FUNCTIONAL GENOMICS STRATEGIES WITH TRANSPOSONS IN RICE

RAFFAELLA GRECO

Promotor:	Prof. dr. ir. M. Koornneef, hoogleraar in de Erfelijkheidsleer, Wageningen Universiteit
Co-promotor:	Dr. Andy Pereira, clusterleider Genoomanalyse, BU Bioscience, Plant Research International, Wageningen UR
Promotiecommissie:	Prof. dr. A.G.M. Gerats, Katholieke Universiteit Nijmegen
	Dr. E. Guiderdoni, CIRAD, Montpellier, France
	Prof. dr. ir. E. Jacobsen, Wageningen Universiteit
	Prof. dr. W.J. Stiekema, Wageningen Universiteit

FUNCTIONAL GENOMICS STRATEGIES WITH TRANSPOSONS IN RICE

RAFFAELLA GRECO

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. dr. ir. L. Speelman,
in het openbaar te verdedigen
op woensdag 5 november 2003
des namiddags te 13.30 uur in de Aula.

R. Greco – Functional genomics strategies with transposons in rice - 2003

Ph.D. Thesis Wageningen University, Wageningen, The Netherlands - 182 p.

ISBN: 90-5808-916-9

Table of Contents

Al mio papa' Benito
Alla mia mamma Rosanna

Chapter 1

General introduction and scope of the thesis.

GENERAL INTRODUCTION

History of rice

Cultivation of rice dates back to the earliest origins of human civilisation. Rice was probably the first cultivated crop and the staple food in Asia, as reflected by its intimate involvement in the culture and religious habits of many Asian countries, where the word “rice” is often synonymous of food or agriculture. Indeed, rice was known in ancient India as “dhanya” or “sustainer of the human race”.

Nowadays, rice is the third largest crop behind maize and wheat but the most important crop for human nutrition, being the staple food of more than half of the world population. Most of the rice consumption is still in Asia, where it supplies up to 70% of the total daily calories in countries like Laos, Cambodia, Myanmar and Bangladesh, with an average of 30% over the whole continent (FAO database 2003, Rome). In Africa as well, countries such as Guinea Bissau, Madagascar and Sierra Leone obtain more than 40% of their calories by rice consumption.

The genus *Oryza* comprises about 24 species spread over the tropical and sub-tropical regions. Nine different genomes were recognised in the genus, with the majority of the species being diploids with few allotetraploids. Cultivated rice belongs to two species, *Oryza sativa* and *Oryza glaberrima*, of which the former is by far the more widely utilized, while the latter is mainly confined to African countries. *O. sativa* comprises *indica* long-grain rice, providing most of the world’s rice, and *japonica* short-grain rice preferred in Japan and some other countries. The *japonica* type consists of tropical (*javanica*) and temperate rice varieties. Variability exists in the fertility of F_1 progenies derived from crosses between *indica* and *japonica* subspecies, ranging from several percent to fully fertile.

When and where rice originated is not known with certainty. While the genus *Oryza* probably emerged at least 130 million years ago in Gondwanaland and successively spread with the drift of the continents (Chang 1976), it is believed that domestication of rice occurred some 10,000 years ago and that the two cultivated species *O. sativa* and *O. glaberrima* originated independently in Asia and Africa from the wild ancestors *Oryza rufipogon* and *Oryza barthii*, respectively (reviewed in Khush 1997). Moreover, domestication of *O. sativa* was also not a monophyletic event, rather *indica* and *japonica* types were separately domesticated from *O. rufipogon* progenitors in Southeast Asia and China. Divergence between the South Asian and Chinese forms of *O. rufipogon* is believed to have commenced more than one million years ago.

There are about 120,000 wild or cultivated varieties of rice in the world, which differ in many morphological and physiological traits. The high degree of divergence among species but also within species, such as between the *indica* and *japonica* subspecies and among the *indica* varieties as well, provides a valuable source of phenotypic and genetic variation that contributed

much to the great achievements in rice genetics and breeding of the past decades. The identification of polymorphism at DNA level and the development of molecular markers allowed the creation of detailed genetic linkage maps, extremely helpful in marker-assisted breeding. Moreover, integration between genetic and physical maps was essential for map-based cloning of important genes and QTLs, and could provide an important starting material for the genome sequencing programs.

Rice as a model cereal

Rice has become an established model system in the last decade and the organism of choice for the study of the cereal genomes. What makes rice an attractive biological system is certainly its small genome size and the extensive conservation of genes and gene order that it shares with the other cereals.

Rice is among the cereals, the species having one of the smallest genomes, with its 430 Mb compared to the 2500 Mb of the maize genome, 5000 Mb of barley or 16000 Mb of wheat, among others (Arumuganathan and Earle 1991). The great variation in genome size among plant species appears to be mainly attributable to differences in the content of repetitive DNA (reviewed in Bennetzen 2000a). Transposable elements, in particular retrotransposons, played a major role in genome expansion as they were found to compose up to 60% or more of the DNA content of species with large genomes such as maize, wheat and barley (Kumar and Bennetzen 2000). In spite of its smaller size, the rice genome as well contains about 45% of repetitive DNA and 20-25% of this repetitive fraction is made up of transposable elements (Goff et al. 2002; Yu et al. 2002).

Comparative genetic mapping among different cereals such as rice, maize, barley, sorghum, wheat, oat and foxtail millet, revealed a significant level of colinearity of molecular markers across regions of several megabases, which suggested that grasses could be considered and studied as a single genetic system (Bennetzen and Freeling 1993; Gale and Devos 1998; Devos and Gale 2000; Bennetzen 2000b; Freeling 2001). A so-called “Circle Diagram” model was proposed in which cereal genomes are represented as aligned circles, with rice being the inner circle from which the other genomes expanded essentially by increase in repetitive regions (Moore et al. 1995; Gale and Devos 1998). In view of these synthenic relations, rice research is believed to provide a significant contribution for the identification of orthologous genes in other cereals by comparative genetics studies.

