

Review

Gene tagging in rice: a high throughput system for functional genomics

Jong-Seong Jeon^a, Gynheung An^{b,*}

^a Department of Plant Pathology, University of California, Davis, CA 95616, USA

^b National Research Laboratory of Plant Functional Genomics, Division of Molecular and Life Sciences, Pohang University of Science and Technology (POSTECH), Pohang 790-784, South Korea

Received 12 January 2001; received in revised form 19 March 2001; accepted 19 March 2001

Abstract

Rice, with its small genome size and well-characterized molecular information, is an ideal model plant for cereal genomics research. Sequence of the rice (*Oryza sativa*) genome will be determined by the International Rice Genome Sequencing Project (IRGSP) in the near future. Therefore, a large population of mutant plants should be required for adequately assigning function to the abundant sequence information. Here we summarize strategies as well as the progress that has been made in producing gene tags that may be invaluable for understanding the functional genomics of rice. © 2001 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Functional genomics; Gene tagging; *Oryza sativa*

1. Introduction

During the last decade, both expressed sequence tags (ESTs) and whole-genome sequencing projects have increased tremendously the amount of molecular information available for the genomes of *Arabidopsis* and rice (*Oryza sativa*). Recently, plant biologists have used sequence information to uncover the functions of many genes. Various mutants, such as gene knockouts or null mutations, are invaluable for providing the biological variability that is essential when assigning functions to the large quantity of sequence information.

Rice has become a model for monocot plants because of the accumulation of molecular information for this species [1–4]; efficiency in transformation [5]; its small (430 Mb) genome [6]; and the

economical importance of this crop [7]. Its gene content is comparable to that of other grass plants, such as wheat, maize, barley, rye, and sorghum [8]. Because of the conservation of gene sequences and order among cereals, the structural and functional analyses of rice have broad practical implications for these other economically important crops. Thus, genomic information for rice should be widely applicable when developing products and technologies in both the rice and non-rice sectors.

Recent studies have demonstrated that results obtained from *Arabidopsis* experiments cannot simply be applied to highly diverse species such as rice [9,10]. For example, whereas *Arabidopsis* *LEAFY* (*LFY*) regulates the formation of floral meristems, *RFL*, a *LFY* homolog of rice, appears to be involved in panicle branch initiation [9]. Studying the co-linearity remaining between two small regions of *Arabidopsis* Chromosome 1 and rice has shown that conservation of gene orders is

* Corresponding author. Fax: +82-54-2792199.
E-mail address: genean@postech.ac.kr (G. An).

no longer identifiable when using comparative mapping [10]. More than 45 000 rice ESTs are currently available to the public. However, only 25% of them show significant homology to known genes, and the function of most of the genes is undetermined. Functional genomics approaches are needed; insertional mutagenesis on the genome scale should provide a high throughput system for future work. In this review we summarize recent research as well as strategies for gene tagging in rice.

2. T-DNA tagging

In *Arabidopsis*, a model dicot species, a number of genes have been isolated from T-DNA-tagged lines [11]. The *Arabidopsis* knockout facility at the University of Wisconsin recently established a population of 60 480 T-DNA-tagged lines [12]. This is a significant step toward the production of genome-wide mutations. T-DNA insertion in *Arabidopsis* probably is a random event, with the inserted sequences being stable through multiple generations [11]. Recently, this T-DNA tagging

strategy has been employed in rice [13]. We have now generated approximately 30 000 fertile rice lines that have been tagged by T-DNA using pGA2144 (Fig. 1A) ([13], unpublished data). Analysis of randomly selected transgenic plants has indicated an average of 1.4 loci of T-DNA inserts per line. Therefore, approximately 42 000 tags are estimated to have been generated.

T-DNA tagging efficiency can be estimated using the following formula, if insertion occurs randomly on the chromosomes:

$$P = 1 - (1 - (L/C))^n f,$$

where P is the probability of finding an insertion within a given gene, L is an average length for a gene, C is the haploid genome size, n is the number of T-DNA tagging lines, and f is the average number of loci inserted per line. Considering that: (1) the rice haploid genome size is 4.3×10^8 bp; (2) an average length for a rice gene is 3.0 kb; and (3) the mean number of T-DNA insertion loci per line is 1.4, our current tagged lines would provide 25% probability of finding an insertion within a given gene. Therefore, 471 000 tagging lines would be required for establishing a popula-

