plate had nine embryos.

Gold particles coated by 10 micrograms of pAA7 (FLPo), pAA8 (FLPe), pUbi-FLP (FLP) were individually bombarded onto the leaves and immature embryos using the standard particle bombardment method. Four replicates were considered for each construct for both types of explants. After 60 hours of incubation, the explants were immersed in GUS staining overnight and blue dots on each explants were counted under microscope. The results were subjected to statistical analysis. Data were analyzed using ANOVA (Analysis of variance) and the differences contrasted using Duncan's multiple range test. The statistical analyses were performed at the level of 5%, using SPSS 15.0 (SPSS Inc. USA).

Results and Discussion

<u>Molecular Strategy</u>

In the present study, a two-step strategy proposed by Srivastava and Ow (2004) was tested for SSI followed by the removal of *FRT*-flanked DNA from the integration locus. The schematic diagram of this strategy is shown in Fig. 3. For this purpose, Cre–*lox* was chosen to integrate a transformation vector into a target locus, and FLP–*FRT* was used to remove the unnecessary elements from the SSI locus. Functionality of both SSR systems have been previously shown in different plant systems by several groups (Zhang *et al.*, 2003; Odell *et al.*, 1994; Kerbach *et al.* 2005; Luo *et al.* 2007; Li *et al.* 2009b). In the strategy, *FRT* sites were placed in the target and donor constructs in a way that allows the excision of undesirable DNA fragments on both ends of the SSI locus upon FLP-mediated excision.

To test the strategy, 3 experimental steps were carried out. First, the target construct, pAK7 containing CRE and FLP genes as well as hygromycin resistance gene (*hpt*) as a selection marker gene, was introduced into rice cells. PAK7 contains a *lox*76 site for Cre-mediated integration reaction, and it is also flanked by oppositely oriented *FRT* sites (Step 1). Callus of the target line (line 1A) was bombarded with the integration vector (pRP4) (Step 2). pRP4 contains two *lox* sites, *lox*P and *lox*75, a promoterless NPT gene and GUS gene cassette flanked by oppositely oriented *FRT* sites. Due to the presence of CRE activity in the target cells, pRP4 was expected to split into two circles:

donor circle containing *lox*75 and backbone circle containing *lox*P. Integration of the donor circle via *lox*75 x *lox*76 recombination is expected to generate a selectable (Geneticin resistance) site-specific integration locus. The resulting integration locus contains a heat-inducible FLP gene and four *FRT* sites: two flanking the whole locus and the remaining two flanking the GUS gene. Upon heat-induction, FLP-mediated recombination between the directly-oriented *FRT* sites is expected to delete the intervening DNA fragments (Step 3).

Heat-inducible FLP gene was incorporated to make the system self-sufficient i.e. to avoid further crosses or retransformation with FLP gene. However, the heat-inducible FLP gene was found to be sub-optimal for DNA excisions, and introduction of strong FLP activity became necessary (as described below).

Target Line

Transformation and Molecular Characterization of the Target Line

Several transformation experiments, each consisting of 10 plates, were carried out to generate the target lines. Only one hygromycin resistant line, 1A, obtained from the first experiment was used as the target line in site-specific integration experiments. Generation of additional lines failed in spite of several attempts. Although low transformation efficiency with pAK7 cannot be ruled out, lack of protocol standardization was most likely the reason of poor efficiency. However, one line, if single-copy, was considered to be enough for further work. The callus of the target line was proliferated on

2N6D plates containing hygromycin (50 mg/L) and carbenicillin (100 mg/L) and sub-cultured every month.

Compared to the particle bombardment, *Agrobacterium*-mediated transformation usually generates transgenic plants with lower copy number, intact T-DNA and stable gene expression (Dai *et al.* 2001). However, it is almost impossible to pre-determine how many copies of the transgene will be integrated into the genome. Also truncations in the T-DNA may occur during insertion, and the transgene may be silenced or its expression may further fluctuate upon integration. These factors necessitate molecular characterization of transgenic plants. Therefore, to determine the integrity of the transgenes, DNA was extracted from the line 1A, and was subjected to PCR, Southern analysis and sequencing.

Analysis of the Target Line: PCR and Sequencing

PCR and sequencing were utilized to confirm the presence of the transgenes, and the integrity of the *FRT* and the *lox* sites in the target locus. Using k (HygF) – m (Hyg B), t (FLP R) – u (FLP FII) and c (CRE UAG) – n (CRE ATG) primer sets, HPT, FLP and CRE genes were amplified, respectively. Primers set h (RB2) – d (RUBI597) were used to amplify the expected 0.7 kb band including the *FRT* site in the RB end, while the the expected 1.4 kb band including the *lox76* site was amplified with primers g (UBI) and z (CRE REVATG) (Fig. 3). After the sizes of the PCR bands were confirmed on an agarose gel, the PCR fragments containing the *FRT* and *lox* sites were excised from the gel, purified, and sequenced. The sequences were analyzed, and the DNA elements, such

as gene sequences, restriction sites and recombination sites were located on these sequences. The analysis verified the exact nucleotide matches of each sequence including that of *FRT* and *lox76* sites.

However, the PCR, using a (LB) – b (Nos1) primer set to amplify the DNA fragment containing the *FRT* site on the LB end of the locus failed, showing LB end was truncated. Because the transgenic locus was not mapped on the rice genome, no other reverse primer was available to amplify the LB end. Therefore, the presence of this *FRT* site was not confirmed, and it was presumed that this *FRT* site is also truncated or absent.

Analysis of the Target Line: Southern Analysis

The target line, 1A, was subjected to Southern Analysis. A blot containing *Eco*RV digested genomic DNA of 1A was hybridized with three radioactive probes, HPT, CRE and FLP (Fig. 4). For the single copy insertion of pAK7, CRE hybridization was expected to generate two bands, one of which was at 2.2 kb and another that was bigger than 2.6 kb, while HPT hybridization was expected to generate a single band bigger than 1.7 kb. The expected band showing the integration of FLP gene was at 1.8 kb. The bands generated by the CRE and HPT hybridization showed that pAK7 integrated into the genome as a single copy, while all three hybridizations confirmed their respective genes, CRE and FLP, were present in 1A.
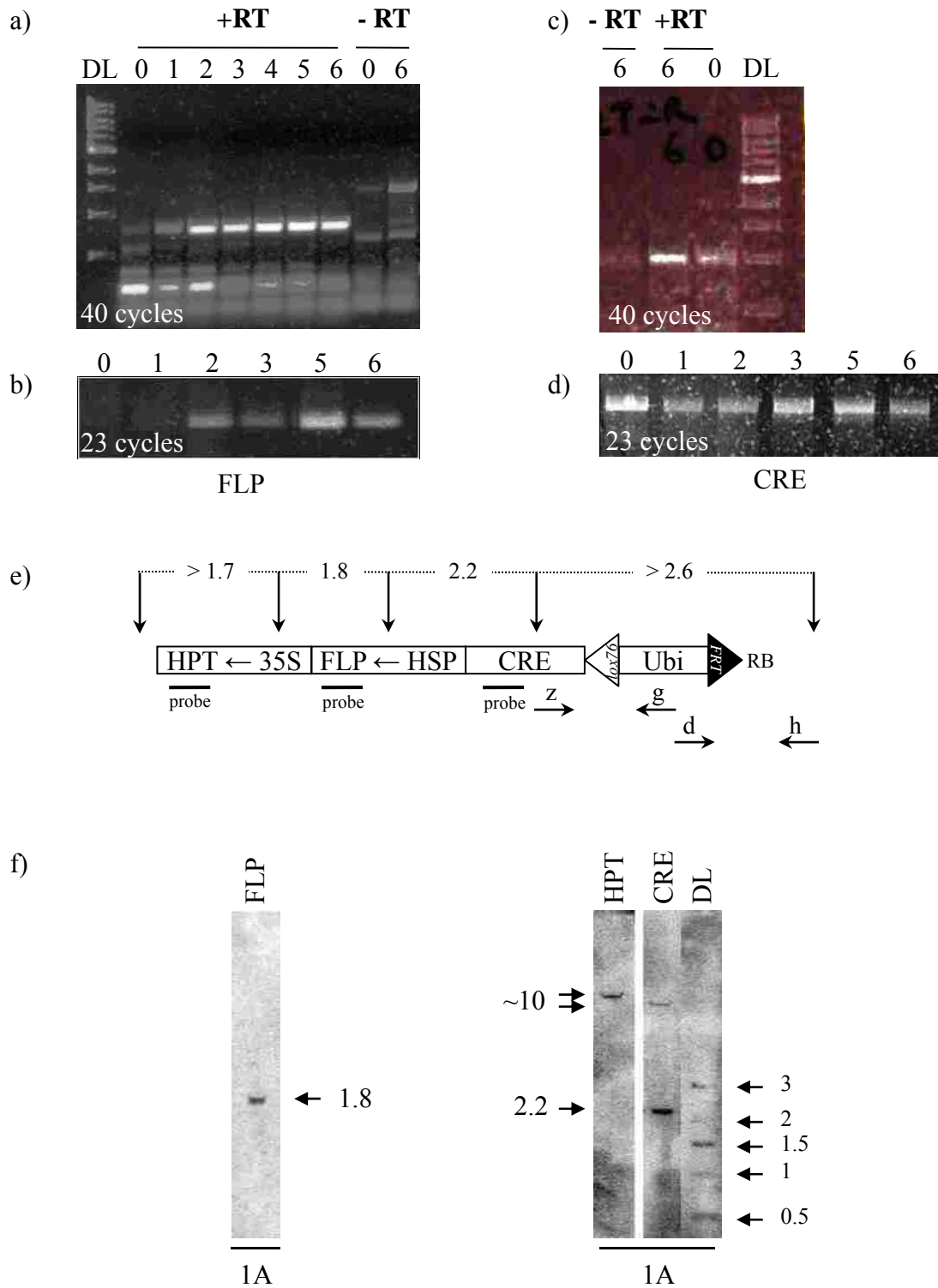
Fig 4. Analysis of target line1A: a – d) RT-PCR 40 cycles or 23 cycles using FLP or Cre primers on RNAs obtained from the heat (42°C) treated (0 – 6h) callus, e) *Eco*RV restriction map of 1A locus, f) Southern blot of *Eco*RV-digested genomic DNA hybridized with HPT, CRE and FLP probes. [DL: 1 Kb DNA Ladder]

Analysis of the Target Line: Expression Analysis of FLP and CRE Genes (RT-PCR)

High level expression of recombinase gene(s) is crucial for efficient recombination, which necessitates the expression analysis of the recombinase gene following the transformation. When compared to the ones driven by a constitutive promoter, transgenes driven by inducible promoters may not produce transcripts abundantly, which makes the expression analysis even more critical.

To test whether the heat shock promoter was functional in the target line, 1A callus was incubated at 42°C for 0 – 6h. FLP expression was detected by 23 cycles or 40 cycles of RT-PCR using total RNA derived from the treated callus (Fig. 4). While FLP expression was undetectable at RT (room temperature) and 1h treated callus in 23-cycle PCR, it was detected in 2 – 6h treated callus. Therefore, 2 hour heat treatment was sufficient to activate the heat shock promoter. The 40 cycle PCR detected low levels of FLP transcripts at RT and 1h treated callus, but a clear up-regulation was seen in 2-6 h treated callus. The band intensity was almost same for 2, 3, 4, 5, 6h time periods, which shows the extended heat exposure had no significant effect on the promoter activity. Expression of the CRE gene, driven by a constitutive promoter, Ubi, was detectable at all times. (Fig. 4).

Integration Lines

Transformation

The integration vector, pRP4 (Fig. 1), was bombarded onto the 1A callus.

Since the CRE gene was expressed in 1A cells, pRP4 was expected to integrate into the target *lox*76 via a two-step reaction: separation of vector construct from its backbone via *lox*P x *lox*75 recombination followed by integration of the circularized vector construct (donor circle) via *lox*75 x *lox*76 recombination (see the molecular strategy section). The bombarded calli were selected and proliferated on 2N6D media plates containing Geneticin (100 mg/l). A total of 7 Geneticin-resistant lines (P1 – 7) were obtained from three experiments, each consisting of 10 plates. The Geneticin resistant lines were regularly sub-cultured on 2N6D media containing Geneticin (100 mg/l). Regeneration of plants was not successful from any of the callus lines. Most of the subsequent analysis was done with P1 and P2 as they were the first two lines available.

Molecular Characterization of the Integration Lines

To confirm the Cre-mediated site-specific integration of pRP4 into the target locus, all Geneticin resistant lines were subjected to histo-chemical staining, PCR, sequencing and Southern analysis. Geneticin resistant lines were screened for GUS expression by histochemical staining method (Jefferson *et al.* 1987). Except for one line, all Geneticin-resistant lines stained blue in GUS stain, indicating the integration and expression of GUS gene (Appx. Fig. 4).

To verify the presence of the site-specific integration locus, polymerase chain reaction (PCR) across the two integration junctions (*cre-lox-FRT-gus* and *npt-lox-ubi*) was carried out using and c (CreUAG) – b (Nos1) and g (Ubi) – f (KanR) primer sets, respectively (Fig. 5).
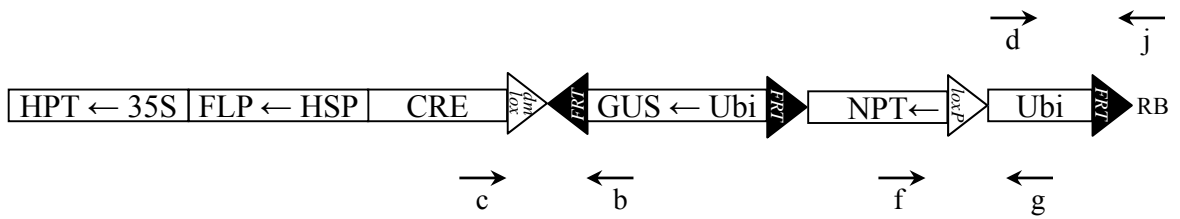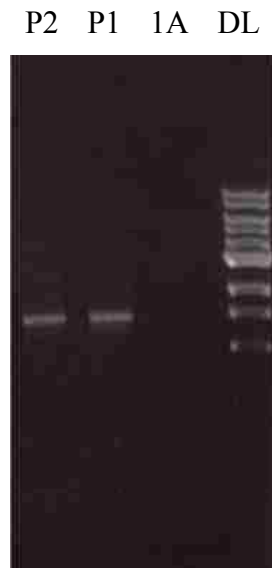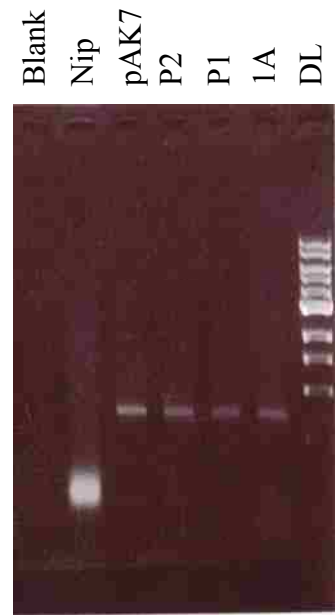
Fig 5. PCR analysis of SSI lines P1 and P2. a) Diagrammatic illustration of P locus; b – d) PCR on lines P1 and P2 using c – b, f – g and d – j primers. Nip (untransformed rice DNA) and 1A served as negative control, while pAK7 served as positive control. [DL: 1 Kb DNA Ladder]

PCR fragments of 1.2 kb and 1.4 kb were expected from precise *gus-cre* and *ubi-npt* junctions, respectively. Both P1 and P2 generated the expected junction bands (Fig. 5). Next, PCR was done to detect the presence of *FRT* sites. Using c (CreUAG) – b (Nos1), d (Rubi597) – e (kanF) and d (RUBI597) – j (RB) primer sets, the fragments including *FRT* sites in P1 and P2 loci were amplified and sequenced to confirm the integrity of the *FRT* sites (Fig. 5). PCR using a (LB) – b (Nos1) or a (LB) and k (HygF) primer sets to amplify the fragment containing the *FRT* site on the LB end failed. The reason behind this was most probably a truncation in the LB end past the HPT gene, as 1A cells are Hygromycin resistant. Additionally, the HPT gene could be amplified using k (Hyg F) and m (Hyg B) primers.