Great advances have been made in the last decades in rice genetics that contributed greatly to its choice as a model system. Genetic linkage maps have been developed using several types of molecular markers such as RFLP, CAPS, AFLP or SSR (Causse et al. 1994; Cho

et al. 1998; Harushima et al. 1998), exploiting the high level of intra and inter-specific polymorphism existing between the several rice varieties. One of the major outcomes is the production of a linkage map containing 3,267 RFLP/CAPS markers scattered over all the 12 chromosomes, which to date represents the most detailed genetic map available for a given organism (Rice Genome Research Program 2000; http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html). In addition, about 2700 SSR markers were also developed and almost all anchored to the rice chromosomes (McCouch et al. 2002). Integrated maps of DNA and phenotypic markers have also been constructed (Yoshimura et al. 1997).

The availability of dense genetic maps, together with the development of large-insert clone (YAC, BAC, PAC) genomic libraries, provided important tools for the development of physical maps of the rice genome (Tao et al. 2001; Chen et al. 2002; Saji et al. 2001). Chen et al. (2002) integrated their BAC-based physical map with genetic markers, providing coverage of 90.6% of the genome. Moreover, a rice transcript map was produced in which more than 6500 ESTs deriving from 19 libraries from various tissues and treatments were mapped on YAC contigs covering 80.8% of the genome (Wu et al. 2002). Recently, a more extensive effort resulted in the mapping of approximately 27,000 full-length cDNA clones on the rice genomic sequences (Kikuchi et al. 2003). Genome-wide physical mapping has been of tremendous help for gene isolation by positional cloning and for genome sequencing, providing the possibility of identifying a minimal tiling path of genomic clones for each chromosome chosen for sequencing.

The establishment of efficient techniques for rice transformation was also an essential and relatively recent achievement in rice research. The first transgenic rice plants were obtained by direct DNA uptake from protoplasts (Toriyama et al. 1988). Shortly thereafter, transformation of several *indica* and *japonica* varieties by particle bombardment of immature embryos was reported by Christou et al. (1991). Biolistic techniques proved to work quite efficiently regardless of the genotype used, and are of particular utility for the transformation of recalcitrant species. However, they are still not devoid of all the disadvantages connected to direct DNA transformation, like multiple DNA integrations promoting rearrangements and transgene silencing (Dai et al. 2001). Genetic manipulation of rice was boosted by the establishment of Agrobacterium transformation as a routine protocol for the attainment of transgenic plants (Hiei et al. 1994, Sallaud et al. 2003), which allowed a more reliable analysis of transgene effects and opened the route to the development of large-scale (T-DNA and heterologous transposon) insertional mutagenesis for gene tagging.

Recent advances in rice genomics

All these features led rice to become the first monocot crop genome to be completely sequenced, and the second plant genome after Arabidopsis. Two draft sequences of the *japonica* Nipponbare cultivar were produced in the past years by Monsanto (Barry 2001) and Syngenta (Goff et al. 2002) that were made available to the research community under restricted conditions, while the draft sequence of the *indica* 93-11 variety was publicly released by the Beijing Genomic Institute (Yu et al. 2002). These drafts were obtained based on a whole-genome shot-gun sequencing approach that, although undoubtedly faster, leads to higher levels of inaccuracy and does not allow a direct anchoring of the sequenced clones to the rice chromosomes. Concurrently, the International Rice Genome Sequencing Project (IRGSP) decided to adopt a map-based clone-by-clone sequencing approach (after the definition of a minimum tiling path covering each chromosome), which resulted in the recent completion of a high-quality draft sequence of the Nipponbare genome (http://rgp.dna.affrc.go.jp/rgp/Dec18_NEWS.html) and the release of finished-quality sequence of chromosomes 1, 4 and 10 (Feng et al. 2002; Sasaki et al. 2002; The Rice Chromosome 10 Sequencing Consortium 2003).

Preliminary estimates revealed a surprisingly high number of predicted genes, about 62000, much higher than Arabidopsis and any other organism sequenced up to date. An interesting outcome from the data obtained is that, while more than 80% of the Arabidopsis predicted ORFs have homologues in rice, a significant set of rice genes do not have a homologous partner in Arabidopsis and can therefore be regarded, if prediction is confirmed, as grass-specific genes. This fraction accounts for about half of the overall predicted genes whose function disclosure can give insights into the biology and evolution of the grasses. On the other hand, the functional genomics resources available for Arabidopsis can also be used to gather information about the function of the homologues rice genes. Comparison between these will in turn shed light into the level of conservation of gene function during evolution of monocot and dicot plant species. Moreover, about 35% of the conserved genes between rice and Arabidopsis are not found in any other sequenced organisms and are likely to represent plant-specific genes. Yet, there is a considerable percentage of predicted genes annotated as “unknown” or “hypothetical” whose function is completely obscure. It has however to be considered that annotation is not complete and annotation tools are still under improvement, so a considerable proportion of genes may have been wrongly predicted.

Functional redundancy is expected in rice as well as in Arabidopsis because of the large extent of gene duplications demonstrated to have occurred in the rice genome (Yu et al. 2002). Amplified gene copies account for about 25% of all rice genes according to preliminary estimates, compared to 17% in Arabidopsis (Song et al. 2002; The Rice Chromosome 10 Sequencing Consortium 2003). Rice is indeed believed to be an ancient polyploid, in which a genome

duplication event may have occurred about 40-50 million years ago (Goff et al. 2002). Nevertheless, lack of large-scale duplication events evident in rice, in contrast to Arabidopsis, will undoubtedly facilitate the comparison among cereal genomes.