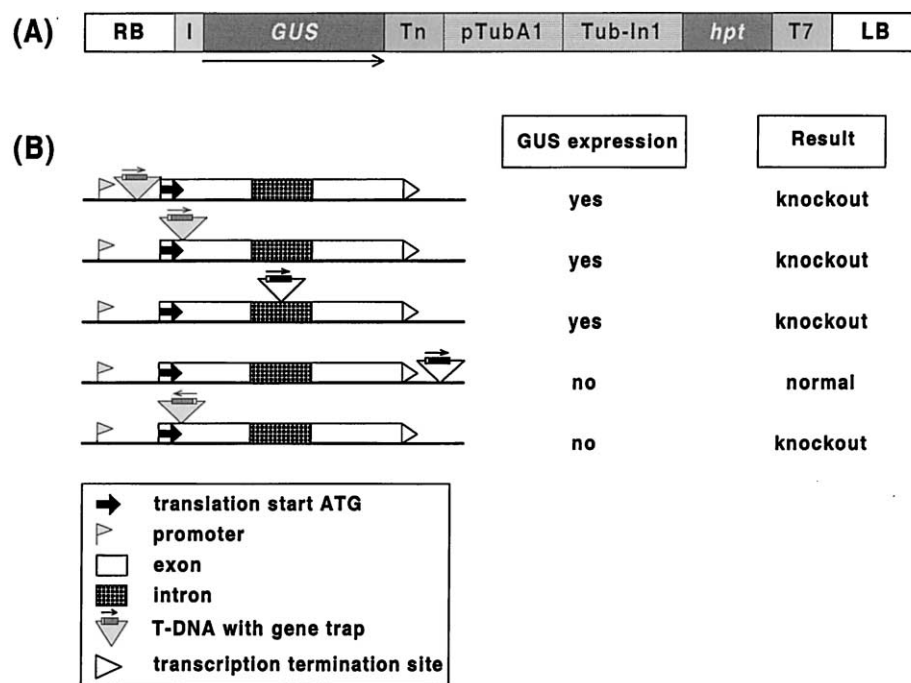


Fig. 1. (A) Map of the T-DNA tagging vector pGA2144. The gene trap of T-DNA consisted of three putative splicing acceptor and donor sites and the *GUS* gene. RB and LB represent the right and left borders of T-DNA, respectively. *GUS*, β -glucuronidase; Tn, *nopaline synthase* (*nos*) terminator; *hpt*, *hygromycin phosphotransferase*; T7, transcription termination region of Gene 7 of the pTiA6; pTubA1, the promoter of the rice α -tubulin gene *OsTubA1* [14]; Tub-In1, the first intron of *OsTubA1*; I, the *OsTubA1* intron 3 carrying three putative splicing acceptor and donor sites; (B) Different outcomes of transgenic plants. Thin arrow indicates the direction of *GUS* gene transcription.

tion in which a T-DNA insertion could be found within a given gene at 99% probability.

The number of tagging lines necessary for saturating all the rice genes may be smaller than the estimated value because T-DNA insertions have been found to occur preferentially in gene-rich regions in rice [15,16]. Nuclear DNAs isolated from a pool of independent transformants of both *Arabidopsis* and rice were fractionated by preparative centrifugation in Cs₂SO₄ density gradients in the presence of 3,6-bis(acetatomercurimethyl)-1,4-dioxane (BAMD) [16]. Hybridization of the GUS sequence present in the T-DNA with the Cs₂SO₄/BAMD fractions showed that the signals were localized in fractions, which corresponds to gene-rich regions in both species. Because genes occupy most of the *Arabidopsis* genome, the T-DNA appears to be integrated essentially everywhere in the genome. In contrast, integration of the T-DNA in rice was detected only in the gene space, i.e. transcriptionally active regions that correspond to 100 Mb, or 25% of its entire genome. This is approximately the size of the *Arabidopsis* genome.

If rice contains 30 000 genes that are randomly distributed on its chromosomes, one gene should be found every 14 kb [17]. However, DNA sequencing analyses have revealed that the distance between genes is shorter. For example, the complete sequence of 340 kb DNA around the *Adh1–Adh2* region on Chromosome 11 has shown 33 putative genes, which corresponds to one gene per 10.3 kb [18]. Within the 7 Mb area of the genomic sequences on Chromosomes 1, 2, 3, and 8 (as published in the DDBJ/EMBL/GenBank databases), 1370 putative genes have been identified either functionally or by exon prediction algorithms. This corresponds to one gene every 5.3 kb. Because the Rice Genome Project (RGP) used EST markers to analyze DNA sequences of PAC clones, the gene-rich regions may have been selected preferentially [19]. This suggests that gene density differs significantly in rice chromosomes, probably due to uneven distribution of repetitive DNA that constitutes at least 50% of the genome [20].

One great difficulty in establishing a population of tagged lines is obtaining enough seeds from the first generation. In our experiments with 10 743 primary transgenic plants, 39% bore fewer than 50 seeds and only 50% yielded more than 100 [13]. Yin and Wang [21] had similar results from their

analysis of 2633 transgenic rice plants. Although the majority of the transgenic plants became fully fertile in the next generation, 22% of the lines showed less than 50% fertility. Reduced fertility is probably due, in part, to the lethal effects of T-DNA insertion. However, a considerable number of lethal mutations likely are induced during the T-DNA transformation process. Chromosome rearrangement or modification might also occur during transformation. Nevertheless, those lines that show low fertility must be individually amplified before being used for further analysis.

T-DNA-tagged lines are useful resources for studying gene function. First, they can be used to identify mutants whose function is altered by T-DNA insertion. Because most mutations are recessive, their phenotypes are easily detected in a segregating population. However, because rice plants are relatively tall and wide, compared with *Arabidopsis*, they need more growing space. Therefore, such a forward screening of phenotypes can be employed only for limited lines and times. In analyzing 1600 lines at the T2 generation, we observed various mutant phenotypes during the vegetative stages (unpublished data). The most common characteristics were the dwarf phenotype (7.0%) and leaf-pigment mutations (9.5%), such as albino, pale green, chlorina, striped, or zebra (transverse green and chlorotic bands). Spotted leaves (1.0%) and leaf-morphology mutations (1.2%) were also found. Seedling mortality was 1.1%. The T-DNA-tagged population also carried low frequencies of mutants in the reproductive organs, e.g. depressed paleae, filamentous flowers, extra glumes, and long sterile glumes. About 1.8% of the lines were completely sterile, and flowering-time mutations included early- (0.2%) and late- (0.1%) flowering phenotypes.