Subsequently, Southern analysis was conducted to determine the structure of the integration locus and, to find out, if any random integrations of pRP4 had occurred. For this purpose, *Eco*RI- and *Eco*RV-digested DNAs of the integration lines were used to prepare a Southern Blot, and hybridized with three different radioactive probes, GUS, NPT and CRE (Fig. 6 & 7).

Lines P1 and P2 were analyzed by digesting their DNA with *Eco*RI and *Eco*RV and hybridizing with CRE, GUS and NPT probes. The GUS hybridization of the *Eco*RI blot was expected to generate a single band at 4.1 kb showing full-length integration of *ubi-gus* fragment. The NPT and CRE hybridizations were expected to generate a 1.6 kb and 1.1 kb bands respectively, showing site-specific integration of pRP4 derived donor circle into the target locus (Fig. 6). Both P1 and P2 generated the expected bands in all hybridizations. However, an extra band at ~3.5 kb was seen in P1 with NPT hybridization, indicating the presence of a random integration of a truncated fragment.
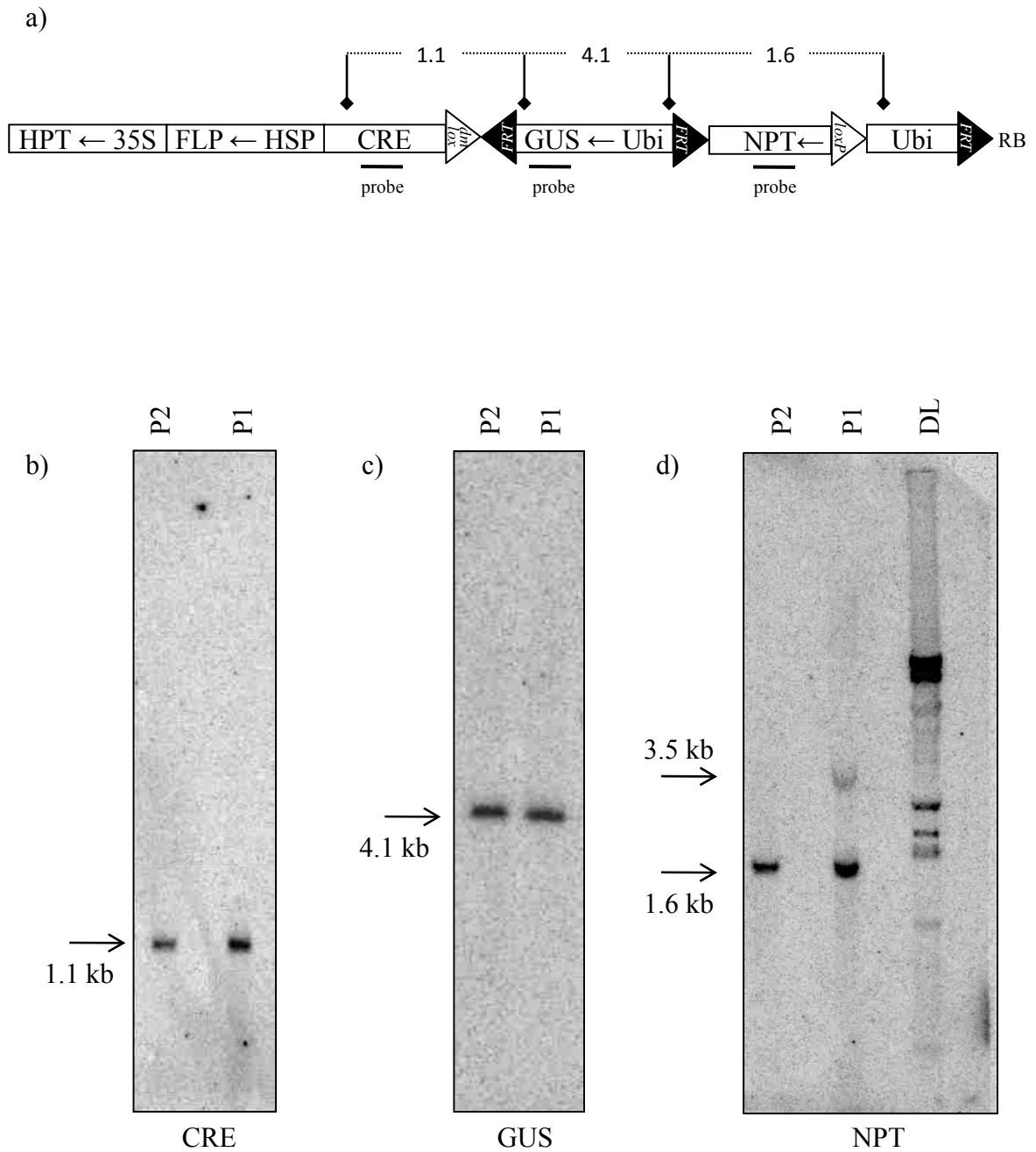
Fig 6. Southern analysis of P1 and P2 lines: a) *Eco*RI restriction map of P1 and P2 loci; b-d) Southern blot of *Eco*RI restricted genomic DNA hybridized with CRE, GUS and NPT probes [DL: 1 Kb DNA Ladder]
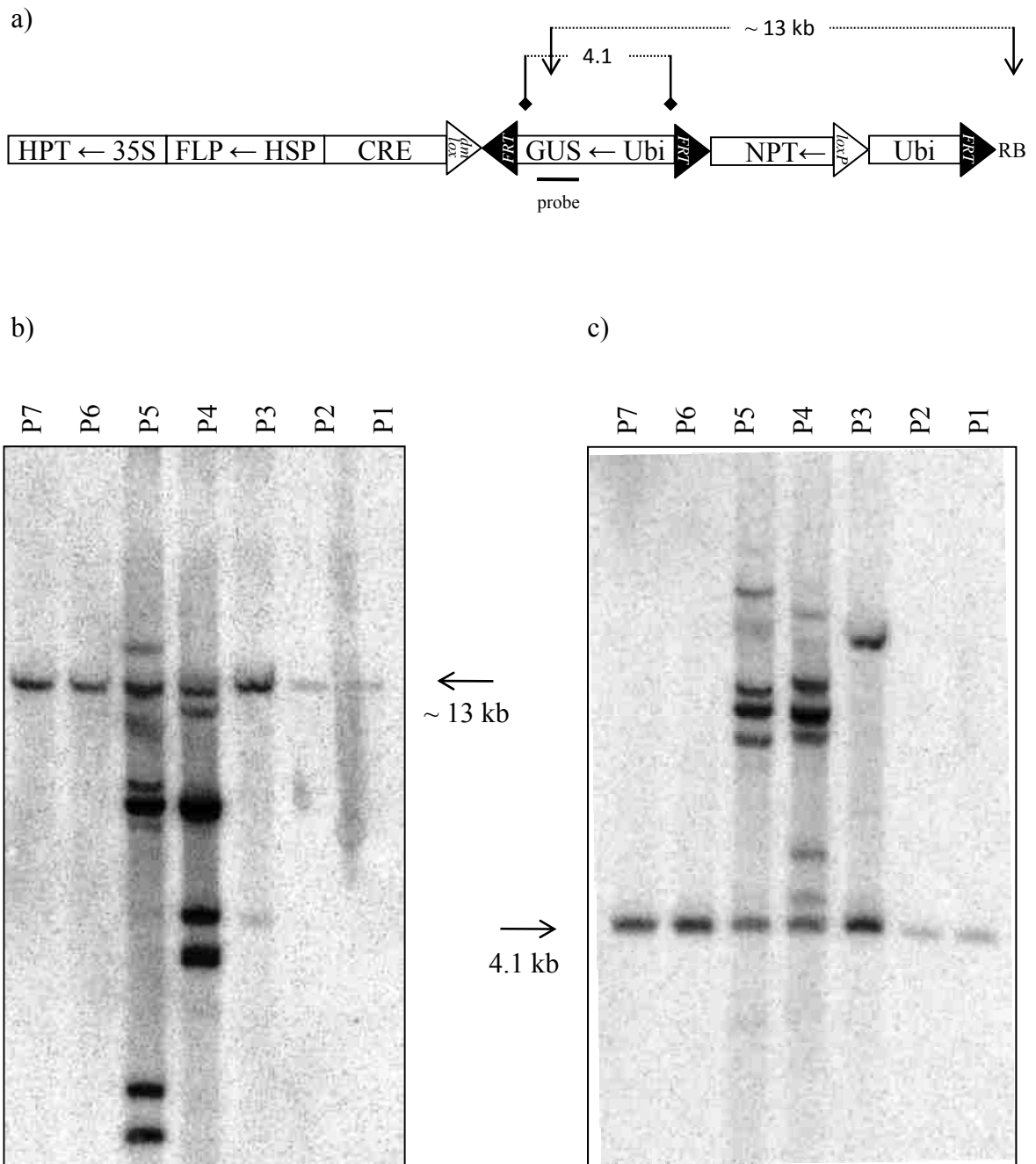
Fig 7. Southern blot of P lines hybridized with GUS probe. a) *Eco*RV (↓) and *Eco*RI (◆) restriction map of P locus; b – c) Southern blot of *Eco*RV (b) and EcoRI (c) digested genomic DNA of P lines with GUS Probe.

P2 line appeared to contain only SSI integration. On *Eco*RV blot, P1 and P2 generated a ~13 kb band indicative of the transgene-host DNA junction (Fig. 7). No other bands were seen in any of the above hybridizations on P1 and P2 genomic DNA, suggesting the presence of only the SSI locus and the absence of random integrations.

Two more Southern blots were done with *Eco*RI- and *Eco*RV- digested genomic DNA of lines P1 – P7. Hybridization of *Eco*RV blot with a GUS probe showed a band of ~13 kb in each line, indicative of transgene-host DNA junction, and the *Eco*RI blot displayed a 4.1 kb band in each line indicating the full-length integration of Ubi-GUS fragment. These hybridizations revealed that P1, P2, P6 and P7 were single copy lines, while P3, P4 and P5 contained additional random integrations (Fig. 7).

In conclusion, hybridizations confirmed that P1 and P2 contained the defined junctions (Fig. 6), and a single copy of the integration construct (pRP4) integrated into the 1A locus. (Fig. 7, Table 4).

Attempts for Inducing Marker Gene Excision

Heat Shock Experiments

Various heat exposure times (1 – 6h) at 42°C followed by different incubation periods (24 – 120 h) at 28°C were used to activate the heat inducible FLP gene to obtain excision of *FRT*-flanked DNA fragments in P1, P2, P6, and P7 cells. FLP expression in the treated callus was assayed by RT-PCR, and the genomic DNA, isolated from the treated callus, was subjected to PCR to detect the excision footprint: *gus-FRT-rb*.

Table 4. Molecular analysis of integrant lines

	PCR		Southern analysis (MC/SC)		
	ubi-npt	*gus-cre*	*GUS*	*NPT*	*CRE*
P1	+	+	SC	SC*	SC
P2	+	+	SC	SC	SC
P3	ND	+	MC	ND	ND
P4	ND	+	MC	ND	ND
P5	ND	+	MC	ND	ND
P6	ND	+	SC	ND	ND
P7	ND	+	SC	ND	ND

ND: Not Determined; MC: Multi-copy; SC: Single copy; *: Also contains a truncated copy of NPT gene.
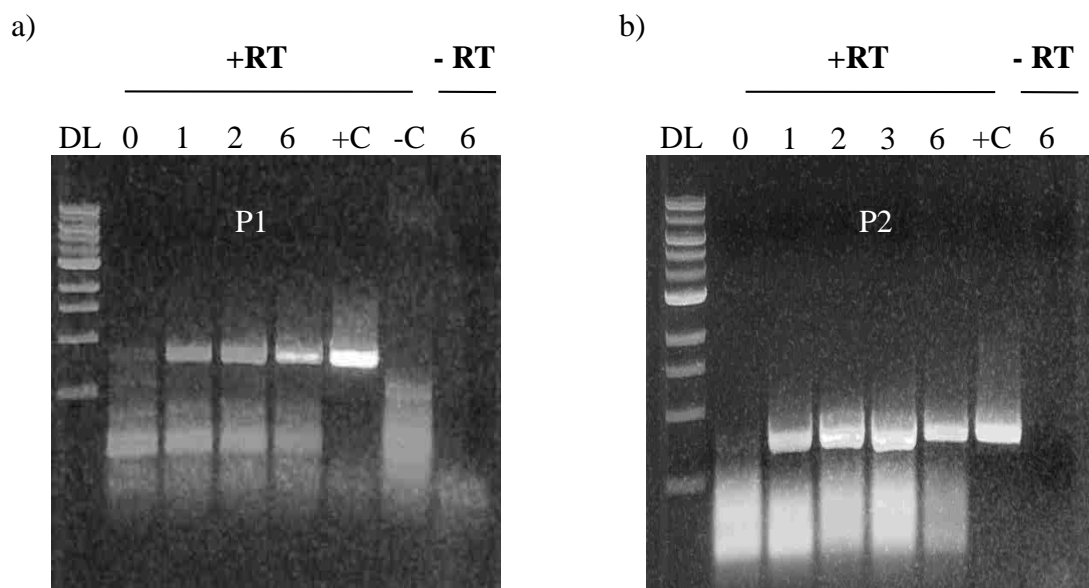
Fig 8. FLP expression analysis of P1 and P2 calli a) FLP RT-PCR (40 cycles) on RNAs obtained from the heat (42°C) treated (0, 1, 2 and 6h) P1 callus b) FLP RT-PCR (40 cycles) on RNAs obtained from the heat (42°C) treated (0, 1, 2 – 6h) P2 callus. – C: negative control, Nipponbare RNA;+C: positive control, pAK7 DNA.