Comparisons of rice genomic sequences with available sequences of other cereals confirmed the existence of macro-colinearity deduced by recombinatorial mapping. However, it also revealed that colinearity at gene level is often disrupted due to small rearrangements such as deletions, insertions, duplications or inversions (Tarchini et al. 2000; Dubcovsky et al. 2001; SanMiguel 2002; Song et al. 2002; reviewed in Bennetzen and Ma 2003). On the other hand, synthenic relations between rice and Arabidopsis seem to be confined to the existence of several regions of micro-colinearity, while extensive colinearity at the genome level is lacking (Devos et al. 1999; van Dodeweerd et al. 1999; Liu et al. 2001; Mayer et al. 2001; Salse et al. 2002; Vandepoele et al. 2002; Goff et al. 2002).

These findings reinforce the idea that information gathered solely from the study of the Arabidopsis genome would definitely have a smaller biological impact on the understanding of the more complex cereal crop species and largely justifies the parallel investigation of the more similar rice genome.

Rice functional genomics

Gene expression analysis

Random EST sequencing provides important information about the actual number of genes that are expressed. Large-scale EST sequencing projects have been undertaken by several labs using different rice varieties (Bohnert et al. 2001; Reddy et al. 2002; Wu et al. 2002; Yamamoto and Sasaki 1997; Beijing Genomics Institute at http://btn.genomics.org.cn/rice/index.php). At present, more than 200,000 ESTs have been released in public databases, accounting for about 60,000 unique expressed sequences according to the TIGR Rice Gene Index (Quackenbush et al. 2001). Ongoing efforts for the creation of a database of full-length cDNA clones identified already more than 28,000 unique clones defining around 19,000-20,500 transcription units in the rice genome and will offer a significant contribution for genome annotation (Kikuchi et al. 2003).

Concurrently, microarray analyses to monitor global changes in gene expression in response to biotic (such as pathogens) and abiotic stresses (such as drought, cold or salinity), as well as developmental cues are being carried out. Both oligo-chips (Syngenta: rice GeneChip microarray containing 500,000 oligonucleotide probes for 51,000 rice genes) and cDNA microarrays (Genoplante, France: planned 12,000 unique genes, Chandelier et al. 2003; NIAS,

Japan: microarray of 9,000 stress-related genes at http://www.nias.affrc.go.jp/index_e.html, Yazaki et al. 2000, 2002) are being used.

In parallel, proteomics projects are also undertaken to complement the information gathered at mRNA level with protein expression data (Kerim et al. 2003; Komatsu et al. 2003).

Mutant analysis

The analysis of mutant phenotypes resulting from gene inactivation is an indispensable tool for the study of gene function.

Before the advent of the genomics era, gene function identification was essentially performed through a forward genetics strategy. This implies a prior identification of a mutant of interest, leading eventually to the isolation of the gene whose function is affected. Mutant phenotypes can occur naturally or be induced by several mutagenic agents that include chemical (such as EMS or diepoxybutane), physical (such as gamma-rays or fast neutrons) and biological agents (insertion elements such as T-DNA or transposons). Generally the use of biological mutagens, such as insertion elements, facilitates the process that leads from the identification of a mutant phenotype to the cloning of the affected gene, as the insertion sequence can be used as a tag. In all the other cases, gene cloning has to be carried out essentially by map-based approaches.

The availability of dense genetic maps and their integration with physical maps greatly facilitated map-based gene isolation in rice. Fine mapping of genes and QTLs responsible for several agronomic traits, such as resistance to pests and diseases, drought and submergence tolerance, plant architecture or heading date, allowed in several cases their isolation by positional cloning. This is the case of the *Xa21* and *Xa1* genes conferring resistance to bacterial blight (Song et al. 1995; Yoshimura et al. 1998), the *Pib* and *Pi-ta* genes controlling blast resistance (Wang et al. 1999; Bryan et al. 2000), the dwarf mutant gene *Dwarf1* (Ashikari et al. 1999), the *MOC1* gene controlling tillering (Li et al. 2003) and three QTLs controlling flowering time (Yano et al. 2000; Takahashi et al. 2001; Kojima et al. 2002). However, this procedure may still require time and labour.

Isolation of rice genes by “candidate gene” approach based on homology to genes with known function in other species or functional characterisation by ectopic over-expression or loss-of-function were also reported, which led to the identification of genes controlling flower development and timing, cell death and cold/salt tolerance (i.e. Izawa et al. 2000; Jeon et al. 2000; Kang et al. 1998; Kawasaki et al. 1999; Saijo et al. 2000).

The achievement of the genome sequence provides enormous resources for the discovery of rice genes. On one hand, it facilitates the development of new molecular markers for breeding

(marker-assisted selection) and more detailed integrated genetic and physical maps for map-based cloning, as well as comparative mapping and candidate cloning in other cereal species. On the other hand, it challenges the community for the development of efficient high-throughput methods for gene function identification. As seen for Arabidopsis, the advent of the genomics technologies also changed the approach from forward to reverse genetics, that is from “mutant to gene” to “sequenced gene to function”.

Gene disruption could conceptually be achieved by targeted gene replacement via homologous recombination of the genes identified by sequencing, a well-established technique for functional studies in yeast and mouse that has also proven to effectively work in moss (reviewed by Hanin and Paszkowski 2003). This approach is however not yet feasible in higher plants, as the efficiency at which homologous recombination occurs is very low and not suitable for large-scale gene disruption. A recent report about the successful targeting of the *waxy* gene in rice (Terada et al. 2002) suggests that experimental designs to optimize the parameters that influence the efficiency of this process in plants may in the near future boost the use of this technology. Gene inactivation via classical anti-sense or co-suppression approaches as well does not seem feasible for high-throughput functional analysis. However, massive silencing of sequenced genes is being undertaken based on RNA interference using vectors designed for high-throughput cloning, procedure that may be particularly useful for the functional analysis of redundant genes (Helliwell et al. 2002, Shimamoto et al. 2002)

At present, the generation of mutants by chemical and physical agents (Leung 2003) but especially by random insertional mutagenesis using T-DNA or transposons seems to be the most efficient strategy for reverse genetics in rice (Jeon and An 2001).