One must determine whether these phenotypes induced by the rice transformation are due to the T-DNA insertion. Once the mutant phenotype co-segregates with T-DNA, the sequence flanking T-DNA can then be isolated by PCR-based methods, such as thermal asymmetric interlaced (TAIL) PCR [22], adapter-ligated PCR [23], inverse PCR [24], a universal biotinylated adapter amplification procedure [25], a panhandle PCR [26], or a plasmid rescue system [27].

Second, the T-DNA insertional line can be identified in a given gene via PCR-based screening [12,28]. Using a gene-specific primer and a primer

located near the end of the T-DNA, a DNA fragment flanking the inserted T-DNA can be amplified and its sequence then determined. Because screening of individual lines requires great effort, DNA pools of a large number of lines are commonly used for the reverse genetic approach. For example, in *Arabidopsis* at least 200 copies of a T-DNA insert were required in a complex pool of DNA to ensure that the insert DNA was reliably detected by PCR. The efficiency was attenuated when more than 125 ng of pooled DNA was used. These limitations, therefore, set the maximum useful pool size at ~ 2350 lines per pool [12]. Because the rice genome is about four times larger than *Arabidopsis* genome, a pool of 500 lines should be sufficient for PCR-based screening of the knockout mutations in rice.

Third, random sequence analysis of the fragments flanking T-DNA can help identify gene function. Flanking sequences from *Ds* insertions of 931 independent transgenic lines in *Arabidopsis* have been characterized [29]. Likewise, direct-sequence characterization of 761 *Mutator*-tagged fragments in maize has shown that random sequencing of transposon-tagged fragments is capable of producing significant numbers of interesting transposon-tagged genes and mutant plant lines [25]. Construction of such a database will be valuable for determining gene function on a large scale. Developing efficient methods for finding the DNA sequences that are flanked by T-DNA will facilitate this approach.

Finally, T-DNA tagging lines can be used for screening mutants caused by endogenous transposons. One such element is a retrotransposon, *Tos17*, that is activated during tissue-culture stress [[30], see below]. Our T-DNA tagged lines possess an average of four new copies of *Tos17* (unpublished data).

3. Transposon tagging

Transposon tagging has become a powerful tool for isolating new genes since the controlling element was first recognized by McClintock [31]. A number of genes have been isolated using endogenous transposons as tags [32]. The first successful cloning of a plant gene was achieved via the *Ac/Ds* (Activator–Dissociation) transposon system [33]. Other transposon systems, such as *En/Spm* (En-

hancer/Suppressor-mutator) and *Mu* (Mutator), have been used for cloning several genes of maize [26,34].

The maize *Ac/Ds* system has been tested for gene tagging in rice. First, the autonomous *Ac* element was cloned between a promoter and the *hygromycin phosphotransferase*-coding region. The construct was then introduced into rice chromosomes by direct transformation. Transposition of the *Ac* element was proven by recovering hygromycin-resistant plants [35,36]. Enoki et al. [37] analyzed the behavior of 559 plants from four transgenic rice families through three successive generations; 18.9% of the plants contained newly transposed *Ac* insertions. They tested a PCR-based, three-dimensional matrix of gene knockouts in 6000 *Ac* plants. Of the 14 randomly selected genes, two knockouts were identified, one of which encodes the rice cytochrome P450 (CYP86) gene. Molecular analysis of *Ac* transmission to the progeny revealed that their germinal transmission was detected in nine of the 12 insertions, evidence that some *Ac* transmissions occurred in somatic tissues that would not transmit to the progeny. Sequence analysis of 99 flanking regions over the 5' region of *Ac* indicated that it preferentially transposed into protein-coding regions.

The non-autonomous *Ds* element has been transposed in the presence of *Ac* transposase via the direct gene transfer method [38,39]. Germinal transposition of *Ds* was observed at high frequency in the R2 progeny when a transgenic plant containing the *Ds* element was crossed with a transgenic plant carrying *Ac* transposase under the control of the cauliflower mosaic virus (CaMV) 35S promoter. A wide spectrum of mutations, affecting growth, morphogenesis, flowering time, and disease resistance, was observed in the *Ds* population [40]. Whether these mutations are due to *Ds* must still be determined. The frequency of *Ds* transposition significantly declined in subsequent generations, even in the presence of *Ac* transposase. However, transposition could be recovered in certain lines through protoplast regeneration.

Similar *Ac/Ds* gene tagging systems have been developed for rice by employing the *Agrobacterium*-mediated gene delivery method [41,42]. This method mediates transfer of one to a few copies of T-DNA into plant cell chromosomes,

thereby making the transgene more stable compared with the direct gene transfer method [43]. As many as 80% of the *Ds* elements have been excised from the original integration sites in the presence of *Ac* transposase activity that was provided in *trans*. The germinal excision frequency was no more than 40% [42]. Repetitive ratoon culturing caused new transposition, at about 30% frequency, thereby demonstrating that the procedure is one of the strategies capable of producing a large population of mutants [41].