Although RT-PCR clearly showed induction of FLP expression in all treated samples (Fig. 8), the excision footprint (0.4 kb: *gus-FRT-rb* fragment) was not detected from most samples except for a single sample that had undergone 6h heat treatment every 24h for 5 days (Fig. 9).

Sequencing of this band confirmed the presence of the 'excision footprint' in this sample (Fig. 9). However, this result was neither reproducible nor verifiable by Southern analysis. This finding suggests that excision occurred rarely in a very few cells, while the SSI locus stayed intact in the rest of the cells. The reason behind this excision difference might be that FLP protein is not as efficient as CRE protein in recombination reactions, which is shown in previous research (Buchholz *et al.*, 1996; Ringrose *et al.,* 1998), or FLP transcript level could not reach the threshold in the majority of the cells to fulfill the excision due to the weak of HSP promoter activity.
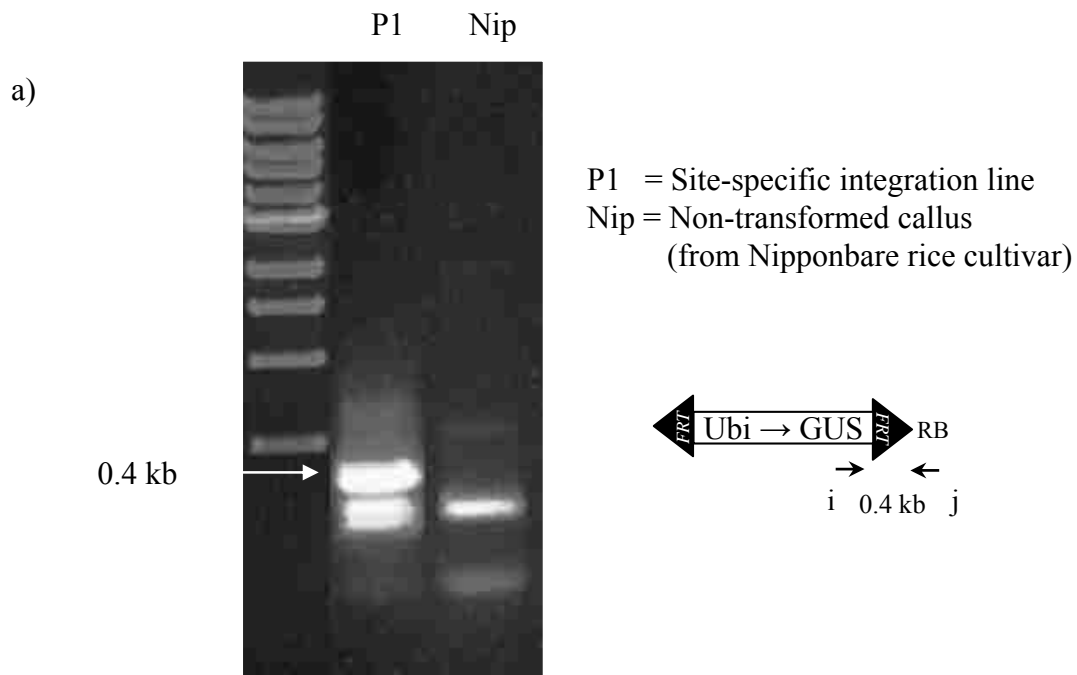
Introduction of Strong FLP Activity into P1 and P2 Cells

Based on the results discussed above, it was concluded that the heat shock promoter did not produce enough FLP transcripts, and hence, not enough FLP protein, to initiate excision of the *FRT* flanked DNA. Therefore, P1 cells were re-transformed with a Ubi-FLP gene in which a strong promoter derived from maize ubiquitin-1 gene drives FLP expression. For this purpose, pUbi-FLP was co-bombarded with pUbi-Bar that served as a selection vector. Integration of the Ubi-FLP gene was detected by PCR analysis in 4 out of 7 lines (P1F lines) obtained in 2 different experiments (Fig. 10). To differentiate Hsp-FLP from Ubi-FLP, PCR was done with a1 (Ubi1812) and b1 (FLPRII) primers.

85

However, expression analysis of the UbiFLP gene in P1F lines gave disappointing results: RT-PCR showed that FLP expression in these lines was only slightly higher than the untreated P1 line. Comparison of this data with the treated P1 and P2 data, suggested that UbiFLP expression was even lower than the expression of heat-treated P1 and P2 calli. Consequently, PCR and Southern analysis revealed that excision of FRT-flanked DNA did not occur in P1F lines (Data not shown). As Ubi-FLP integration occurred in complex patterns in all of the lines analyzed by Southern analysis, poor FLP expression is most likely related to the gene silencing associated with complex integrations.

New Generation FLP Genes: FLPo and FLPe

Wild type FLP protein, which has low thermo-stability, was found to be much less efficient in recombination assays as compared to CRE protein (Raymond and Soriano 2007; Buchholz *et al.* 1996). To improve the recombinase property of FLP gene, Buchholz *et al.* (1998) randomly mutated the coding sequence of FLP, and screened them in *Escherichia coli* for improved recombinases. After successive rounds of screening and DNA shuffling they collected the best clone from the eight generation, and called this thermostable FLP gene as FLPe (enhanced FLP) gene. FLPe was found to have a 4-fold more efficiency at 37°C and 10-fold at 40°C than the wild-type FLP protein in recombination assays. The recombination efficiency of FLPe was similar to that of Cre in these two experiments. On the other hand, differences in codon usage between prokaryotic and eukaryotic cells or the presence of cryptic splicing sites in foreign genes can cause poor gene expression (Raymond and Soriano, 2007).

a)

P1   Nip

P1 = Site-specific integration line
Nip = Non-transformed callus
         (from Nipponbare rice cultivar)

0.4 kb

i   0.4 kb   j

b)

TGTAATAATTAACATGTAATGCATGACGTTATTTAGAGATGGGTTTTTATGATTA
GAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAA
ACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAATTCA
AGCTTGATGGGGTACC**GAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC**CT
CGAGTTGATGAAAGAATACGTTATTCTTTCATCAAAAGCTTANCTTGNGCTTGG
ATCANATTGTCGTTTA

Fig 9. PCR analysis of heat-treated P1 line. (a) Genomic DNA isolated from the heat-treated P1 callus (6h 42°C every 24h for a period of 120h) was subjected to PCR with i – j primers; (b) The amplified band (0.4 kb) was sequenced with i and j primers. The presence of GUS, nos3′ downstream of GUS (blue)-FRT (black)-RB (red) sequences found, part of which is displayed here.

87

Use of FLP and FLPe genes in plants is subject to this phenomenon, since they originate from a lower eukaryote, *Saccharomyces cerevisiae*. To deal with this problem, Raymond and Soriano (2007) engineered FLPe gene *de novo* with mouse codon usage, and obtained FLPo gene (FLP optimized), which they found to be more efficient than FLPe gene in mammalian cells. Therefore, we tested recombination efficiencies of FLPe and FLPo genes driven by ubi promoter in rice cells. This work was done to find out which version of FLP gene would work best in rice.

Comparison of Transiently Expressed FLP, FLPe and FLPo Genes

To compare the recombination efficiencies of transiently expressed FLP, FLPe and FLPo genes, pUbi-FLP (FLP), pAA7 (Ubi-FLPo) and pAA8 (Ubi-FLPe) (Fig. 2) were bombarded on the young leaves and immature embryos obtained from *FRT* target line, 1.7D. Line 1.7D contains an *FRT* flanked NPT-nos3′ fragment, which blocks the expression of GUS gene. Expression of FLP, FLPe and FLPo genes were expected to delete the *npt-nos* fragment, and activate GUS gene by fusing it with Ubi promoter (Fig. 11). pHPT (hygromycin phosphotransferase) served as negative control (Fig. 2). Four replicates of nine immature embryos and three leaves (1 cm long) were bombarded, for each construct. After 60 h incubation at 37°C, explants were immersed in GUS stain overnight. Upon GUS staining it was clear that FLPe- and FLPo-bombarded leaves and embryos had more blue dots as compared to the ones bombarded with FLP. Four representative embryos and one leaf for each construct are shown in Fig. 11. Blue dots on each leaves and embryos were counted under microscope and subjected to statistical analysis.
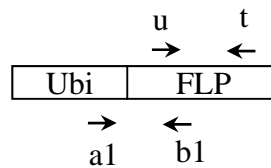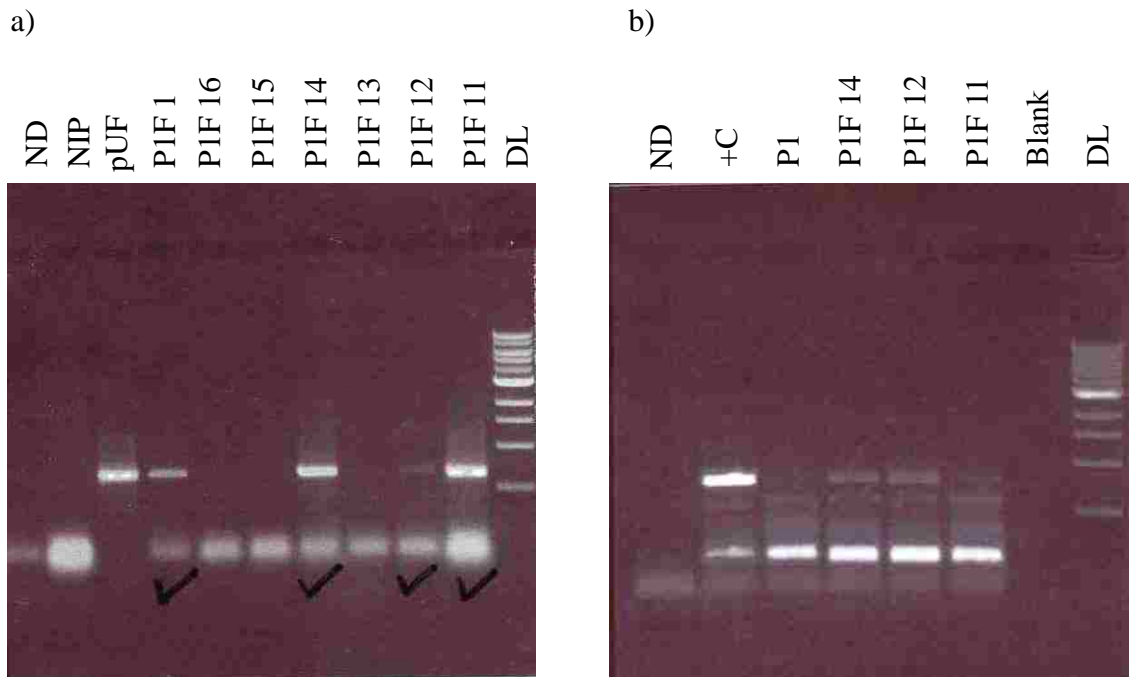
Fig 10. PCR and RT-PCR analysis of P1F lines. a) PCR analysis on genomic DNA using a1 (UBI1812) and b1 (FLPRII) primers. b) RT-PCR analysis on total RNA using FLP primers (u and t). + C: Positive control (P1 DNA); ND: No DNA; Nip: Untransformed rice DNA; pUF: Ubi-FLP plasmid [DL: 1 Kb DNA Ladder]
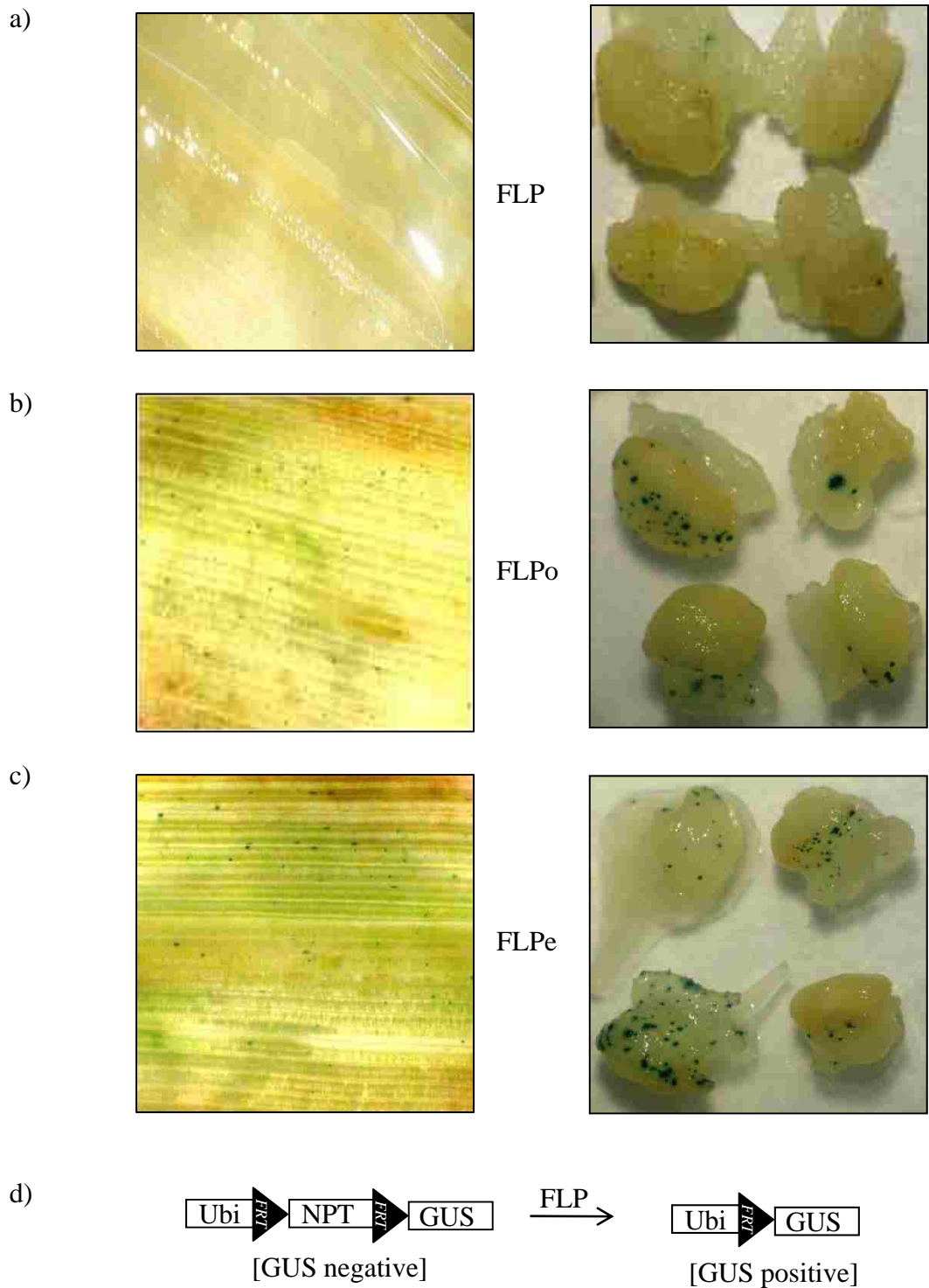
a)

FLP

b)

FLPo

c)

FLPe

d)

Fig 11. Recombination efficiency of different versions of FLP protein. Bombardment of a) pUbi-FLP (FLP), b) pAA7 (FLPo) and c) pAA8 (FLPe) on 1.7D leaf and immature embryos. 1.7D contains a single-copy integration of pRP9 (see Fig. 1b). Representative leaves and embryos are presented in the figure d) Recombination assay.