Insertional mutagenesis

Large mutant collections are being generated in rice using T-DNA, endogenous transposons (mainly the tissue-culture activated retrotransposon *Tos17*) and heterologous transposons (based on the maize *Ac/Ds* system; reviewed by Jeon and An 2001; see Chapter 9 for an overview of the main resources available).

Heterologous transposon tagging can be more advantageous as it offers the opportunity to regulate transposition by engineering the transposon components and introducing selectable markers. An optimal scheme usually employs a two-component system, in which a defective transposon is mobilised by a transposase source supplied either in *cis* (on the same T-DNA) or in *trans* (on a different T-DNA introduced by co-transformation or by crossing). The transposase itself consists of an immobilised autonomous element, which is deprived of the *cis*-requirements for transposition and is usually driven by a constitutive promoter. In more sophisticated systems,

the transposase can be put under control of inducible promoters to achieve spatial and temporal regulated transposition. Markers to monitor excision and select for independent transposition events are incorporated in the construct to maximise their efficiency.

Insertion tags can be designed for knockout mutagenesis, gene detection and activation tagging. Knockout mutagenesis is based on the expectation that the insertion of a transposon or a T-DNA in a gene will disrupt its function and produce a mutant phenotype. However, functional redundancy, as well as multiple roles a gene may play in different developmental stages, or gametophytic and embryonic lethality will decrease the chance to observe a loss-of-function phenotype with knockout approaches. In Arabidopsis, less than 5% of the genes are estimated to reveal a visible phenotype upon disruption (Long et al. 1993) and the rice genome is expected to reveal a similar behaviour. Addition of reporter genes inside the insertion elements that can "trap" and display the pattern of expression of neighbouring tagged host genes can complement knockout mutagenesis and help unravelling their function (gene detection; reviewed by Springer 2000). Usually, a GUS reporter gene is utilised to function as gene, promoter or enhancer trap, but recently developed systems employ non-destructive assays based on markers such as GFP and luciferase (Johnson et al. 2003; Yamamoto et al. 2003). Similarly, the placement of strong enhancer sequences next to one of the borders of the insertion tag will cause overexpression of adjacent genes, providing a very efficient way to reveal novel phenotypes, which proved to be extremely powerful when studying genes with overlapping function like those belonging to gene families.

In general, the use of T-DNA has the advantage of directly yielding transformed plants with stable insertions, while transposons need to be stabilised by eliminating the transposase source (in case of two-component systems). On the other hand, multiple integration causing rearrangements at the borders often make the isolation of genes tagged by T-DNA inserts a difficult task. In case of gene detection, these complex integrations can also alter the pattern of expression of the reporter gene leading to mis-interpretation of the results. Moreover, as large collections of T-DNA inserts require the generation of high numbers of independent transformants that in rice are essentially obtained through tissue culture, a high degree of somaclonal variation is expected to increase the number of untagged mutations. In contrast, new transposon insertions can be generated by crossing or by simple propagation of lines, without need of tissue culture manipulations. Since transposons propagate in the genome as single clean inserts, they allow easy isolation of flanking sequences. The possibility to re-mobilize a transposon from a tagged gene, causing reversion of the mutant phenotype or a weaker mutant allele, provides a direct confirmation that the gene was indeed tagged. Furthermore, the property of certain transposon systems like the maize *Ac/Ds* to reinsert at linked genomic position can also be exploited for targeted tagging of genes of interest by re-mobilization of nearby transposons.

Once a T-DNA- or a transposon-mutagenised population has been generated, identification of a plant containing an insertion in a gene of interest can be carried out. With the availability of the genome sequence, large-scale reverse genetics strategies are now essentially oriented towards the random isolation of T-DNA or transposon flanking sequences to generate libraries that could then be screened *in silico* for insertions in any given gene. Examples are coming from Arabidopsis, for which several insertional mutant collections have been established and databases of flanking sequences can be electronically screened (Parinov and Sundaresan 2000; The Arabidopsis Information Resource at http://www.arabidopsis.org). The success of this strategy relies on the efficient high-throughput recovery of the genomic sequences flanking the insertions. Several techniques have been developed that could effectively be employed for this purpose, such as TAIL-PCR (Liu end Whittier 1995), inverse-PCR (Earp et al. 1990), adapter-ligation PCR (Siebert et al. 1995; Devic et al. 1997) or AFLP-based methods such as Transposon Insertion Display (Van den Broeck et al. 1998; Tissier et al. 1999).

Alternatively, reverse genetics screenings can be carried out based on PCR on pooled DNA from several insertion lines, using a gene specific primer in combination with a primer specific for the insertion tag. Particularly efficient is the use of bi- or three-dimensional pooling schemes that allow efficient screening of large number of plants in a relative lower number of reactions (i.e. Koes et al. 1995; Tissier et al. 1999).

SCOPE OF THE THESIS

The scope of the research work described in this thesis was to establish a set of “starter” transposon lines as a foundation for the further development of saturated tagging populations for the functional analysis of the rice genome. The choice to employ the heterologous *Ac/Ds* and *En/Spm* maize transposon systems rests in the notion of their effectiveness in other plant systems like Arabidopsis and in the early studies on their transposition in rice, limited however to the *Ac/Ds* system. The use of different transposon systems will assure a reduced insertion bias, as transposition to closely linked sites typical for *Ac/Ds* in maize was confirmed also in other heterologous systems such as Arabidopsis (Bancroft and Dean 1993; Keller et al. 1993), while more scattered transpositions were revealed for *En/I* (Aarts et a. 1995).