Linkage analysis of the empty donor site and the transposed *Ds* insertion site revealed that four of five insertion sites were linked to the donor site [42]. This indicates that most of the transposition events occurred in a tightly linked site, as had been reported for *Arabidopsis* [32,44]. In that species, the physical map positions of 356 *Ds* insertions showed significant preference for transposition to the regions adjacent to the nucleolus organizer regions [29]. For any given gene, insertions preferentially occurred at the 5' end of a gene. Therefore, the short-range and highly preferential transposition system can be effectively utilized for the targeted mutagenesis of closely linked genes.

In contrast, transpositions of another mobile element, *En/spm*, were well distributed over different chromosomes and were not clustered in a few genomic locations in *Arabidopsis* [45]. However, the tendency of the *En/Spm* element to amplify can potentially complicate interpretation of phenotypes and molecular analyses. *Ac/Ds* in conjunction with *Spm/dSpm* has been proposed for use in developing a four-element system for directed tagging of crop-specific alleles [46]. In the first step, stocks carrying *Ds* within *dSpm* would be crossed to a line with *Spm* to facilitate random transposition of the transposon. In the second step, individual lines would be crossed to a line with *Ac* to induce local transpositions of the *Ds* element. This four-element system would, therefore, exploit the natural tendency of unlinked jumps of *dSpm* and linked jumps of *Ds*.

Unfortunately, the *Ac* element in this system would undergo frequent excision from the target gene, causing variegation. To overcome this, a self-stabilizing *Ac* derivative has been developed that undergoes autonomous transposition but is stable after integration [47]. Charng et al. [48] have designed an inducible transposon system using an element that contains a *PR-1a* pro-

moter::transposase fusion (*PR-1a::Tpase*). Treatment with salicylic acid (SA) induces expression of transposase, resulting in the transposition of *Ds* in somatic tissues. Germinal transposition events also have been observed after SA treatment of flowers. Development of inducible transposable element systems will also be valuable for rice.

4. Retrotransposon tagging

Another effort toward gene tagging in rice involves introducing tobacco retrotransposon *Tto1*, and demonstrating its autonomous transposition through reverse transcription [49]. A transposable element, *Tag1*, from *Arabidopsis* has also been analyzed in rice [50]. There, transcription and excision behavior of *Tag1* were similar in both rice and *Arabidopsis*. However, the excision was tissue-specific only in rice.

A survey of 73 000 sequence-tagged-connectors (STC), corresponding to nearly 50 Mb of the rice genome, showed that retroelements are randomly distributed with respect to potential genes [51]. Retrotransposons are both functionally and structurally different from well-characterized transposable elements such as *Ac* and *Spm*. These retrotransposons are believed to be involved in gene duplication as well as regulation of gene expression. Transposition of some retroelements can be induced by stresses caused by pathogen infection, cell culture, and wounding [30,52].

Such endogenous elements in the rice genome have provided excellent tools for gene tagging. One such element is *Tos17*, which is activated during tissue culture and amplified up to 30 copies [30]. A gene knockout system using *Tos17* has been developed for identifying insertional mutations in a number of genes. After screening 550 plants that were mutagenized by *Tos17*, Sato et al. [28] have identified a mutation in the homeobox gene *OSH15*. Rice phytochrome A (*phyA*) mutant lines have also been isolated from *Tos17*-induced mutant population using a three-dimensional DNA-pooling system [53]. By screening for viviparous mutants, Agrawal et al. [54] have identified *Tos17* insertions in the rice zeaxanthin epoxidase gene (*OsABA1*) and in a novel *OsTATC* gene, which shows a weak homology with bacterial Sec-independent translocase *TATC*. More than 32 000 *Tos17* insertion lines have been pro-

duced and over 8600 independent sequence flanking insertion sites have been determined [55]. Analysis of insertion points in the rice genome sequence have revealed that insertions in the genic regions (exon and intron) were two-fold higher than in the intergenic region. This suggests that *Tos17* prefers the genic region for target sites. In addition, hot spots of *Tos17* insertions were observed, with some of them being clustered. This may reduce the efficiency of genome-wide gene disruption.

5. Activation tagging

Conventional strategies, such as T-DNA mutagenesis or transposon tagging, are not efficient for analyzing the function of redundant genes. Likewise, neither T-DNA tagging nor transposon tagging will identify the genes that are required during multiple stages of a life cycle and whose loss of function results in early embryonic or gametophytic mortality [27]. Fewer than 10% of the genes tagged in the *Arabidopsis* genome are likely to generate a visible phenotypic change. Therefore, complementing technologies are needed for assessing the function of the remaining genes.

One such technique is the activation-tagging system developed in *Arabidopsis* and successful for cloning several genes [27,56–58]. This system uses T-DNA vectors that contain multimerized CaMV 35S transcriptional enhancers positioned near the right T-DNA border [59]. A transposon-mediated activation-tagging system has also been developed, using the *Ds* element that carries the tetramerized CaMV 35S enhancer [60]. From the T-DNA activation-tagging pools of *Arabidopsis*, Weigel et al. [27] have characterized over 30 dominant mutants with various phenotypes. Analysis of a subset of the mutants has shown that the tagging vector causes over-expression of the gene immediately adjacent to the inserted enhancer.

Gain-of-function mutagenesis in rice by activation tagging requires a strong enhancer element. Our current research demonstrates that the CaMV 35S enhancer element has enhancer activity in rice cells (unpublished data). We have developed the binary vector, pGA2715, that carries the tetramerized 35S enhancers and the hygromycin-resistant gene. This vector is being used for generating activation-tagging pools for rice.