However, the embryos that had no blue dots were not included in the statistical analysis, because 1.7D plants were hemizygous for *FRT* locus, and therefore produced segregating embryos. Data were analyzed using ANOVA (<u>An</u>alysis <u>of</u> <u>var</u>iance) and the differences contrasted using Duncan's multiple range test at 5%. The mean of the blue dots on each leaf were 4.82, 14.30 and 18.73 for FLP, FLPo and FLPe respectively. Similarly, the mean of the blue dots on each embryo were 5.06, 14.94 and 15.61 for FLP, FLPo and FLPe, respectively. Both in the leaves and the embryos, the efficiency of transiently expressed FLPe and FLPo gene were significantly higher than that of FLP in catalyzing site-specific recombination on the chromosomal target in 1.7D (Table 5).

Excision via Re-transformation with FLPe

Based on the above data, Ubi-FLPe was chosen for driving excision in P1 and P2 loci. To introduce the FLPe gene into P1 and P2 cells, pAA8 (FLPe) and pUbi-Bar were co-bombarded onto P1 and P2 calli. pUbiBar was used for selection purpose as *bar* gene confers resistance to bialaphos. Thirty-two transformants (E lines) were selected on media containing biolaphos (5mg/L). All of them were screened by PCR for the presence of the FLPe gene, but eleven of the E transformants were PCR positive for FLPe (Fig. 12c).

Molecular analysis of E lines: PCR & Southern hybridization

These E lines were then subjected to PCR with y (GUSF3380) and j (RB) primers to detect the excision footprint (Fig. 12). The amplification of a 0.8 kb band in 8 of the 11

lines indicated the occurrence of *FRT* x *FRT* recombination, presumably excision of *FRT* flanked DNA, generating a *gus-FRT-rb* junction. A representative PCR result is shown in Fig. 12b.

Since P1 and P2 locus contained 3 *FRT* sites, multiple recombination products were possible (shown in Fig. 13). However, these recombinations could only change the structure between *dmlox* and RB, generating 4 possible structures designated as B − E. The presence of structure B cannot be easily determined by PCR as both contain Ubi-*FRT*-RB junctions. Presence of C and D can easily be determined by PCR because a new junction GUS-*FRT*-RB is formed. As described above, most E lines contained this structure. To analyze further, *Eco*RV Southern blot of E lines was hybridized with different DNA probes. This analysis indicated the occurrence of (1) excision and inversion, (2) excision without inversion, (3) inversion without excision. The detailed analysis is given below:

Hybridization of *Eco*RV-cut genomic DNA of P1 and P2 lines with 5′-GUS probe generated a ~13 kb fragment (Fig. 14). E lines on the same blot showed one of the following combinations: (1) presence of 3.3 kb and 10 kb bands (E2, E4, E15, E21), indicating occurrence of both inversion and excision (structure C); (2) 3.3 kb and 11.7 kb bands (E13, E14), indicating occurrence of inversion without excision (structure E); (3) presence of only 10 kb (E12), indicating excision without inversion (structure B); presence of only 13 kb band (E6, E10, E11), indicating lack of recombination (structure A: P1 or P2).

Table 5. Efficiency test of FLP genes in *FLP-FRT* mediated excision

Gene	N (Sample Size) (Exp.1 / Exp. 2)[*]	Experiment 1 (Leaf)[**]	Experiment 2 (Immature Embryo)[**]
FLP	3/9	4.82[a]	5.06[a]
FLPe	3/9	18.73[b]	15.61[b]
FLPo	3/9	14.30[b]	14.94[b]
Hygromycin (Control)	3/9	0.09[a]	0.00[a]

Values with identical letters in the same column are not significantly different ($P<0.05$) as compared by Duncan test.
*: Experiments were conducted with four replications.
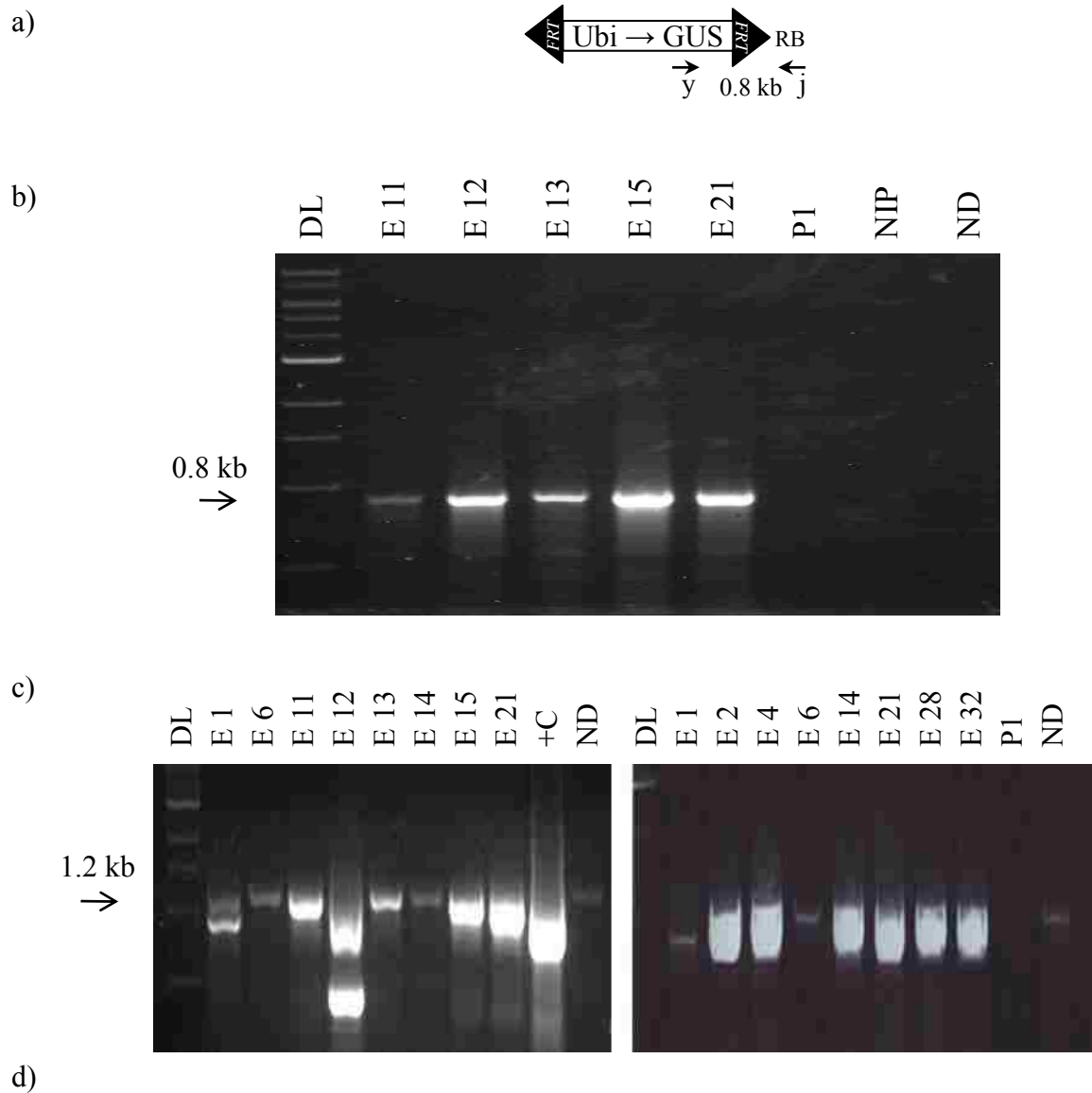**: The means of the blue dots on each explant

a)

b)

0.8 kb →

c)

1.2 kb →

d)

TGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTA
GAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCG
CAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAAT
TCAAGCTTGATGGGGTACC**GAAGTTCCTATTCTCTAGAAAGTATAGGAACTT
C**CTCGAGTTGATGAAAGAATACGTTATTCTTTCATCAAAAGCTTAGCTTGAGCT
TGGATCAGATTGTCGTTTA

Fig 12. Molecular evidences for FLP-mediated excision in E lines: a) Schematic diagram of the excision locus, b) PCR on selected E lines using y (GUS F3380) – j (RB) primers, c) PCR on selected E lines using v (FLPe F) – w (FLPe R) primers, d) The sequence of the amplified band from 5 E lines with y (GUSF3380) primer, showed presence of excision footprint *GUS-**FRT**-RB*. [DL: 1 Kb DNA Ladder; NIP: Untransformed rice (var. Nipponbare) DNA; ND: No DNA (– Control)]
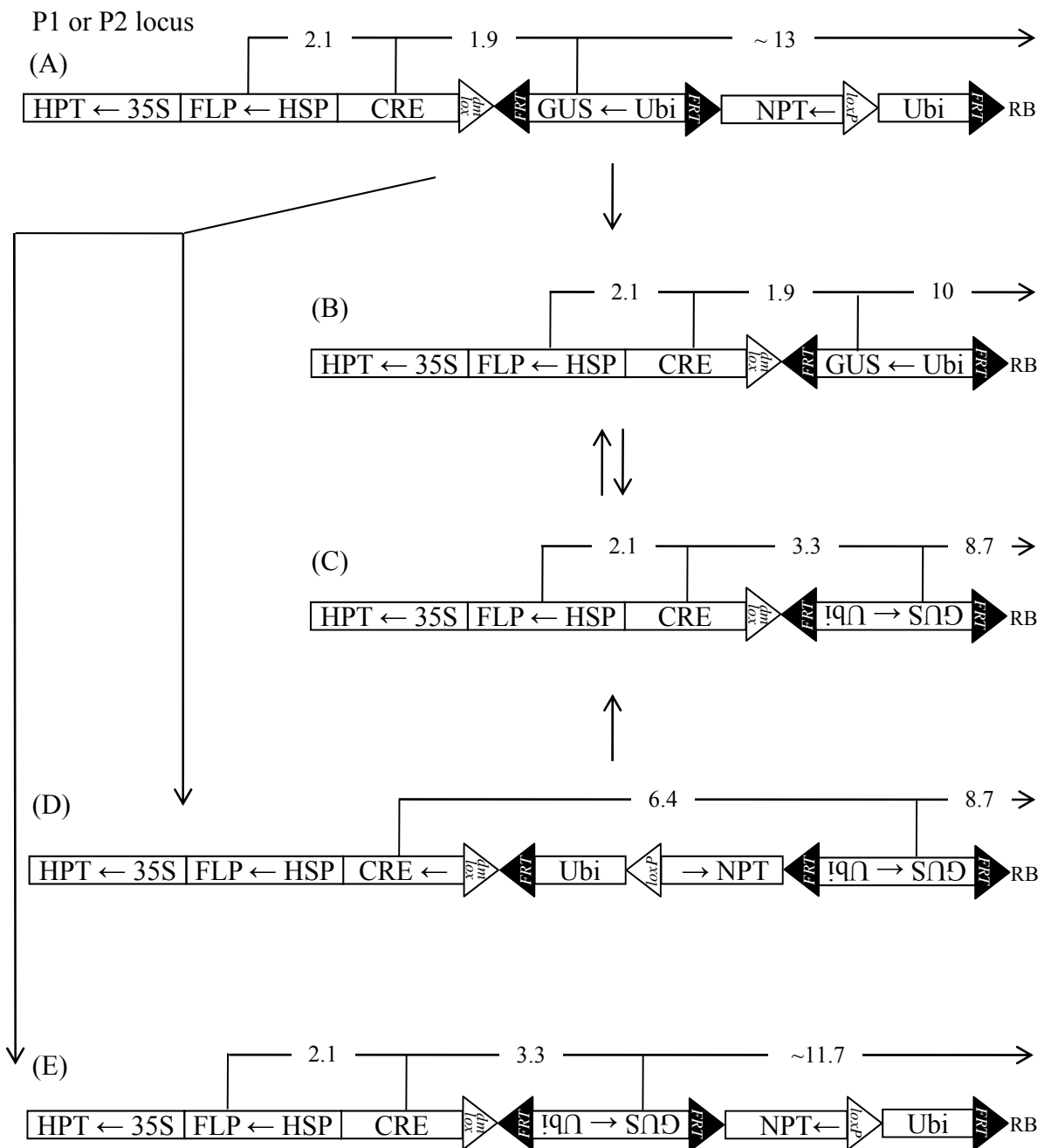
Fig 13. Possible locus structures which may occur as a result of FLPe-mediated site-specific recombinatons A) P1 or P2 locus (No recombination), B) Excision of NPT fragment, C) Excision of the NPT fragment and inversion in the GUS gene cassette, D) Inversion in the fragment flanked by the farthest FRT sites, E) Inversion in the GUS cassette without excision of NPT fragment. *Eco*RV restriction sites are marked on the locus structures and intervening fragment sizes are given in kb.
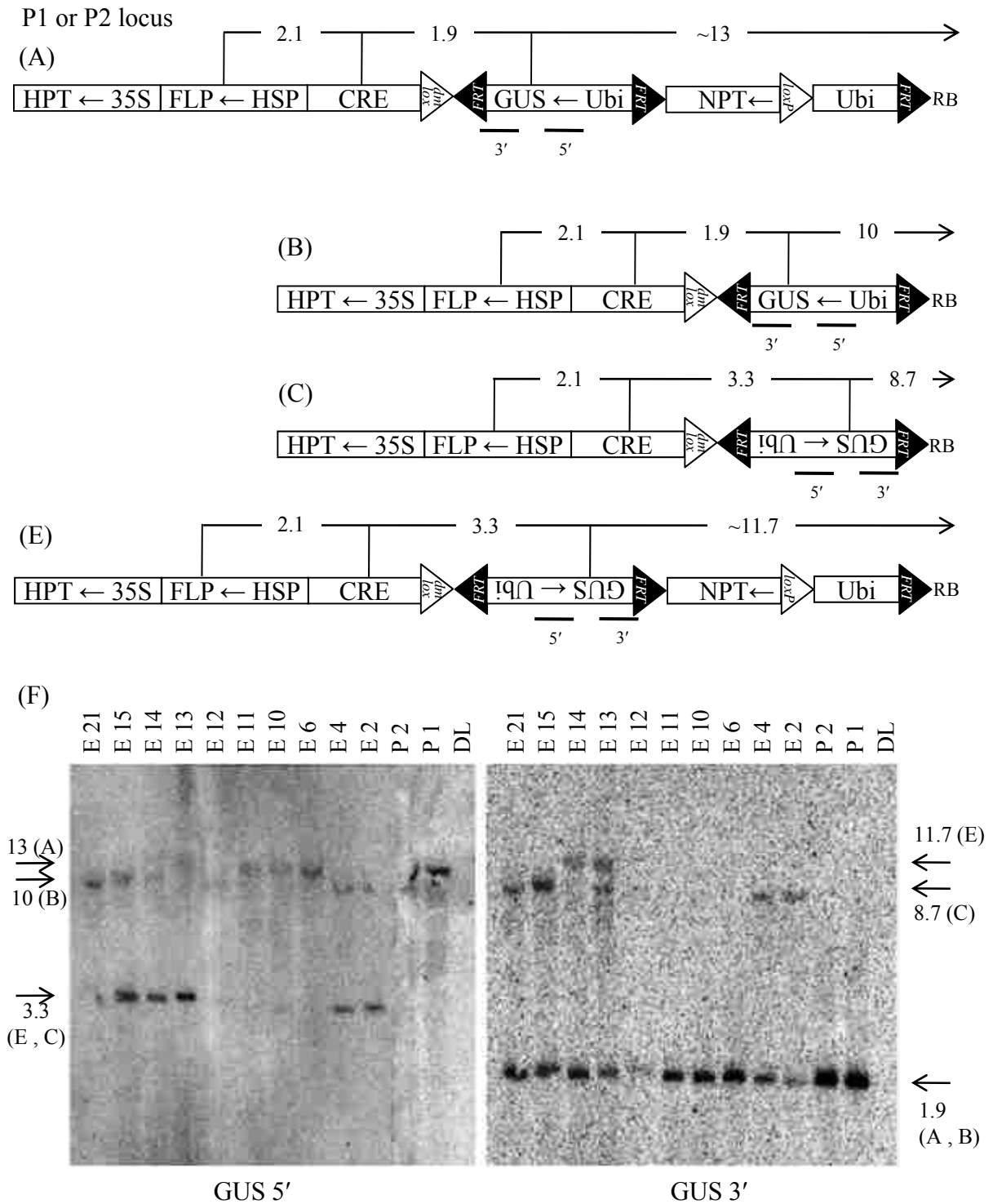
Fig 14. Locus structures generated as a result of FLPe-mediated site-specific recombinatons, A) P1 and P2 locus (No recombination) B) Excision in NPT fragment, C) Excision in the NPT fragment and inversion in the GUS gene cassette, E) Inversion in the GUS cassette. *Eco*RV restriction sites are marked on each locus structure. F) Southern blot of *Eco*RV-digested genomic DNA of selected E lines hybridized with GUS 5′ and 3′ probes. DNA fragment lengths are given in kb. Association of each band with locus structure (A – E) is also shown.