At the time this research work started, reports on transposon mutagenesis in rice were limited to the attainment of transgenic plants by direct DNA transformation containing either the autonomous *Ac* element or a defective *Ds* that could be activated by a transposase source supplied in trans (Izawa et al. 1991; Murai et al. 1991; Shimamoto et al. 1993; Izawa et al. 1997). This transposase source consisted of the autonomous *Ac* element itself or a truncated *Ac* version

in which transposase expression was driven by the constitutive CaMV 35S promoter. In particular, *Ac* was demonstrated to transpose stably in subsequent generations with high frequency of germinal transposition, while preliminary reports on the behaviour of two component *Ac/Ds* systems were revealing a tendency of the *Ds* transposon to undergo inactivation in later generations. However, at the time this work was initiated, reports on a very limited number of subsequent generations were available, and efforts directed towards large-scale production of *Ds* transposon lines were beginning to be carried out. On the other hand, no reports about the use of the *En/Spm* transposon system in rice were available at that time.

In this work, several transposon constructs were developed and transformed into rice cultivars essentially by Agrobacterium-mediated transformation or particle bombardment in case of *indica* varieties. Constructs employing the autonomous *Ac* transposon were developed for knockout mutagenesis and used for transformation of *japonica* (**Chapter 3**) and *indica* (**Chapter 4**) rice cultivars. Their transposition behaviour is depicted and discussed in terms of usefulness for transposon tagging. As previously mentioned, gene redundancy in rice is also expected to reduce the efficacy at which knockout lines will display mutant phenotypes. The advantages of integrating knockout mutagenesis with gene detection are discussed in **Chapter 2** and in **Chapter 5**. Gene detection constructs were mainly developed in the form of enhancer traps using two-component *Ac/Ds* and *En/I* transposon systems and their use is discussed in **Chapter 6** and **Chapter 7**, respectively. Transposition of the *En/Spm* transposon system in rice was demonstrated to occur at low efficiency. Among the possible causes, interaction between the exogenous maize transposon and endogenous related elements, demonstrated to be present in the rice genome in a relatively high abundance, was investigated (**Chapter 8**). During the course of this research, several reports on the use of similar *Ac* and *Ac/Ds* tagging systems in rice by other groups have been published. A description of these systems and their relative advantages and disadvantages in comparison to the work reported here are discussed in the concerned chapters.

LITERATURE CITED

Aarts MGM, Corzaan P, Stiekema WJ, Pereira A (1995) A two-element *enhancer-inhibitor* transposon system in *Arabidopsis thaliana*. Mol Gen Genet 247: 555-564

Arumuganathan K, Earle E (1991) Nuclear DNA content of some important plant species. Plan Mol Biol Rep 9: 208-218

Ashikari M, Wu JZ, Yano M, Sasaki T, Yoshimura A (1999) Rice gibberellin-insensitive dwarf mutant gene *Dwarf 1* encodes the alpha-subunit of GTP-binding protein. Proc Natl Acad Sci 96: 10284-10289

Bancroft I, Dean C (1993) Transposition pattern of the maize element *Ds* in *Arabidopsis thaliana*. Genetics 134: 1221-1229

Barry G (2001) The use of the Monsanto draft rice genome sequence in research. Plant Physiol. 125: 1164-1165

Bennetzen JL (2000a) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42: 251-269

Bennetzen JL (2000b) Comparative sequence analysis of plant nuclear genomes:microcolinearity and its many exceptions. Plant Cell 12: 1021-1029

Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: genome composition, collinearity and compatibility. Trends Genet 9: 259-261

Bennetzen JL, Ma JX (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol 6: 128-133

Bohnert H, Ayoubi P, Borchert C, Bressan R, Burnap R, Cushman J, Cushman M, Deyholos M, Fischer R, Galbraith D (2001) A genomics approach towards salt stress tolerance. Plant Physiol Biochem 39: 295-311

Bryan GT, Wu KS, Farrall L, Jia YL, Hershey HP, McAdams SA, Faulk KN, Donaldson GK, Tarchini R, Valent B (2000) A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell 12: 2033-2045

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, et al. (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138: 1251-1274

Chandelier C (2003) The genoplante microarrays for rice. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Chang T (1976) The origin, evolution, cultivation, dissemination and diversification of Asian and African rices. Euphytica 25: 435-444

Chen MS, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B et al. (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14: 537-545

Cho YG, McCouch SR, Kuiper M, Kang MR, Pot J, Groenen JTM, Eun MY (1998) Integrated map of AFLP, SSLP and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). Theor Appl Genet 97: 370-380

Christou P, Ford T, Koffron M (1991) Production of transgenic rice (*Oryza sativa* L.) plants from agronomically important indica and japonica varieties via electrodischarge particule acceleration of exogenous DNA into immature zygotic embryos. Biotechnology 9: 957-962

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

Devic M, Albert S, Delseny M, Roscoe T (1997) Efficient PCR walking on plant genomic DNA. Plant Physiol Biochem 35: 331-339

Devos KM, Beales J, Nagamura Y, Sasaki T (1999) Arabidopsis - Rice: Will colinearity allow gene prediction across the eudicot-monocot divide? Genome Res 9: 825-829

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell 12: 637-646

Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan LL, Shiloff BA, Bennetzen JL (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol 125: 1342-1353

Earp DJ, Lowe B, Baker B (1990) Amplification of genomic sequences flanking transposable elements in host and heterologous plants: a tool for transposon tagging and genome characterization. Nucleic Acids Res 18: 3271-3278

Feng Q, Zhang Y, Hao P, Wang S, Fu G, et al. (2002) Sequence and analysis of rice chromosome 4. Nature 420: 316-320

Freeling M (2001) Grasses as a single genetic system: reassessment 2001. Plant Physiol 125: 1191-1197

Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. Science 282: 656-659

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

Hanin M, Paszkowski J (2003) Plant genome modification by homologous recombination. Curr Opin Plant Biol 6: 157-162

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148: 479-494

Helliwell C, Wesley V, Wielopolska A, Wu R, Bagnall D, Waterhouse P (2002) High throughput gene silencing in plants. Abstract of the International Worhshop "Towards building a global rice gene machine". CSIRO Plant Industry, Canberra, Australia