6. Entrapment tagging

An entrapment tagging system allows for monitoring gene activity by creating fusions between tagged genes and a reporter gene, such as β -glucuronidase (*GUS*) and green fluorescent protein (*GFP*). Insertion of the promoterless reporter will not only destroy normal gene function but also activate expression of the reporter gene (Fig. 1B). Three entrapment systems are available: enhancer, promoter, and gene trap [61]. The devices for entrapment can be transferred into plant cells as a part of T-DNA or transposons [32,61,62]. This approach has been successfully applied to genes that are difficult to identify by traditional methods. Some examples include regulatory sequences that drive reporter gene expression in nematode feeding structures [63]; molecular markers for embryogenesis [64]; or regulatory sequences that mediate guard-cell-specific expression [65].

Activation of a reporter gene in a promoter trap vector can be as high as 30% [32]. In rice, at least 5% of the T-DNAs and 10% of the transposed *Ds* elements become activated in various tissues, e.g. roots, leaves, flowers, and seeds [13,41]. If reporter gene activation by certain environmental conditions or by chemicals such as growth substances is included, total tagging efficiency is higher. Some of the tags have displayed tissue- or organ-specific reporter expression, while others have exhibited ubiquitous expression patterns. One must study whether these lines will manifest any mutant phenotypes in organs where the reporter has been activated, and whether the phenotypes co-segregate with the inserted reporter gene.

The *GUS* gene has frequently been used for gene trapping in plants because of the accurate detection of its gene product and tolerance for the N-terminal translational fusions in its enzyme activity. However, one problem with the *GUS* assay is the destructive nature of its staining and destaining procedures [66]. Non-invasive and non-destructive reporter genes, such as *GFP* or luciferase, have not yet been widely used for gene trapping in plants. Recently, a visualization system with a charge-coupled device camera, band-pass filters, and a light source was used to demonstrate that green fluorescence emitted from *GFP* could be visualized in the calli, dry seeds, roots, and seedlings of transgenic rice plants [67]. Such an efficient visualization system should facilitate use

of a non-destructive visual selection marker for entrapment tagging in rice. We are currently investigating the efficiency of GFP expression in rice plants transformed with the vector pGA2717 that carries the promoterless *GFP* next to the left T-DNA border (unpublished data).

A new genomic DNA-based signal sequence trap method, the signal-exon trap (SET), may help identify genes that encode secreted and membrane-bound proteins [68]. SET is based on the coupling of an exon trap to the translation of captured exons, thus allowing the screening of exon-encoded polypeptides for signal peptide function. This system may be helpful in the discovery of novel members of known secretory gene clusters as well as for other positional cloning approaches.

7. Prospects

As Syngenta and Myriad Genetics announced the completion of the rice genome sequence on January 26, 2001 [69], the International Rice Genome Sequencing Project (IRGSP) decided to produce a complete sequence of the rice genome in the shortest time possible. The working draft of the rice genome that had been achieved by Monsanto in April 2000 would be a tremendous help to this sequencing project, thereby accelerating its completion. Therefore, entire sequences of the rice genome should soon be publicly available. This will produce a number of genes of which their possible identification is simply theoretically. One of the most efficient means for obtaining information on the function of a gene is by creating a loss-of-function mutation and studying the phenotype of the resulting mutant. A large population of mutagenized rice plants would be invaluable for assigning function to their uncharacterized genes.

With the increasing availability of rice genome sequences from public databases, it is worthwhile to construct a database for the flanking sequences. The introduction of effective methods for analyzing sequences adjoining the insertion would facilitate development of this database. These methods would also be extremely valuable for future gene discovery and functional genomics.

Acknowledgements

We thank Priscilla Licht for critical editing of the manuscript. The gene tagging project for functional genomics of rice in the An laboratory is supported, in part, by grants from the Korea Institute of Science and Technology Evaluation and Planning (KISTEP) and from Novartis.