Hybridization of this blot with 3′-GUS probe generated a 1.9 kb band on P1 and P2 genomic DNA, indicating single orientation of Ubi-GUS gene (as depicted in structure A) in P1 and P2. All E lines also showed this 1.9 kb band, indicating the presence of structures A and/or B. As E6, E10, E11 did not contain any additional band; they must contain only structure A (i.e., P1 or P2). Presence of an 11.7 kb band, which is indicative of inversion w/o excision (structure E), was seen in E13 and E14; however E13 also contained 8.7 kb band that indicates excision (structure C). Presence of a clear 8.7 kb band was seen in four other lines (E2, E4, E15, E21), indicating excision with inversion (structure C).

Next, the *Eco*RV blot was hybridized with the NPT probe. Both P1 and P2 showed ~13 kb band (Fig. 15a). An extra band of 3.5 kb was seen in P1 but not P2. This band represents the extra truncated NPT fragment in P1 genome (discussed earlier; see Fig. 6). Thus, 3.5 kb band is also seen in all P1-derived E lines (E2, E4, E6, E10, E11, E12, E21). This band is not expected to undergo recombination. Absence of 13 kb NPT band was observed in E2, E4, E15 and E21, suggesting the excision of *FRT*-flanked Ubi-NPT fragment. Occurrence of inversion (structures C or E) was detected by hybridizing this blot with Cre probe. Cre-hybridization on P1 and P2 genomic DNA generated 1.9 and 2.1 kb bands, indicating structure A. E lines either generated a pattern identical to P1 and P2 or a new combination of 3.3 kb and 2.1, indicating the presence of structure E, i.e. inversion.

In summary, excision footprints (10 kb and 8.7 kb GUS bands) were clearly seen in five lines: E2, E4, E13, E15 and E21. Four of these lines (E2, E4, E15 and E21) did not hybridize with NPT gene located in SSI locus (13 kb band), suggesting high rate of

excision in these lines. E13 on the other hand displayed the presence of NPT gene, suggesting low rate of excision. Of the four efficiently excising lines, three (E2, E4, E15) also underwent inversion, while the fourth line (E21) did not. The above analysis also indicated that structure D (i.e. recombination between farthest *FRT* sites resulting in Inversion) did not occur in any of the E lines. The predominant structure seen in the E lines is structure C (Table 6).

Finally, using 50 ng RNA-free DNA, RT-qPCR was done to calculate percentage excision on two selected lines, E12 and E15. P1 served as a control. Dilution series (0.05, 0.5, 5 and 50 ng) of P1 was used in RT-qPCR to generate standard curves (Appx. Fig.5). Using the primers sets in Table 3, DNA fragments belonging to CRE, GUS and NPT genes were amplified, and Ct values were obtained (Fig. 16). Results showed that GUS and CRE genes were present at 100% in all lines, while NPT gene was completely excised from E12. The excision efficiency was 75 % for E15 (Fig. 17).

Table 6. Summary of E lines in the *FLP-FRT* mediated recombination types occurred

|	Locus of origin	(A) No Recombination	(B) Excision	(C) Excision + Inversion	(E) Inversion
E2	P1			X	
E4	P1			X	
E6	P1	X			
E10	P1	X			
E11	P1	X			
E12	P1	x		X	
E13	P2			X	X
E14	P2				X
E15	P2			X	
E21	P1			X	

x: Occurred, X: Predominantly occurred
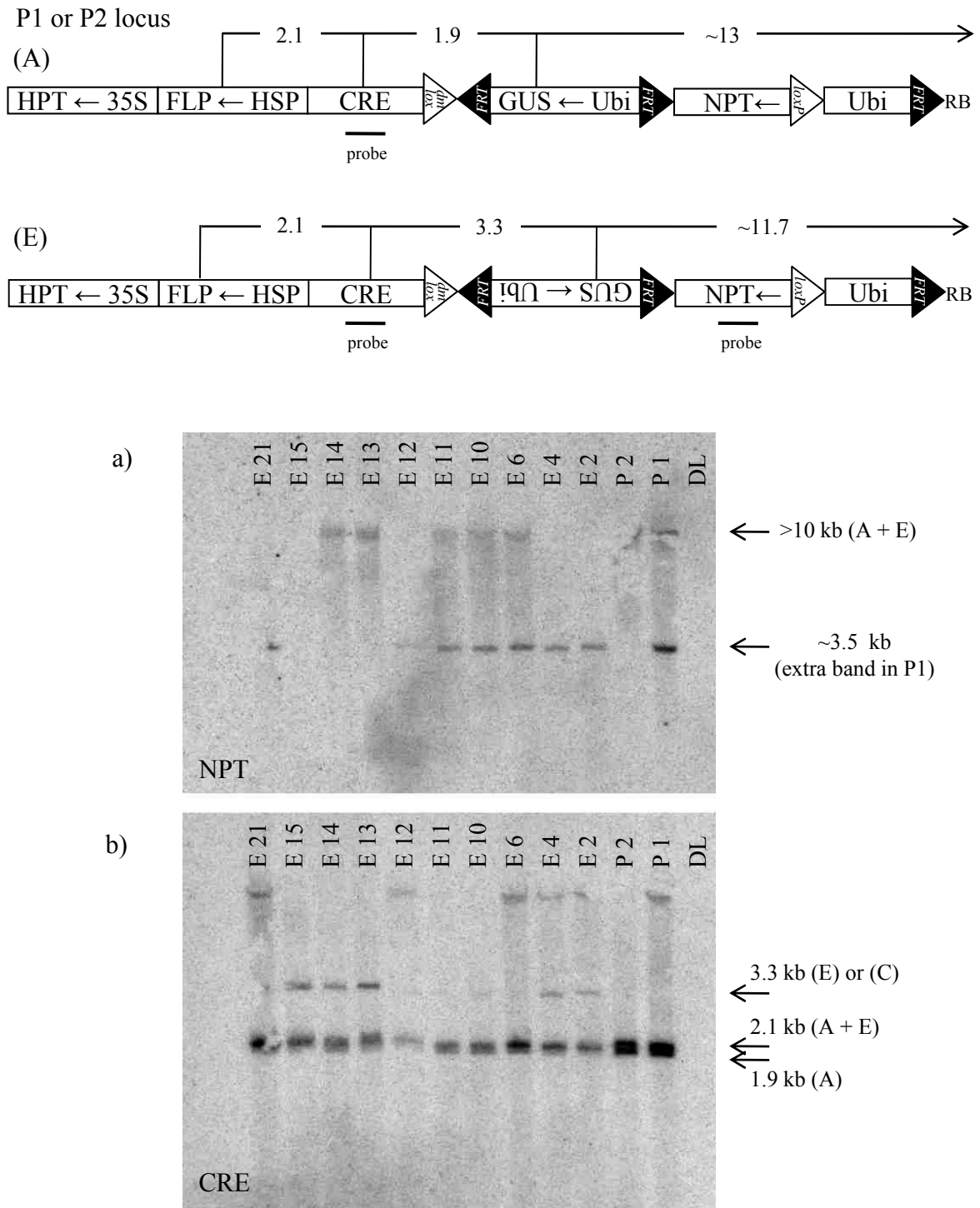
Fig 15. Southern blot of selected E lines showing the pattern of NPT and CRE genes. *Eco*RV restriction map of detectable locus structures, (A) and (E), with NPT and CRE hybridizations. CRE and NPT probes are shown as thick line a) Southern blot of *Eco*RV-cleaved genomic DNA hybridized with NPT probe, b) Southern blot of *Eco*RV-cleaved genomic DNA hybridized with CRE probe.
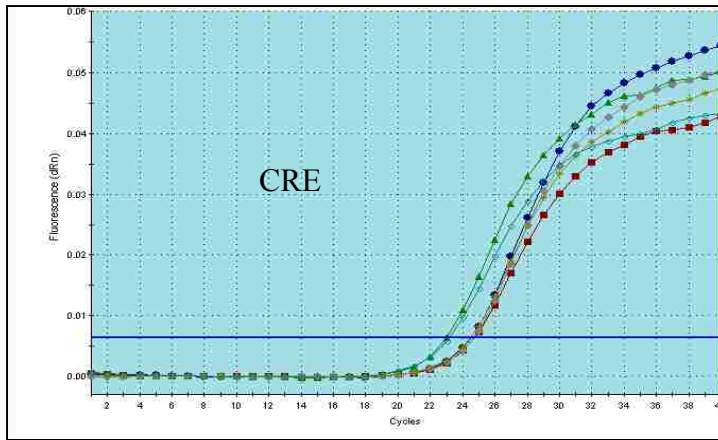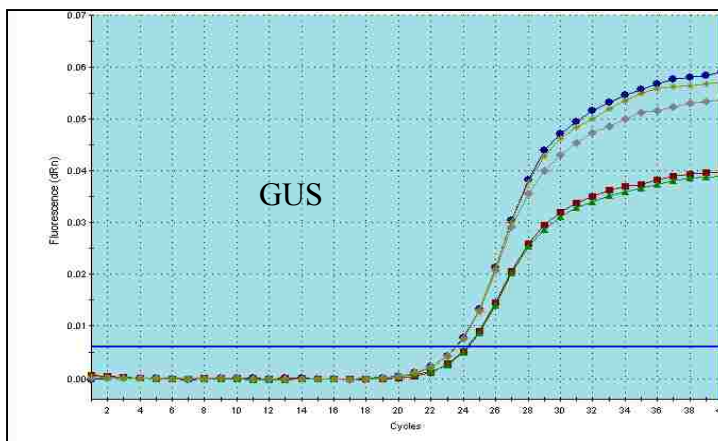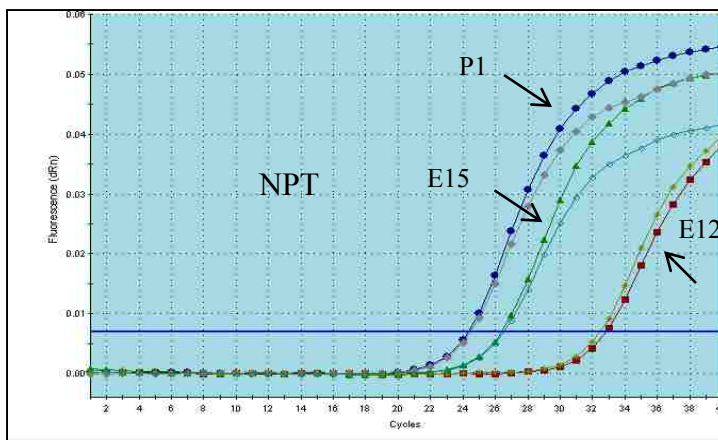
Fig 16. Results of RT-qPCR analysis on 50 ng DNAs of P1, E12, E15. Each line was replicated twice as shown in different colors. a) Using q1 (CRE F2128) – q2 (CRE R2388 b) Using q3 (GUS F2130) – q4 (GUS R2534) c) Using q5 (NPT F265) – q6 gene (NPT R418)
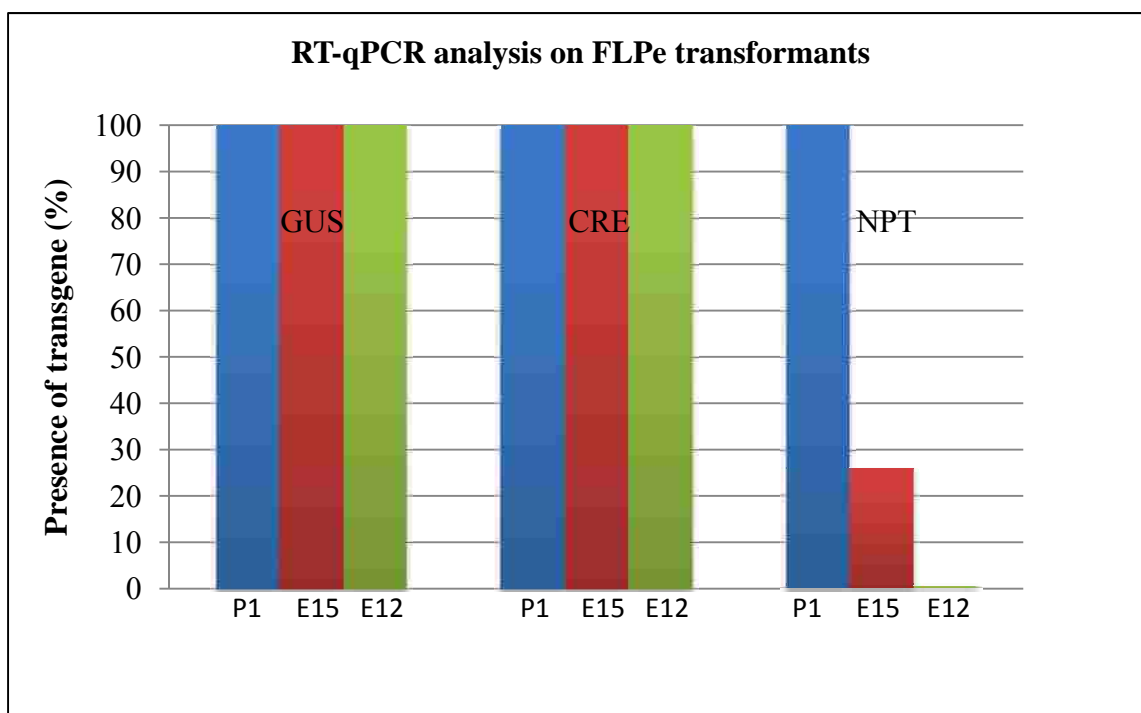
Fig 17. Results of RT-qPCR analysis for determining excision rates on E12, E15 and P1. *GUS*: β- glucuronidase gene; *CRE:* Cre gene; *NPT:* neomycin phosphotransferase II gene

Conclusion

To develop a "clean" transgene locus in the rice genome, combination of two site-specific recombination systems, FLP–*FRT* and Cre–*lox* was validated in this study. Cre–*lox* was used for site-specific integration of a transgene, GUS, and FLP–*FRT* system was utilized to remove all undesired elements, such as antibiotic resistance genes and recombinase genes, in the site- specific integration locus following integration. The soybean heat-shock promoter (HSP17.5E) was used for heat-inducible FLP activity. The efficiency of the system was validated in stably transformed callus cultures.