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

Izawa T, Miyazaki C, Yamamoto M, Terada R, Iida S, Shimamoto K (1991) Introduction and transposition of the maize transposable element Ac in rice (*Oryza sativa* L.). Mol Gen Genet 227: 391-396

Izawa T, Ohnishi T, Nakano T, Ishida N, Enoki H, Hashimoto H, Itoh K, Terada R, Wu C, Miiyazaki C, Endo T, Iida S, Shimamoto K (1997) Transposon tagging in rice. Plant Mol Biol 35: 219-229

Izawa T, Oikawa T, Tokutomi S, Okuno K, Shimamoto K (2000) Phytochromes confer the photoperiodic control of flowering in rice (a short-day plant). Plant J 22: 391-399

Jeon J, An G (2001) Gene tagging in rice: A high throughput system for functional genomics. Plant Sci. 161: 211-219

Jeon J, Jang S, Lee S, Nam J, Kim Cea (2000) *leafy hull sterile1* is a homeotic mutation in a rice MADS box gene affecting rice flower development. Plant Cell 12: 871-884

Johnson A, Tester M, Hibberd J, Gay C, Sallaud C, Guiderdoni E (2003) Generation of enhancer trap lines of rice (*Oryza sativa* L.) expressing GAL4 and GFP in specific cell types. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Kang HG, Jeon JS, Lee S, An GH (1998) Identification of class B and class C floral organ identity genes from rice plants. Plant Mol Biol 38: 1021-1029

Kawasaki T, Henmi K, Ono E, Hatakeyama S, Iwano M, Satoh H, Shimamoto K (1999) The small GTP-binding protein Rac is a regulator of cell death in plants. Proc Natl Acad Sci 96: 10922-10926

Keller J, Lim E, Dooner HK (1993) Preferential transposition of *Ac* to linked sites in Arabidopsis. Theor Appl Genet 86: 585-588

Kerim T, Imin N, Weinman J, Rolfe B (2003) Proteome analysis of male gametophyte development in rice anthers. Proteomics 3: 738-751

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35: 25-34

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, et al. (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. Science 301: 376-379

Koes R, Souer E, Van Houwelingen A, Mur L, Spelt C, Quattrocchio F, Wing J, Oppedijk B, Ahmed S, Maes T, Gerats T, Hoogeveen P, Meesters M, Kloos D, Mol Joseph NM (1995) Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. Proc Natl Acad Sci 92: 8149-8153

Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) *Hd3a*, a rice ortholog of the Arabidopsis *FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol 43: 1096-1105

Komatsu S, Konishi H, Shen S, Yang G (2003) Rice proteomics: a step toward functional analysis of the rice genome. Mol Cell Proteom 2: 2-10

Kumar A, Bennetzen JL (2000) Retrotransposons: central players in the structure, evolution and function of plant genomes. Trends Plant Sci 5: 509-510

Leung H (2003) Public research platform in rice functional genomics: applications of mutational analysis and allele mining for trait improvement. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, Wang X, Liu X, Teng S, Hiroshi F, Yuan M, Luo D, Han B, Li J (2003) Control of tillering in rice. Nature 422: 618-621

Liu H, Sachidanandam R, Stein L (2001) Comparative genomics between rice and Arabidopsis shows scant collinearity in gene order. Genome Res 11: 2020-2026

Liu YG, Whittier RF (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25: 674-681

Long D, Martin M, Sundberg E, Swomburne J, Puangsomlee P, Coupland G (1993) The maize transposable element system *Ac/Ds* as a mutagen in Arabidopsis: Identification of an *albino* mutation induced by *Ds* insertion. Proc Natl Acad Sci 90: 10370-10374

Mayer K, Murphy G, Tarchini R, Wambutt R, Volckaert G, Pohl T, Dusterhoft A, Stiekema W, Entian KD, Terryn N, Lemcke K, Haase D, Hall CR, van Dodeweerd AM, Tingey SV, Mewes HW, Bevan MW, Bancroft I (2001) Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. Genome Res 11: 1167-1174

McCouch SR, Teytelman L, Xu YB, Lobos KB, Clare K, Walton M, Fu BY, Maghirang R, Li ZK, Xing YZ, Zhang QF, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9: 199-207

Moore G, Foote T, Helentjaris T, Devos K, Kurata N, Gale M (1995) Was there a single ancestral cereal chromosome? Trends Genet 11: 81-82

Murai N, Li Z, Kawagoe Y, Hayashimoto A (1991) Transposition of the maize activator element in transgenic rice plants. Nucleic Acids Res 19: 617-622

Parinov S, Sundaresan V (2000) Functional genomics in Arabidopsis: large-scale insertional mutagenesis complements the genome sequencing project. Curr Opin Biotechnol 11: 157-161

Quackenbush J, Liang F, Holt I, Pertea G, Upton J (2000) The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res 28: 141-145

Reddy A, Ramakrishna W, Sekhar A, Ithal N, Babu P, Bonaldo M, Soares M, Bennetzen J (2002) Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. *indica* cv. Nagina 22). Genome 45: 204-211

Saijo Y, Hata S, Kyozuka J, Shimamoto K, Izui K (2000) Over-expression of a single Ca2+-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. Plant J 23: 319-327

Saji S, Umehara Y, Antonio BA, Yamane H, Tanoue H, Baba T, Aoki H, Ishige N, Wu J, Koike K, Matsumoto T, Sasaki T (2001) A physical map with yeast artificial chromosome (YAC) clones covering 63% of the 12 rice chromosomes. Genome 44: 32-37

Sallaud C, Meynard D, Van Boxtel J, Gay C, Bes M, Brizard J, Larmande P, Ortega D, Raynal M, Portefaix M, Ouwerkerk P, Rueb S, Delseny M, Guiderdoni E (2003) Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics. Theor Appl Genet 106: 1396-1408