References

- [1] Y. Harushima, M. Yano, A. Shomura, M. Sato, T. Shimano, Y. Kuboki, T. Yamamoto, S.Y. Lin, B.A. Antonio, A. Parco, H. Kajiya, N. Huang, K. Yamamoto, Y. Nagamura, N. Kurata, G.S. Khush, T. Sasaki, A high-density rice genetic linkage map with 2275 markers using a single F2 population, *Genetics* 148 (1998) 479–494.
- [2] K. Sakata, B.A. Antonio, Y. Mukai, H. Nagasaki, Y. Sakai, K. Makino, T. Sasaki, INE: a rice genome database with an integrated map view, *Nucleic Acids Res.* 28 (2000) 97–101.
- [3] S. Temnykh, W.D. Park, N. Ayres, S. Cartinhour, N. Hauck, L. Lipovich, Y.G. Cho, T. Ishii, S.R. McCouch, Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.), *Theor. Appl. Genet.* 100 (2000) 697–712.
- [4] Q. Yuan, F. Liang, J. Hsiao, V. Zismann, M.-I. Benito, J. Quackenbush, R. Wing, R. Buell, Anchoring of rice BAC clones to the rice genetic map *in silico*, *Nucleic Acids Res.* 28 (2000) 3636–3641.
- [5] A.K. Tyagi, A. Mohanty, Rice transformation for crop improvement and functional genomics, *Plant Sci.* 158 (2000) 1–18.
- [6] K. Arumuganathan, E.D. Earle, Nuclear DNA contents of some important plant species, *Plant Mol. Biol. Rep.* 9 (1991) 208–218.
- [7] C.C. David, The world rice economy: challenges ahead, in: G.S. Khush, G.A. Toenniessen (Eds.), *Rice Biotechnology*, International Rice Research Institute, 1991, pp. 1–18.
- [8] M.D. Gale, K.M. Devos, Comparative genetics in the grasses, *Proc. Natl. Acad. Sci. USA* 95 (1998) 1971–1974.
- [9] J. Kyoizuka, S. Konishi, K. Nemoto, T. Izawa, K. Shimamoto, Down-regulation of *RFL* the *FLO/LFY* homolog of rice accompanied with panicle branch initiation, *Proc. Natl. Acad. Sci. USA* 95 (1998) 1979–1982.
- [10] K.M. Devos, J. Beales, Y. Nagamura, T. Sasaki, *Arabidopsis*-rice: will colinearity allow gene prediction across the eudicot–monocot divide?, *Genome Res.* 9 (1999) 825–829.
- [11] R. Azpiroz-Leehan, K.A. Feldmann, T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth, *Trends Genet.* 13 (1997) 152–156.
- [12] P.J. Krysan, J.C. Young, M.R. Sussman, T-DNA as an insertional mutagen in *Arabidopsis*, *Plant Cell* 11 (1999) 2283–2290.