Because the efficiency of the heat-inducible FLP gene was found low in deletion of *FRT* flanked DNA fragments in the SSI locus, cells of the integration line (P1) were re-transformed with FLP gene driven by a constitutive promoter, Ubi, to increase the FLP protein amount in the cells. FLP expression in the re-transformed cells, however, was found even less than that of in P1, which may have resulted from HDGS (Day *et al.*, 2000).

Finally, P1 cells were re-transformed by an efficient version of the FLP gene (FLPe), resulting in the generation of E-lines.

In conclusion;

1- Soybean heat-shock promoter, HSP17.5E, is suitable for inducing FLP activity. While FLP expression was undetectable at RT and 1h treated callus, FLP expression was detected at 2 - 6 h treated callus. On the other hand, expression of CRE gene, driven by a constitutive promoter, was detectable at all times.

2-   PCR revealed that FLP-mediated recombination reaction resulted in deletion of *FRT*-flanked DNA; however, this result was not confirmed by Southern Analysis, showing the deletion occurred in a limited number of the cells.

3-   Cre-mediated site-specific integration was successful in precise, single copy transgene integration into a target line.

4-   Multi-copy integrations of FLP gene reversely correlated with FLP expression in the rice genome.

5-   FLPo and FLPe genes were more efficient than wild-type FLP in deletion of *FRT*-flanked DNA fragments.

6-   Expression of FLPe gene resulted in deletion of a 5-kb *FRT* flanked DNA fragment in the rice genome with a high efficiency.

7-   Replacing FLP gene with FLPe gene would result in efficient marker gene deletion using the present approach.

OVERALL CONCLUSION

The utility of site-specific recombination systems for expressing transcription units from an integration locus and removing marker genes from the rice genome was assessed in this study.

In the first part of the study, it was investigated if a positive correlation of gene dosage with expression level could be obtained from a locus containing 1 – 3 full-length copies in a tandem manner. To this end, Cre–*lox* system was used to integrate 1–3 copies of GUS or GFP gene cassettes into the rice cell. In conclusion, this study demonstrated that:

1- Precise integrations of tandem copies can be obtained by site-specific integration approach

2- Direct repeat structures remain stable in rice cells

3- Integrated gene copies confer additive effect on gene expression

The findings of this study could be used in development of technologies which target (a) boosting gene expression, (b) expression of different genes from the same locus. The latter would be especially useful in pathway engineering, which requires expression of each gene in a biochemical pathway in a predicted level.

The extension of this study could be:

1- Further increase in transgene copy number

2- Use of different genes

In the second part of the study, to develop a "clean" transgene locus in the rice genome, combination of two site-specific recombination systems, FLP–*FRT* and Cre–*lox* was validated. Cre–*lox* was used for site-specific integration of a transgene, GUS, and FLP–*FRT* system was utilized to remove all undesired elements, such as antibiotic resistance genes and recombinase genes, in the site- specific integration locus following integration. The soybean heat-shock promoter (HSP17.5E) was used for heat-inducible FLP activity. The efficiency of the system was validated in stably transformed callus cultures.

Because the efficiency of the heat-inducible FLP gene was found low in deletion of *FRT* flanked DNA fragments in the SSI locus, cells of the integration line (P1) were re-transformed with FLP gene driven by a constitutive promoter, Ubi, to increase the FLP protein amount in the cells. FLP expression in the re-transformed cells, however, was found even less than that of in P1.

Finally, P1 cells were re-transformed by an efficient version of the FLP gene (FLPe), resulting in the generation of E-lines. The conclusion highlights of this study are that;

1- Cre-mediated site-specific integration was successful in precise, single copy transgene integration into a target line.

2- FLPo and FLPe genes were more efficient than wild-type FLP in deletion of *FRT*-flanked DNA fragments.

3- Expression of FLPe gene resulted in deletion of a 5-kb *FRT* flanked DNA fragment in the rice genome with a high efficiency.

4-   Replacing FLP gene with FLPe gene would result in efficient marker gene deletion using the present approach.

Conclusions of both parts of this dissertation were presented in detail at the end of Chapters II and III.

REFERENCES

Ahmad R, Kim YH, Kim MD, Phung MN, Chung WI, Lee HS, Kwak SS and Kwon SY (2008) Development of Selection Marker-free Transgenic Potato Plants with Enhanced Tolerance to Oxidative Stress. *Journal of Plant Biology* **51**:401-407.

Ajay K, Leech M, Vain P, Laurie DA and Christou P (1998) Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hot spots. *Proceedings of the National Academy of Sciences of the United States of America* **95**:7203-7208.

Albert H, Dale EC, Lee E and Ow DW (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant Journal* **7**:649-659.

Artelt P, Grannemann R, Stocking C, Friel J, Bartsch J and Hauser H (1991) The prokaryotic neomycin-resistance-encoding gene acts as a transcriptional silencer in eukaryotic cells. *Gene* **99**:249-254.

Assaad FF, Tucker KL and Signer ER (1993) Epigenetic repeat-induced gene silencing (RIGS) in Arabidopsis. *Plant Molecular Biology* **22**:1067-1085.

Bozas A, Beumer KJ, Trautman JK and Carroll D (2009) Genetic Analysis of Zinc-Finger Nuclease-Induced Gene Targeting in Drosophila. *Genetics* **182**:641-651.

Brown LR, Flavin C and French HF (1999) State of the World, New York: W.W. Norton & Co Inc.

Buchholz F, Angrand PO and Stewart AF (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nature Biotechnology* **16**:657-662.

Buchholz F, Ringrose L, Angrand PO, Rossi F and Stewart AF (1996) Different thermostabilities of FLP and Cre recombinases: Implications for applied site-specific recombination. *Nucleic Acids Research* **24**:4256-4262.

Carroll D, Beumer KJ, Morton JJ, Bozas A and Trautman JK (2008) Gene targeting in Drosophila and Caenorhabditis elegans with zinc-finger nucleases. *Methods in Molecular Biology*:63-77.

Chawla R, Ariza-Nieto M, Wilson AJ, Moore SK and Srivastava V (2006) Transgene expression produced by biolistic-mediated, site-specific gene integration is

consistently inherited by the subsequent generations. *Plant Biotechnology Journal* **4**:209-218.

Chen LJ, Lee DS, Song ZP, Suh HS and Lu BR (2004) Gene flow from cultivated rice (Oryza sativa) to its weedy and wild relatives. *Annals of Botany* **93**:67-73.

Chen Y and Rice PA (2003) New insight into site-specific recombination from Flp recombinase-DNA structures. *Annual Review of Biophysics and Biomolecular Structure* **32**:135-159.

Cotsaftis O, Sallaud C, Breitler JC, Meynard D, Greco R, Pereira A and Guiderdoni E (2002) Transposon-mediated generation of T-DNA- and marker-free rice plants expressing a Bt endotoxin gene. *Molecular Breeding* **10**:165-180.

Dai SH, Zheng P, Marmey P, Zhang SP, Tian WZ, Chen SY, Beachy RN and Fauquet C (2001) Comparative analysis of transgenic rice plants obtained by Agrobacterium-mediated transformation and particle bombardment. *Molecular Breeding* **7**:25-33.

Dale EC and Ow DW (1991) Gene-transfer with subsequent removal of the selection gene from the host genome. *Proceedings of the National Academy of Sciences of the United States of America* **88**:10558-10562.

Daniell H and Dhingra A (2002) Multigene engineering: dawn of an exciting new era in biotechnology. *Current Opinion in Biotechnology* **13**:136-141.

Daniels RE and Sheail J (1999) Genetic pollution: concepts, concerns and transgenic crops. *Gene Flow and Agriculture: Relevance for Transgenic Crops*:65-72.

Day CD, Lee E, Kobayashi T, Holappa LD, Albert H and Ow DW (2000) Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. *Genes & Development* **14**:2869-2880.

De Maagd RA, Bosch D and Stiekema W (1999) Bacillus thuringiensis toxin-mediated insect resistance in plants. *Trends in Plant Science* **4**:9-13.

De Majnik J, Joseph RG, Tanner GJ, Larkin PJ, Djordjevic MA, Rolfe BG and Weinman JJ (1997) A convenient set of vectors for expression of multiple gene combinations in plants. *Plant Molecular Biology Reporter* **15**:134-140.

De Vetten N, Wolters AM, Raemakers K, van der Meer I, ter Stege R, Heeres E, Heeres P and Visser R (2003) A transformation method for obtaining marker-free plants of a cross-pollinating and vegetatively propagated crop. *Nature Biotechnology* **21**:439-442.

De Wilde C, Van Houdt H, De Buck S, Angenon G, De Jaeger G and Depicker A (2000) Plants as bioreactors for protein production: avoiding the problem of transgene silencing. *Plant Molecular Biology* **43**:347-359.

Decarvalho F, Gheysen G, Kushnir S, Vanmontagu M, Inze D and Castresana C (1992) Suppression of beta-1,3-glucanase transgene expression in homozygous plants. *Embo Journal* **11**:2595-2602.

EFB (2001) European Federation of Biotechnology, Antibiotic resistance markers in genetically modified (GM) crops. Available from <www.efb-central.org/images/uploads/AntibioticRM_English.pdf>

Elmayan T and Vaucheret H (1996) Expression of single copies of a strongly expressed 35S transgene can be silenced post-transcriptionally. *Plant Journal* **9**:787-797.

FAO (2009) Food and Agricultural Organization statistics database (FAOSTAT) Available from <www.fao.org>.

Frame BR, Shou HX, Chikwamba RK, Zhang ZY, Xiang CB, Fonger TM, Pegg SEK, Li BC, Nettleton DS, Pei DQ and Wang K (2002) Agrobacterium tumefaciens-mediated transformation of maize embryos using a standard binary vector system. *Plant Physiology* **129**:13-22.

Gealy DR, Mitten DH and Rutger JN (2003) Gene flow between red rice (Oryza sativa) and herbicide-resistant rice (O-sativa): Implications for weed management. *Weed Technology* **17**:627-645.

Gilbertson L (2003) Cre-lox recombination: Cre-ative tools for plant biotechnology. *Trends in Biotechnology* **21**:550-555.

Gilland B (2002) World population and food supply - Can food production keep pace with population growth in the next half-century? *Food Policy* **27**:47-63.

Grant JE, Cooper PA and Dale TM (2004) Transgenic Pinus radiata from Agrobacterium tumefaciens-mediated transformation of cotyledons. *Plant Cell Reports* **22**:894-902.

Halpin C (2005) Gene stacking in transgenic plants - the challenge for 21st century plant biotechnology. *Plant Biotechnology Journal* **3**:141-155.

Halpin C, Barakate A, Askari BM, Abbott JC and Ryan MD (2001) Enabling technologies for manipulating multiple genes on complex pathways. *Plant Molecular Biology* **47**:295-310.

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. *Plant Journal* **6**:271-282.

Hoa TTC, Bong BB, Huq E and Hodges TK (2002) Cre/lox site-specific recombination controls the excision of a transgene from the rice genome. *Theoretical and Applied Genetics* **104**:518-525.

Hobbs SLA, Kpodar P and Delong CMO (1990) The effect of T-DNA copy number, position and methylation on reporter gene-expression in tobacco transformants. *Plant Molecular Biology* **15**:851-864.

Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKelver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, Meng XD, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD and Jaenisch R (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature Biotechnology* **27**:851-U110.

Hohn B, Levy AA and Puchta H (2001) Elimination of selection markers from transgenic plants. *Current Opinion in Biotechnology* **12**:139-143.

Huang SS, Gilbertson LA, Adams TH, Malloy KP, Reisenbigler EK, Birr DH, Snyder MW, Zhang Q and Luethy MH (2004) Generation of marker-free transgenic maize by regular two-border Agrobacterium transformation vectors. *Transgenic Research* **13**:451-461.

Jefferson RA, Kavanagh TA and Bevan MW (1987) GUS fusions - beta-glucuronidase as a sensitive and versatile gene fusion marker in higher-plants. *Embo Journal* **6**:3901-3907.

Kerbach S, Lorz H and Becker D (2005) Site-specific recombination in Zea mays. *Theoretical and Applied Genetics* **111**:1608-1616.

Khan MS and Maliga P (1999) Fluorescent antibiotic resistance marker for tracking plastid transformation in higher plants. *Nature Biotechnology* **17**:910-915.

Kilby NJ, Davies GJ, Snaith MR and Murray JAH (1995) FLP recombinase in transgenic plants - constitutive activity in stably transformed tobacco and generation of marked cell clones in *Arabidopsis*. *Plant Journal* **8**:637-652.

Kohli A, Griffiths S, Palacios N, Twyman RM, Vain P, Laurie DA and Christou P (1999) Molecular characterization of transforming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. *Plant Journal* **17**:591-601.

Komari T, Hiei Y, Saito Y, Murai N and Kumashiro T (1996) Vectors carrying two separate T-DNAs for co-transformation of higher plants mediated by Agrobacterium tumefaciens and segregation of transformants free from selection markers. *Plant Journal* **10**:165-174.

Kononov ME, Bassuner B and Gelvin SB (1997) Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: Evidence for multiple complex patterns of integration. *Plant Journal* **11**:945-957.

Kooter JM, Matzke MA and Meyer P (1999) Listening to the silent genes: transgene silencing, gene regulation and pathogen control. *Trends in Plant Science* **4**:340-347.

Li BC, Xie C and Qiu H (2009a) Production of selectable marker-free transgenic tobacco plants using a non-selection approach: chimerism or escape, transgene inheritance, and efficiency. *Plant Cell Reports* **28**:373-386.