Salse J, Piegu B, Cooke R, Delseny M (2002) Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. Nucleic Acids Res 30: 2316-2328

SanMiguel P, Ramakrishna W, Bennetzen J, Busso C, Dubcovsky J (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). Funct Integr Genom 2: 70-80

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, et al. (2002) The genome sequence and structure of rice chromosome 1. Nature 420: 312-316

Shimamoto K, Miyazaki C, Hashimoto H, Izawa T, Itoh K, Terada R, Inagaki Y, Iida S (1993) 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: 354-360

Shimamoto K, Moritoh S, Miki D (2002) RNAi as a complementing method for functional genomics of rice. Abstract of the International Worhshop "Towards building a global rice gene machine". CSIRO Plant Industry, Canberra, Australia

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA (1995) An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res 23: 1087-1088

Song R, Llaca V, Messing J (2002) Mosaic organization of orthologous sequences in grass genomes. Genome Res 12: 1549-1555

Song W, Wang G, Chen L, Kim H, Pi L, al. e (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270: 1804-1806

Springer P (2000) Gene traps: Tools for plant development and genomics. Plant-cell,-The 12: 1007-1020.

Takahashi Y, Shomura A, Sasaki T, Yano M (2001) *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. Proc Natl Acad Sci 98: 7922-7927

Tao Q, Chang YL, Wang J, Chen H, Islam Faridi MN, Scheuring C, Wang B, Stelly DM, Zhang HB (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158: 1711-1724

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12: 381-391

Terada R, Urawa H, Inagaki Y, Tsugane K, Iida S (2002) Efficient gene targeting by homologous recombination in rice. Nat Biotechnol 20: 1030-1034

The Rice Chromosome 10 Sequencing Consortium (2003) In-depth view of structure, activity, and evolution of rice chromosome 10. Science 300: 1566-1569

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JD (1999) Multiple independent defective *suppressor-mutator* transposon insertions in Arabidopsis: a tool for functional genomics. Plant Cell 11: 1841-1852

Toriyama K, Arimoto Y, Uchimiya H, Hinata K (1988) Transgenic rice plants after direct gene transfer into protoplasts. Biotechnology 6: 1072-1074

Van den Broeck D, Maes T, Sauer M, Zethof J, De Kuekeleire P, D'Hauw M, Van Montagu M, Gerats T (1998) Transposon display identifies individual transposable elements in high copy number lines. Plant J 13: 121-129

van Dodeweerd AM, Hall CR, Bent EG, Johnson SJ, Bevan MW, Bancroft I (1999) Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. Genome 42: 887-892

Vandepoele K, Saeys Y, Simillion C, Raes J, Van de Peer Y (2002) The automatic detection of homologous regions (ADHoRe) and its application to microcolinearity between Arabidopsis and rice. Genome Res 12: 1792-1801

Wang ZX, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, Katayose Y, Sasaki T (1999) The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J 19: 55-64

Wu JZ, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, Fujii F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano M, Matsumoto T, Sasaki T (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 14: 525-535

Yamamoto K, Sasaki T (1997) Large-scale EST sequencing in rice. Plant Mol Biol 35: 135-144

Yamamoto Y, Tsuhara Y, Gohda K, Suzuki K, Matsui M (2003) Gene trapping of the Arabidopsis genome with a firefly luciferase reporter. Plant J 35: 273-283

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. Plant Cell 12: 2473-2483

Yazaki J, Kishimoto N, Ishikawa M, Endo D, Kojima K, MicroArray Center, Kikuchi S (2002) The Rice Expression Database (RED): gateway to rice functional genomics. Trends Plant Sci 7: 563-564

Yazaki J, Kishimoto N, Nakamura K, Fujii F, Shimbo K, Otsuka Y, Wu J, Yamamoto K, Sakata K, Sasaki T, Kikuchi S (2000) Embarking on rice functional genomics via cDNA microarray: use of 3' UTR probes for specific gene expression analysis. DNA Res 7: 367-370

Yoshimura A, Ideka O, Iwata N (1997) Linkage map of phenotype and RFLP markers in rice. Plant Mol Biol 35: 49-60

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang ZX, Kono I, Kurata N, Yano M, Iwata N, Sasaki T (1998) Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci 95: 1663-1668

Yu J, Hu S, Wang J, Wong GK, Li S, et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296: 79-92

Chapter 2

Transposons and functional genomics in rice.

Raffaella Greco[1], Pieter B.F. Ouwerkerk[2], Christophe Sallaud[3], Ajay Kohli[4], Cristina Favalli[5], Thierry Beguiristain[6], Lucia Colombo[5], Enrico Pè[5], Pere Puigdomènech[6], Emmanuel Guiderdoni[3], Paul Christou[4], J.Harry C. Hoge[2†] and Andy Pereira[1]

[1]Plant Research International, Wageningen, The Netherlands
[2]Institute of Biology, Leiden University, Leiden, The Netherlands
[3]CIRAD-BIOTROP, Montpellier, France
[4]John Innes Centre, Norwich, UK
[5]Dept. of Biology, University of Milan, Milano, Italy
[6]CID-CSIC, Barcelona, Spain

In: Rice Genetics IV, pp. 263-277 (2001)
Edited by Khush GS, Brar D and Hardy B
IRRI and Science Publishers Inc.