- [13] J.-S. Jeon, S. Lee, K.-H. Jung, S.-H. Jun, D.-H. Jeong, J. Lee, C. Kim, S. Jang, K. Yang, J. Nam, K. An, M.-J. Han, R.-J. Sung, H.-S. Choi, J.-H. Yu, J.-H. Choi, S.-Y. Cho, S.-S. Cha, S.-I. Kim, G. An, T-DNA insertional mutagenesis for functional genomics in rice, *Plant J.* 22 (2000) 561–570.
- [14] J.-S. Jeon, S. Lee, K.-H. Jung, S.-H. Jun, C. Kim, G. An, Tissue-preferential expression of a rice (α -tubulin gene, *OsTubA1*, mediated by the first intron, *Plant Physiol.* 123 (2000) 1005–1014.
- [15] A. Barakat, N. Carels, G. Bernardi, The distribution of genes in the genomes of Gramineae, *Proc. Natl. Acad. Sci. USA* 94 (1997) 6857–6861.
- [16] A. Barakat, P. Gallois, M. Raynal, D. Mestre-Ortega, C. Sallaud, E. Guiderdoni, M. Delseny, G. Bernardi, The distribution of T-DNA in the genomes of transgenic *Arabidopsis* and rice, *FEBS Lett.* 471 (2000) 161–164.
- [17] T. Sasaki, The rice genome project in Japan, *Proc. Natl. Acad. Sci. USA* 95 (1998) 2027–2028.
- [18] R. Tarchini, P. Biddle, R. Wineland, S. Tingey, A. Rafalski, The complete sequence of 340 kb of DNA around the rice *Adh1–Adh2* region reveals interrupted colinearity with maize chromosome 4, *Plant Cell* 12 (2000) 381–391.
- [19] T. Sasaki, Rice genome research: putting the pieces together, 6th International Congress of Plant Molecular Biology, 2000, pp. S1–74.
- [20] H. Nagano, L. Wu, S. Kawasaki, Y. Kishima, Y. Sano, Genomic organization of the 260 kb surrounding the *waxy* locus in a japonica rice, *Genome* 42 (1999) 1121–1126.
- [21] Z. Yin, G.-L. Wang, Evidence of multiple complex patterns of T-DNA integration into rice genome, *Theor. Appl. Genet.* 100 (2000) 461–470.
- [22] Y.-G. Liu, R. Whittier, Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking, *Genomics* 25 (1995) 674–681.
- [23] S. Balzergue, B. Dubreucq, S. Chauvin, I. Le-Clainche, F. Le Boulair, R. de Rose, F. Samson, V. Biauudet, A. Lechary, C. Cruaud, J. Weissenbach, M. Caboche, L. Lepiniec, Improved PCR-walking for large-scale isolation of plant T-DNA borders, *Biotechniques* 30 (2001) 496–504.
- [24] T. Triglia, M.G. Peterson, D.J. Kemp, A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences, *Nucleic Acids Res.* 16 (1988) 8186.
- [25] S. Hanley, D. Edwards, D. Stevenson, S. Haines, M. Hegarty, W. Schuch, K.J. Edwards, Identification of transposon-tagged genes by the random sequencing of mutator-tagged DNA fragments from *Zea mays*, *Plant J.* 23 (2000) 557–566.
- [26] V. Walbot, Saturation mutagenesis using maize transposons, *Curr. Opin. Plant Biol.* 3 (2000) 103–107.
- [27] D. Weigel, J.H. Ahn, M.A. Blazquez, J.O. Borevitz, S.K. Christensen, C. Fankhauser, C. Ferrandiz, I. Kardailsky, E.J. Malancharuvil, M.M. Neff, J.T. Nguyen, S. Sato, Z.Y. Wang, Y. Xia, R.A. Dixon, M.J. Harrison, C.J. Lamb, M.F. Yanofsky, J. Chory, Activation tagging in *Arabidopsis*, *Plant Physiol.* 122 (2000) 1003–1013.
- [28] Y. Sato, N. Sentoku, Y. Miura, H. Hirochika, H. Kitano, M. Matsuoka, Loss-of-function mutations in the rice homeobox gene *OSH15* affect the architecture of internodes resulting in dwarf plants, *EMBO J.* 18 (1999) 992–1002.
- [29] S. Parinov, M. Sevugan, Y. De, W.C. Yang, M. Kumaran, V. Sundaresan, Analysis of flanking sequences from dissociation insertion lines. A database for reverse genetics in *Arabidopsis*, *Plant Cell* 11 (1999) 2263–2270.
- [30] H. Hirochika, K. Sugimoto, Y. Otsuki, H. Tsugawa, M. Kanda, Retrotransposons of rice involved in mutations induced by tissue culture, *Proc. Natl. Acad. Sci. USA* 93 (1996) 7783–7788.
- [31] N.V. Fedoroff, S. Wessler, M. Shure, Isolation of the transposable maize controlling elements *Ac* and *Ds*, *Cell* 35 (1983) 235–242.
- [32] V. Sundaresan, Horizontal spread of transposon mutagenesis: new uses of old elements, *Trends Plant Sci.* 1 (1996) 184–191.
- [33] N.V. Fedoroff, D.B. Furtek, O.E. Nelson, Cloning of the *bronze* locus in maize by a simple and generalizable procedure using the transposable element Activator (*Ac*), *Proc. Natl. Acad. Sci. USA* 81 (1984) 3825–3829.
- [34] V. Walbot, Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis, *Annu. Rev. Plant Physiol.* 43 (1992) 49–82.
- [35] T. Izawa, C. Miyazaki, M. Yamamoto, R. Terada, S. Iida, K. Shimamoto, Introduction and transposition of the maize transposable element *Ac* in rice (*Oryza sativa* L.), *Mol. Gen. Genet.* 227 (1991) 391–396.
- [36] N. Murai, Z. Li, Y. Kawagoe, A. Hayashimamoto, Transposition of the maize Activator element in transgenic rice plants, *Nucleic Acids Res.* 19 (1991) 617–622.
- [37] H. Enoki, T. Izawa, M. Kawahara, M. Komatsu, S. Koh, J. Kyozuka, K. Shimamoto, *Ac* as a tool for the functional genomics for rice, *Plant J.* 19 (1999) 605–613.
- [38] K. Shimamoto, C. Miyazaki, H. Hashimoto, T. Izawa, K. Itoh, R. Terada, Y. Inagaki, S. Iida, Trans-activation and stable integration of the maize transposable element *Ds* cotransfected with the *Ac* transposase gene in transgenic rice plants, *Mol. Gen. Genet.* 239 (1993) 354–360.
- [39] K. Sugimoto, Y. Otsuki, S. Saji, H. Hirochika, Transposition of the maize *Ds* element from a viral vector to the rice genome, *Plant J.* 5 (1994) 863–871.
- [40] T. Izawa, T. Ohnishi, T. Nakano, N. Ishida, H. Enoki, H. Hashimoto, K. Itoh, R. Terada, C. Wu, C. Miyazaki, T. Endo, S. Iida, K. Shimamoto, Transposon tagging in rice, *Plant Mol. Biol.* 35 (1997) 219–229.
- [41] H.G. Chin, M.S. Choe, S.H. Lee, S.H. Park, J.C. Koo, N.Y. Kim, J.J. Lee, B.G. Oh, G.H. Yi, S.C. Kim, H.C. Choi, M.J. Cho, C. Han, Molecular analysis of rice plants harboring an *Ac/Ds* transposable element-mediated gene trapping system, *Plant J.* 19 (1999) 615–623.
- [42] Y. Nakagawa, C. Machida, Y. Machida, K. Toriyama, Frequency and pattern of transposition of the maize transposable element *Ds* in transgenic rice plants, *Plant Cell Physiol.* 41 (2000) 733–742.
- [43] S. Dai, P. Zheng, P. Marmey, S. Zhang, W. Tian, S. Chen, R.N. Beachy, C. Fauquet, Comparative analysis of transgenic rice plants obtained by *Agrobacterium*-mediated transformation and particle bombardment, *Mol. Breed.* 7 (2001) 25–33.