Li ZS, Xing AQ, Moon BP, McCardell RP, Mills K and Falco SC (2009b) Site-Specific Integration of Transgenes in Soybean via Recombinase-Mediated DNA Cassette Exchange. *Plant Physiology* **151**:1087-1095.

Lloyd A, Plaisier CL, Carroll D and Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **102**:2232-2237.

Lloyd AM and Davis RW (1994) Functional expression of the yeast FLP/*FRT* site-specific recombination system in *Nicotiana*-tabacum. *Molecular & General Genetics* **242**:653-657.

Luo KM, Duan H, Zhao DG, Zheng XL, Deng W, Chen YQ, Stewart CN, McAvoy R, Jiang XN, Wu YH, He AG, Pei Y and Li Y (2007) 'GM-gene-deleter': fused loxP-FRT recognition sequences dramatically improve the efficiency of FLP or CRE recombinase on transgene excision from pollen and seed of tobacco plants. *Plant Biotechnology Journal* **5**:263-274.

Lyznik LA, Gordon-Kamm WJ and Tao Y (2003) Site-specific recombination for genetic engineering in plants. *Plant Cell Reports* **21**:925-932.

Maliga P (2002) Engineering the plastid genome of higher plants. *Current Opinion in Plant Biology* **5**:164-172.

Matthews PR, Wang MB, Waterhouse PM, Thornton S, Fieg SJ, Gubler F and Jacobsen JV (2001) Marker gene elimination from transgenic barley, using co-transformation with adjacent 'twin T-DNAs' on a standard Agrobacterium transformation vector. *Molecular Breeding* **7**:195-202.

Matzke AJM and Matzke MA (1998) Position effects and epigenetic silencing of plant transgenes. *Current Opinion in Plant Biology* **1**:142-148.

Matzke AJM, Neuhuber F, Park YD, Ambros PF and Matzke MA (1994) Homology-dependent gene silencing in transgenic plants - epistatic silencing loci contain multiple copies of methylated transgenes. *Molecular & General Genetics* **244**:219-229.

Messeguer J, Marfa V, Catala MM, Guiderdoni E and Mele E (2004) A field study of pollen-mediated gene flow from Mediterranean GM rice to conventional rice and the red rice weed. *Molecular Breeding* **13**:103-112.

Meyer P and Saedler H (1996) Homology-dependent gene silencing in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**:23-48.

Miller M, Tagliani L, Wang N, Berka B, Bidney D and Zhao ZY (2002) High efficiency transgene segregation in co-transformed maize plants using an Agrobacterium tumefaciens 2 T-DNA binary system. *Transgenic Research* **11**:381-396.

Nanto K, Sato K, Katayama Y and Ebinuma H (2009) Expression of a transgene exchanged by the recombinase-mediated cassette exchange (RMCE) method in plants. *Plant Cell Reports* **28**:777-785.

Napoli C, Lemieux C and Jorgensen R (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* **2**:279-289.

Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu JR, Warner DC, Anstrom DC, Bensen RJ, Castiglioni PP, Donnarummo MG, Hinchey BS, Kumimoto RW, Maszle DR, Canales RD, Krolikowski KA, Dotson SB, Gutterson N, Ratcliffe OJ and Heard JE (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres, in *Proceedings of the National Academy of Sciences of the United States of America* pp 16450-16455, Washington; USA: National Academy of Sciences.

Nielsen KM, Bones AM, Smalla K and van Elsas JD (1998) Horizontal gene transfer from transgenic plants to terrestrial bacteria - a rare event? *Fems Microbiology Reviews* **22**:79-103.

Odell JT, Hoopes JL and Vermerris W (1994) Seed-specific gene activation mediated by the Cre/lox site-specific recombination system. *Plant Physiology* **106**:447-458.

Oki T and Kanae S (2006) Global hydrological cycles and world water resources. *Science (Washington)* **313**:1068-1072.

Ow DW (2002) Recombinase-directed plant transformation for the post-genomic era. *Plant Molecular Biology* **48**:183-200.

Park J, Lee YK, Kang BK and Chung WI (2004) Co-transformation using a negative selectable marker gene for the production of selectable marker gene-free transgenic plants. *Theoretical and Applied Genetics* **109**:1562-1567.

Pawlowski WP and Somers DA (1998) Transgenic DNA integrated into the oat genome is frequently interspersed by host DNA. *Proceedings of the National Academy of Sciences of the United States of America* **95**:12106-12110.

Pontiroli A, Rizzi A, Simonet P, Daffonchio D, Vogel TM and Monier JM (2009) Visual Evidence of Horizontal Gene Transfer between Plants and Bacteria in the Phytosphere of Transplastomic Tobacco. *Applied and Environmental Microbiology* **75**:3314-3322.

Puchta H, Dujon B and Hohn B (1993) Homologous recombination in plant-cells is enhanced by *in-vivo* induction of double-strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Research* **21**:5034-5040.

Que QD, Wang HY, English JJ and Jorgensen RA (1997) The frequency and degree of cosuppression by sense chalcone synthase transgenes are dependent on transgene promoter strength and are reduced by premature nonsense codons in the transgene coding sequence. *Plant Cell* **9**:1357-1368.

Raymond CS and Soriano P (2007) High-Efficiency FLP and Phi C31 Site-Specific Recombination in Mammalian Cells. *Plos One* **2**.

Remy S, Tesson L, Menoret S, Usal C, Scharenberg AM and Anegon I (2010) Zinc-finger nucleases: a powerful tool for genetic engineering of animals. *Transgenic Res* **19**:363-371.

Ringrose L, Lounnas V, Ehrlich L, Buchholz F, Wade R and Stewart AF (1998) Comparative kinetic analysis of FLP and Cre recombinases: Mathematical models for DNA binding and recombination. *Journal of Molecular Biology* **284**:363-384.

Robinson SJ and Parkin IAP (2009) Bridging the Gene-to-Function Knowledge Gap Through Functional Genomics. *Plant Genomics: Methods and Protocols*:153-173.

Sauer B (1987) Functional expression of the Cre-lox site-specific recombination system in the yeast Saccharomyces-cerevisiae. *Molecular and Cellular Biology* **7**:2087-2096.

Schlake T and Bode J (1994) Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* **33**:12746-12751.

Schwartz CJE and Sadowski PD (1990) FLP protein of 2-Mu circle plasmid of yeast induces multiple bends in the FLP recognition target site. *Journal of Molecular Biology* **216**:289-298.

Seckler D, Barker R and Amarasinghe U (1999) Water scarcity in the twenty-first century. *International Journal of Water Resources Development* **15**:29-42.

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng XD, Choi VM, Rock JM, Wu YY, Katibah GE, Zhifang G, McCaskill D, Simpson MA, Blakeslee B, Greenwalt SA, Butler HJ, Hinkley SJ, Zhang L, Rebar EJ, Gregory PD and Urnov FD (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature* **459**:437-U156.

Sinclair TR (1998) Historical changes in harvest index and crop nitrogen accumulation. *Crop Science* **38**:638-643.

Song ZP, Lu BR, Zhu YG and Chen JK (2003) Gene flow from cultivated rice to the wild species Oryza rufipogon under experimental field conditions. *New Phytologist* **157**:657-665.

Srivastava V, Ariza-Nieto M and Wilson AJ (2004) Cre-mediated site-specific gene integration for consistent transgene expression in rice. *Plant Biotechnology Journal* **2**:169-179.

Srivastava V and Gidoni D (2010) Site-specific gene integration technologies for crop improvement. *In Vitro CellDevBiol - Plant*.

Srivastava V and Ow DW (2002) Biolistic mediated site-specific integration in rice. *Molecular Breeding* **8**:345-350.

Srivastava V and Ow DW (2004) Marker-free site-specific gene integration in plants. *Trends in Biotechnology* **22**:627-629.

Timmons AM, Charters YM, Crawford JW, Burn D, Scott SE, Dubbels SJ, Wilson NJ, Robertson A, Obrien ET, Squire GR and Wilkinson MJ (1996) Risks from transgenic crops. *Nature* **380**:487-487.

Timmons G (1996) Science and education in the first half of the nineteenth century. *Endeavour* **20**:140-143.

Townsend JA, Wright DA, Winfrey RJ, Fu FL, Maeder ML, Joung JK and Voytas DF (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* **459**:442-U161.

USDA (2009) United States Department of Agriculture, Available from <www.usda.gov>.

Vanderkrol AR, Mur LA, Beld M, Mol JNM and Stuitje AR (1990) Flavonoid genes in petunia - addition of a limited number of gene copies may lead to a suppression of gene-expression. *Plant Cell* **2**:291-299.

Vergunst AC, Jansen LET and Hooykaas PJJ (1998) Site-specific integration of Agrobacterium T-DNA in Arabidopsis thaliana mediated by Cre recombinase. *Nucleic Acids Research* **26**:2729-2734.

Wang Y, Chen BJ, Hu YL, Li JF and Lin ZP (2005a) Inducible excision of selectable marker gene from transgenic plants by the Cre/lox site-specific recombination system. *Transgenic Research* **14**:605-614.

Wang Y, Chen BJ, Hu YL, Li JF and Lin ZP (2005b) Inducible excision of selectable marker gene from transgenic plants by the Cre/lox site-specific recombination system. *Transgenic Research* **14**:605-614.

Wright DA, Townsend JA, Winfrey RJ, Irwin PA, Rajagopal J, Lonosky PM, Hall BD, Jondle MD and Voytas DF (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant Journal* **44**:693-705.

Wu J, Kandavelou K and Chandrasegaran S (2007) Custom-designed zinc finger nucleases: What is next? *Cellular and Molecular Life Sciences* **64**:2933-2944.

Ye F and Signer ER (1996) RIGS (repeat-induced gene silencing) in Arabidopsis is transcriptional and alters chromatin configuration. *Proceedings of the National Academy of Sciences of the United States of America* **93**:10881-10886.

Zhang W, Subbarao S, Addae P, Shen A, Armstrong C, Peschke V and Gilbertson L (2003) Cre/lox-mediated marker gene excision in transgenic maize (Zea mays L.) plants. *Theoretical and Applied Genetics* **107**:1157-1168.

Zhao Y, Qian Q, Wang HZ and Huang DN (2007) Co-transformation of gene expression cassettes via particle bombardment to generate safe transgenic plant without any unwanted DNA. *In Vitro Cellular & Developmental Biology-Plant* **43**:328-334.

Zuo JR, Niu QW, Moller SG and Chua NH (2001) Chemical-regulated, site-specific DNA excision in transgenic plants. *Nature Biotechnology* **19**:157-161.
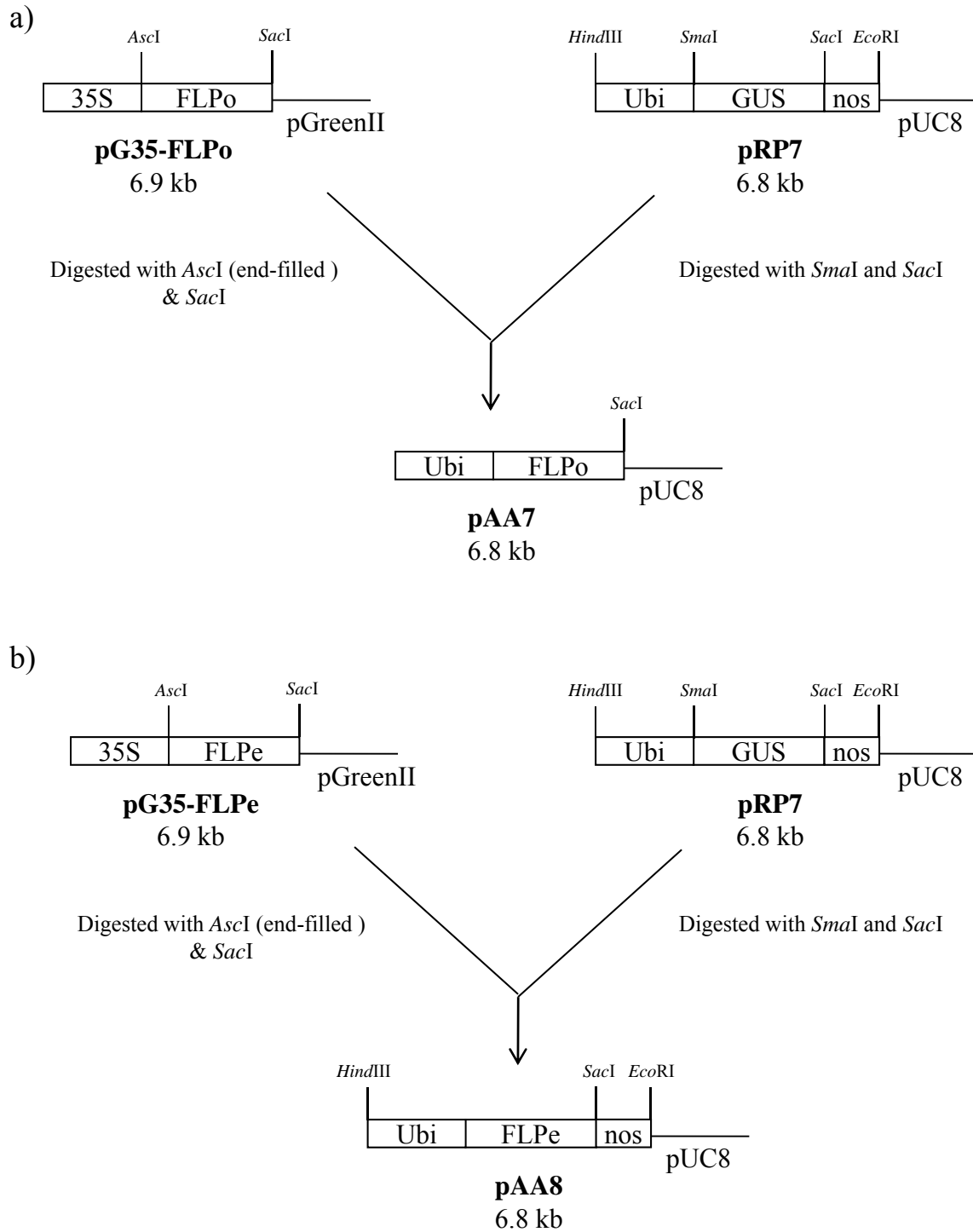
Appx. Fig. 1 Diagrammatic illustration of construction of the plasmids (a) pAA7 and (b) pAA8. *Ubi*: maize ubiquitin promoter; *GUS*: β- glucuronidase gene; *35S*: cauliflower mosaic virus 35S promoter; *nos*: nopaline synthase transcription terminator.
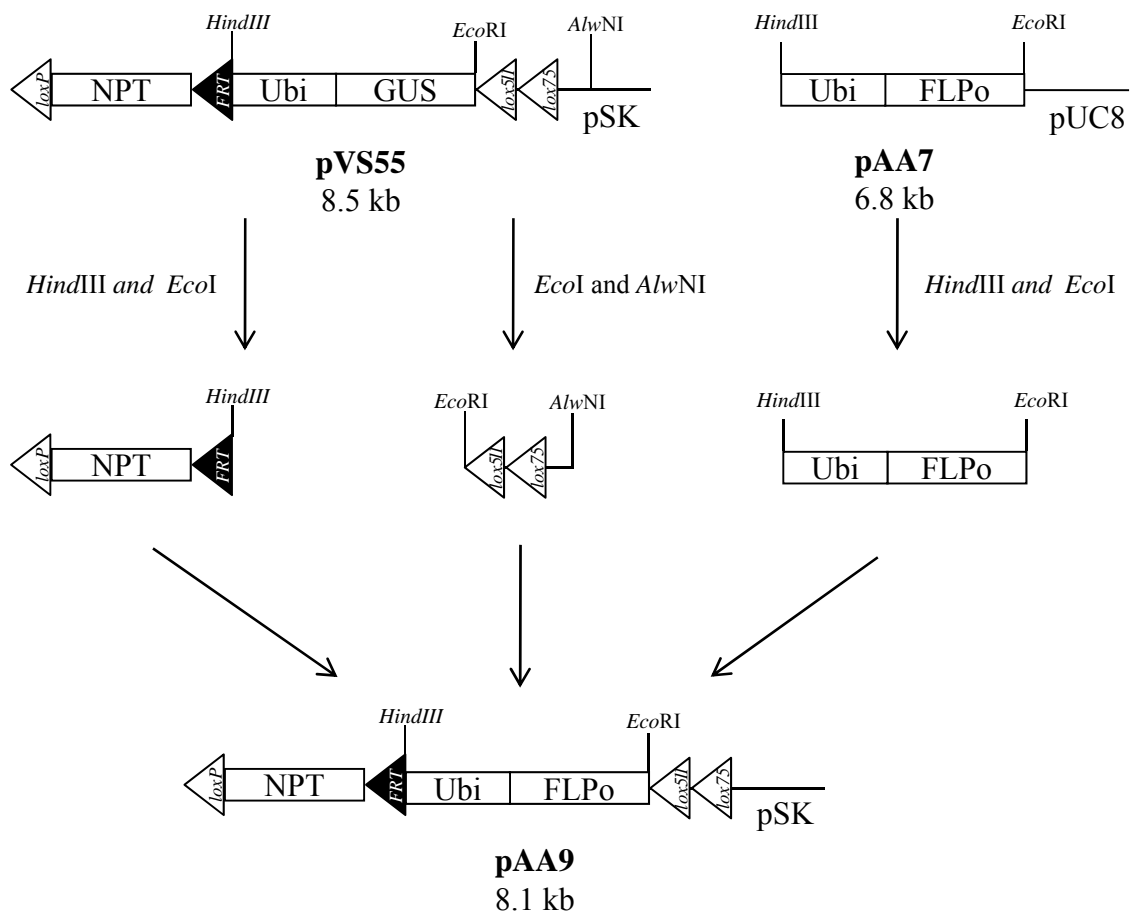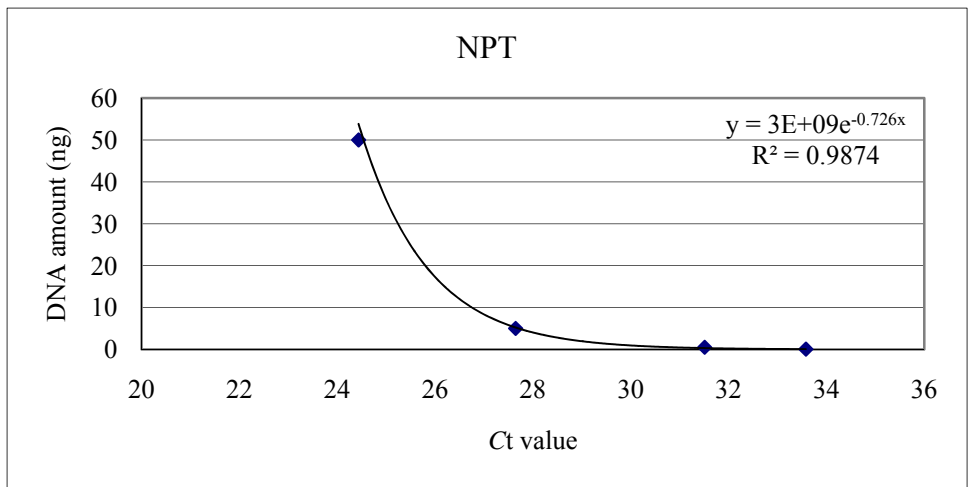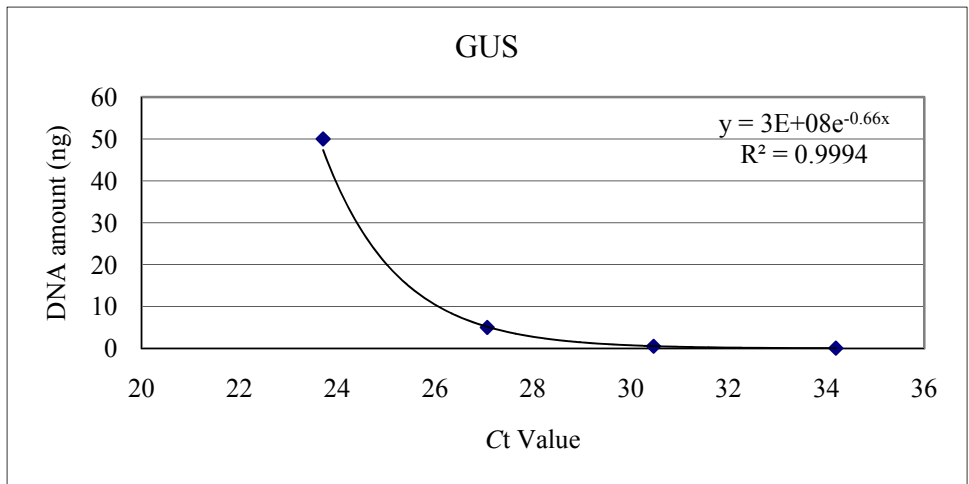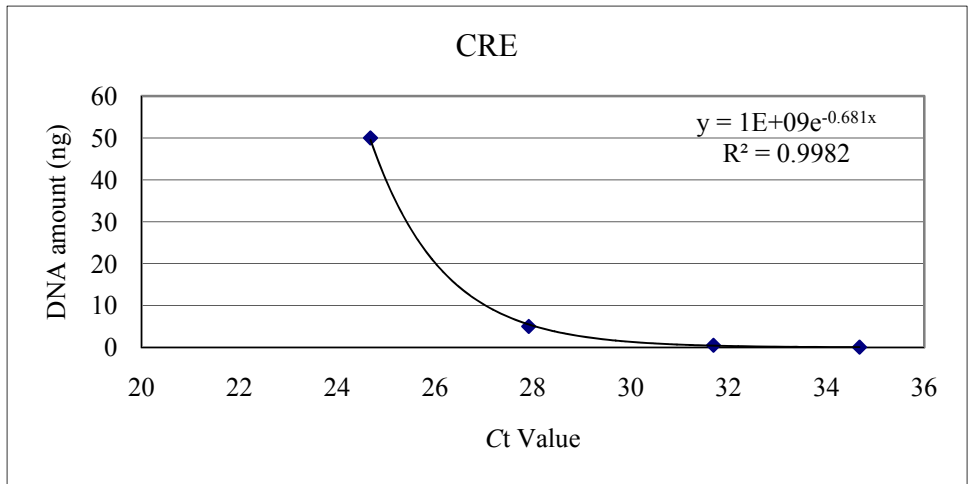
Appx. Fig. 2 Diagrammatic illustration of construction of the plasmid pAA9. *Ubi*: maize ubiquitin promoter; *NPT*: neomycin phosphotransferase II gene; *GUS*: β-glucuronidase gene. Each ORF is followed by *nos3′* (nopaline synthase transcription terminator) (not shown).
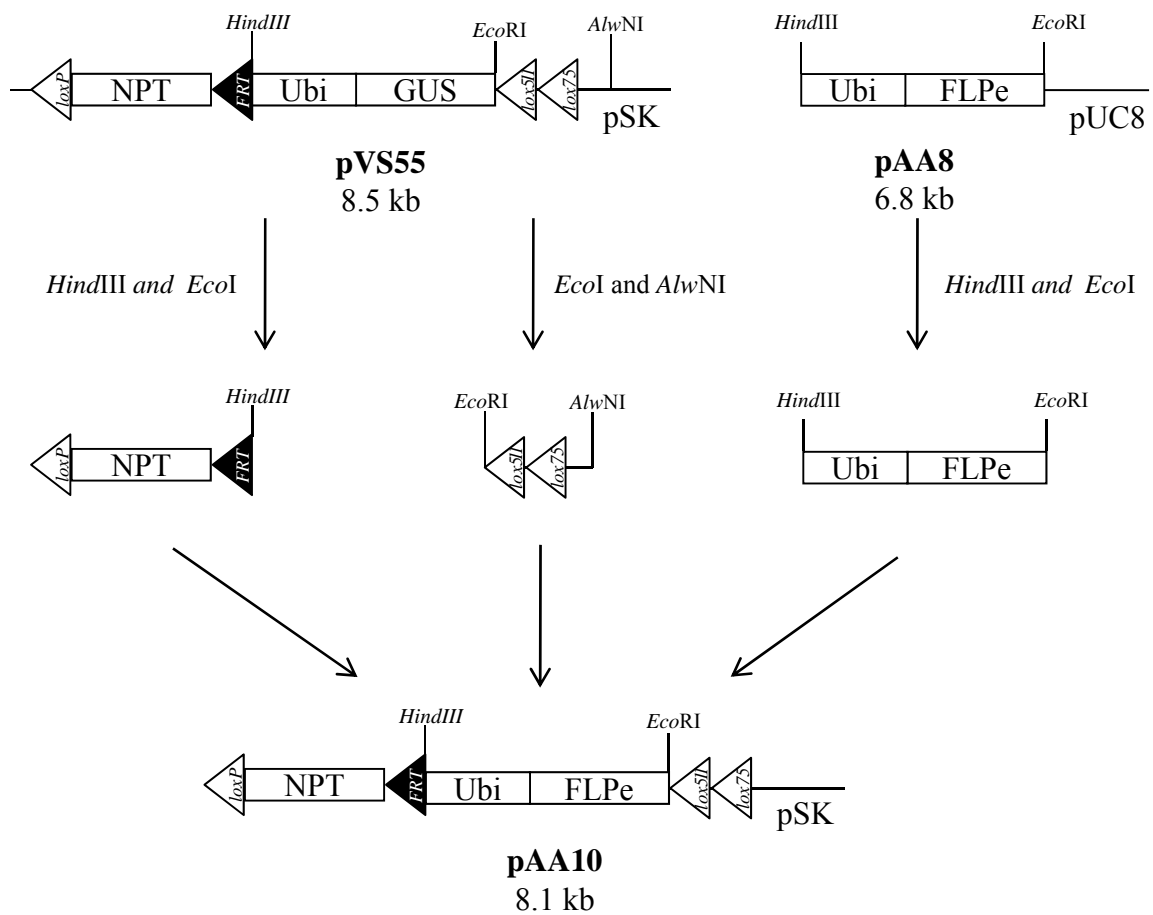
Appx. Fig. 3 Diagrammatic illustration of construction of the plasmid pAA10. *Ubi*: maize ubiquitin promoter; *NPT*: neomycin phosphotransferase II gene; *GUS*: β-glucuronidase gene. Each ORF is followed by *nos3′* (nopaline synthase transcription terminator) (not shown).

P1

P2
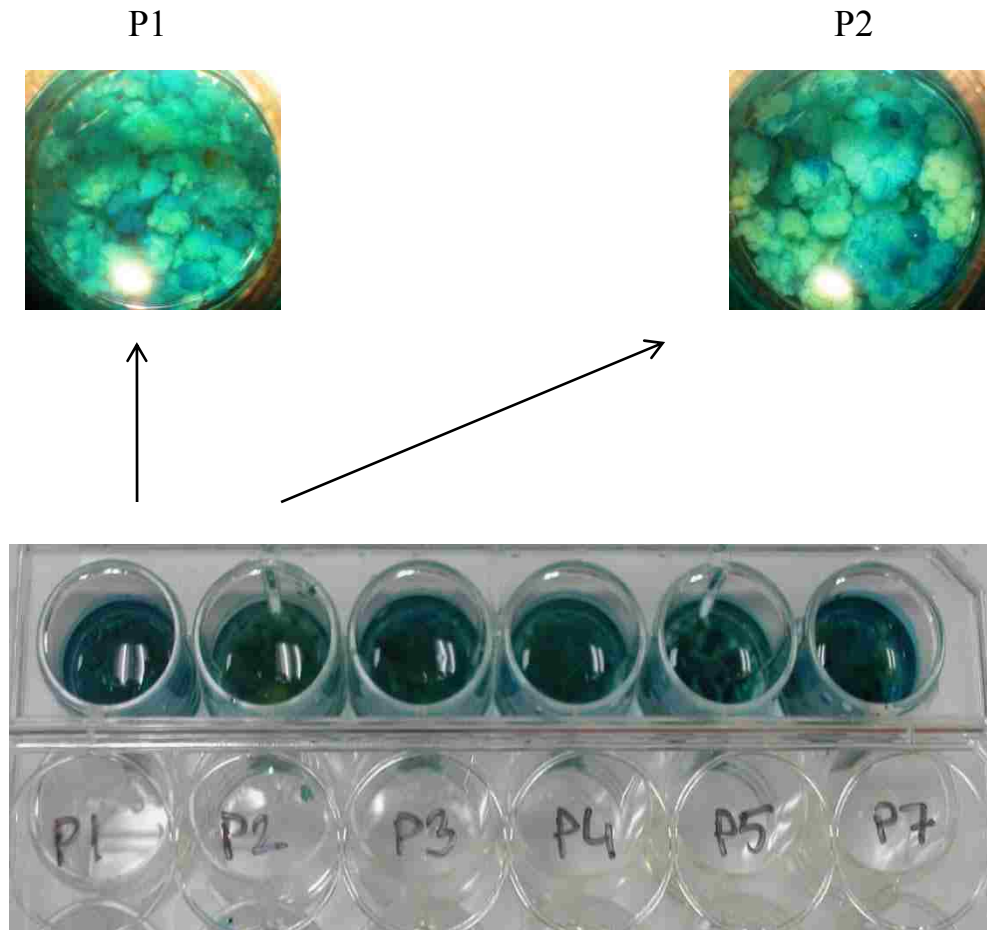
Appx. Fig 4. Histo-chemical staining on P lines, indicating GUS expression in cells.

CRE

$$y = 1E+09e^{-0.681x}$$
$$R^2 = 0.9982$$

GUS

$$y = 3E+08e^{-0.66x}$$
$$R^2 = 0.9994$$

NPT

$$y = 3E+09e^{-0.726x}$$
$$R^2 = 0.9874$$

Appx. Fig 5. Standard curves obtained from CRE, GUS and NPT amplifications using the dilutions of P1 DNA (50, 5, 0.5, 0.05 ng) in RT-qPCR.