- [44] C. Machida, H. Onouchi, J. Koizumi, S. Hamada, E. Semiarti, S. Torikai, Y. Machida, Characterization of the transposition pattern of the *Ac* element in *Arabidopsis thaliana* using endonuclease I-*Sce*I, Proc. Natl. Acad. Sci. USA 94 (1997) 8675–8680.
- [45] E. Wisman, G.H. Cardon, P. Fransz, H. Saedler, The behaviour of the autonomous maize transposable element *En/Spm* in *Arabidopsis thaliana* allows efficient mutagenesis, Plant Mol. Biol. 37 (1998) 989–999.
- [46] S. Phogat, P.K. Burma, D. Pental, A four-element based transposon system for allele specific tagging in plants — theoretical considerations, J. Biosci. 25 (2000) 57–63.
- [47] G. Schmitz, K. Theres, A self-stabilizing *Ac* derivative and its potential for transposon tagging, Plant J. 6 (1994) 781–786.
- [48] Y.-C. Charng, A.J.P. Pfitzner, U.M. Pfitzer, K.-F. Charng-Chang, C. Chen, J. Tu, T.-T. Kuo, Construction of an inducible transposon, *INAc*, to develop a gene tagging system in higher plants, Mol. Breed. 6 (2000) 353–367.
- [49] H. Hirochika, H. Otsuki, M. Yoshikawa, Y. Otsuki, K. Sugimoto, S. Takeda, Autonomous transposition of the tobacco retrotransposon *Tto1* in rice, Plant Cell 8 (1996) 725–734.
- [50] D. Liu, S. Zhang, C. Fauquet, N.M. Crawford, The *Arabidopsis* transposon *Tag1* is active in rice, undergoing germinal transposition and restricted, late somatic excision, Mol. Gen. Genet. 262 (1999) 413–420.
- [51] L. Mao, T.C. Wood, Y. Yu, M.A. Budiman, J. Tomkins, S. Woo, M. Sasinowski, G. Presting, D. Frisch, S. Goff, R.A. Dean, R.A. Wing, Rice transposable elements: a survey of 73 000 sequence-tagged-connectors, Genome Res. 10 (2000) 982–990.
- [52] S. Takeda, K. Sugimoto, H. Otsuki, H. Hirochika, Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate, Plant Mol. Biol. 36 (1998) 365–376.
- [53] M. Takano, H. Kanegae, T. Shinomura, A. Miyao, H. Hirochika, M. Furuya, Isolation and characterization of rice phytochrome A mutants, Plant Cell 13 (2001) 521–534.
- [54] G.K. Agrawal, M. Yamazaki, M. Kobayashi, R. Hirochika, A. Miyao, H. Hirochika, Screening of the rice viviparous mutants generated by endogenous retrotransposon *Tos17* insertion. Tagging of a zeaxanthin epoxidase gene and a novel *OsTATC* gene, Plant Physiol. 125 (2001) 1248–1257.
- [55] A. Miyao, K. Murata, K. Tanaka, K. Onosato, Y. Yamashita, H. Hirochika, Profile of retrotransposon *Tos17* insertions in rice genome sequence, *OsCesA7* disrupted by insertion of the retrotransposon *Tos17*, 9th Conference of Plant and Animal Genome, 2001, pp. P5b–1.
- [56] I. Kardailsky, V.K. Shukla, J.H. Ahn, N. Dagenais, S.K. Christensen, J.T. Nguyen, J. Chory, M.J. Harrison, D. Weigel, Activation tagging of the floral inducer FT, Science 286 (1999) 1962–1965.
- [57] T. Ito, E.M. Meyerowitz, Overexpression of a gene encoding a cytochrome P450, *CYP78A9*, induces large and seedless fruit in *Arabidopsis*, Plant Cell 12 (2000) 1541–1550.
- [58] H. Lee, S.S. Suh, E. Park, E. Cho, J.H. Ahn, S.G. Kim, J.S. Lee, Y.M. Kwon, I. Lee, The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*, Genes Dev. 14 (2000) 2366–2376.
- [59] H. Hayashi, I. Czaja, H. Lubenow, J. Schell, R. Walden, Activation of a plant gene by T-DNA tagging: auxin-independent growth in vitro, Science 258 (1992) 1350–1353.
- [60] M. Mori, K. Sugimoto, K. Satoh, K. Okabe, S. Toki, M. Ugaki, H. Hirochika, S. Kikuchi, Toward activation tagging in rice, 6th International Congress of Plant Molecular Biology, 2000, pp. S1–S57.
- [61] P.S. Springer, Gene traps: tools for plant development and genomics, Plant Cell 12 (2000) 1007–1020.
- [62] R.A. Martienssen, Functional genomics: probing plant gene function and expression with transposons, Proc. Natl. Acad. Sci. USA 95 (1998) 2021–2026.
- [63] N. Barthels, F.M. van der Lee, J. Klap, O.J. Goddijn, M. Karimi, P. Puzio, F.M. Grundler, S.A. Ohl, K. Lindsey, L. Robertson, W.M. Robertson, M. van Montagu, G. Gheysen, P.C. Sijmons, Regulatory sequences of *Arabidopsis* drive reporter gene expression in nematode feeding structures, Plant Cell 9 (1997) 2119–2134.
- [64] J.F. Topping, K. Lindsey, Promoter trap markers differentiate structural and positional components of polar development in *Arabidopsis*, Plant Cell 9 (1997) 1713–1725.
- [65] G. Plesch, E. Kamann, B. Mueller-Roeber, Cloning of regulatory sequences mediating guard-cell-specific gene expression, Gene 249 (2000) 83–89.
- [66] R.A. Jefferson, T.A. Kavanagh, M.W. Bevan, GUS fusions: β -glucuronidase as a sensitive and versatile gene fusion marker in higher plants, EMBO J. 6 (1987) 3901–3907.
- [67] B.C. Chung, J.K. Kim, B.H. Nahm, C.H. Lee, In planta visual monitoring of green fluorescent protein in transgenic rice plants, Mol. Cells 10 (2000) 411–414.
- [68] M. Peterfy, T. Gyuris, L. Takacs, Signal-exon trap: a novel method for the identification of signal sequences from genomic DNA, Nucleic Acids Res. 28 (2000) E26.
- [69] R.J. Davenport, Rice genome: syngenta finishes, consortium goes on, Science 291 (2001) 807.