ABSTRACT

Genome and expressed sequence tag sequencing in rice provides a vast resource of gene sequences whose functions need to be determined by reverse genetics methods for expression and mutational analysis. To develop insertional mutagenesis strategies in rice, we transformed japonica and indica cultivars with maize transposon constructs, for knockout and gene detection insertions. A green fluorescent protein (GFP) excision assay developed enabled the visualization of transposon excision in a variety of tissues. Surprisingly, early *Ac* excision was observed directly after transformation from a construct containing the strong double CaMV enhancer element adjacent to the *Ac* promoter. We identified genotypes with *Ac* amplification events and with a forward transposition rate of 15–50% that are useful for generating lines containing multiple transposons. The sequence of DNA flanking transposed *Ac* provided a resource of *Ac*-tagged sites, which represented about 50% in the target region, indicating insertional specificity appropriate for the identification of mutants of sequenced genes. Clustered *Ac* transposition was revealed by six insertions in 70 kb of chromosome 6. Gene detection *Ac*-*Ds* enhancer trap and activation tag transformants revealed active transposition in about half the lines. These resources for functional genomics are developed by an EU-funded consortium and will be made available to rice researchers worldwide.

Genomic technologies and DNA sequence information have become important in understanding the genetics of plants. This is particularly true for model organisms such as rice, in which the availability of vast amounts of genome information encourages biologists to ask questions in genetics, biochemistry, physiology, and molecular biology in a genomic perspective. *Arabidopsis* genomics and functional genomics have led the way and will provide the blueprint for the structure and function of many plant genes. Some of the remarkable insights emerging from whole-genome analysis (Lin et al 1999, Mayer et al 1999) are the number of gene clusters, large-scale duplications of chromosome segments, as well as a surprisingly high frequency (40%) of newly discovered genes of unknown function.

Functional genomics

The field of functional genomics has emerged to address the function of genes discovered by genome sequencing. In contrast to the previously prevalent gene-by-gene approaches, new high-throughput methods are being developed for expression analysis as well as for the recovery and identification of mutants. The experimental approach is consequently changing from hypothesis-driven to nonbiased data collection and an archiving methodology that makes these data available in relational databases that can be analyzed by bioinformatics tools. The functional genomics methodology is also changing the experimental strategy from a forward genetics (or mutant to gene) approach to a reverse genetics (or sequenced gene to mutant and function) approach. It is expected that the functional genomics of model plants will contribute to the understanding of basic plant biology as well as the exploitation of genomic information for crop plant improvement. This is because a large number of gene functions for generic traits will be functional across species, either directly or after identifying the functional homologues.

With its relatively small genome (450 Mb), rice has been selected to be the model monocot plant to be first sequenced. Although the other cereal genomes will not be sequenced completely soon, the synteny between the monocots (Gale and Devos 1998) will help decipher the structure and function of the more complex genomes.

The major developments in whole-genome analysis have been in the field of transcript expression analysis using a variety of high-throughput methods. In plants, different technologies have been employed (Baldwin et al 1999) based on high-density nylon filters, microarrays, various gel systems, and even a sequencing-based method, SAGE, that has been employed for rice (Matsumura et al 1999). With gene chips or microarrays, the representative genes of an organism can be placed on a solid support such as glass and used in hybridization experiments with different RNA samples to reveal gene expression patterns and help identify pathways by association. However, the expression patterns of genes supply correlative information and do not necessarily prove a causal relationship between gene sequence and function. As mRNA level

alone does not reflect the actual expression of the gene product, additional means are provided by proteomics (Dove 1999), which addresses the protein expression of a cell type, and metabolomics, which reveals the metabolite profile (Trethewey et al 1999) of the cell/tissue type revealing the active biochemical processes. A combination of different expression analysis tools and mutant or overexpression analysis of genes will be able to provide a unique multidimensional picture of genetic circuits and pathways.

Knockout mutagenesis

The phenotype associated with a gene function is often the best clue to its role in the plant. Phenotype of mutants can be broadly defined at the morphological, biochemical, or physiological level and it provides information on the interactions among different processes. Classically, chemical and physical mutagens have provided loss-of-function mutants and have helped define genes involved in specific pathways or processes. To study the interactions among genes, mutants in a particular pathway can be combined and the genetic hierarchy studied. These mutants can be precisely mapped and lead to gene isolation by map-based cloning procedures.

Insertional mutagenesis, with transposable elements or *Agrobacterium*-mediated T-DNA insertions, can generate mutants and also lead directly to gene identification. The insertion often causes a knockout mutation by blocking the expression of the gene and might display a mutant phenotype. The mutant gene tagged by the insertion can be isolated by recovering DNA flanking the insert and subsequently lead to the isolation of the wild-type gene. For forward genetics, T-DNA insertional mutagenesis is only practical on a large scale for *Arabidopsis* (avoiding somaclonal variation) and has resulted in a large number of tagged genes, although its use for reverse genetics in model plants such as rice has a future.

The ability of the well-characterized maize transposon systems *Ac-Ds* and *En-I(Spm*) to transpose in heterologous hosts (Baker et al 1986, Pereira and Saedler 1989) offers new possibilities for transposon tagging, which is an efficient tool for identifying genes. For effective tagging strategies, two component systems comprising a mobile transposon component (*Ds* or *I/dSpm*) and the corresponding stable transposase (*Ac* or *En/Spm*) source have been developed (reviewed in Pereira 1998). The mobile transposon components are often inserted in assayable/selectable marker genes to monitor their excision phenotype (Baker et al 1987) and can also contain other convenient marker/selectable genes to screen for their presence. To control transposition, the transposase is often put under the control of heterologous promoters (Swinburne et al 1992) or segregated out in progeny to yield stable transposon inserts. As transposons move preferentially to closely linked sites (Jones et al 1990) and tag genes efficiently near their original position, efforts are made to map transposons in heterologous systems in order to generate jumping pads all over the genome for efficient targeted mutagenesis.

Ac-Ds transposons were first introduced by electroporation into rice (Izawa et al 1991, Murai et al 1991, Shimamoto et al 1993) and shown to transpose. It was noted, however, that the two-component *Ac-Ds* system often gets inactivated after the first generation (Izawa et al 1997), justifying a systematic investigation of *Ac-Ds* biology in rice. Recently, the behavior of *Ac* has been followed through three successive generations and it has revealed characteristics suitable for functional genomics strategies (Enoki et al 1999). Another knockout mutagenesis tool in rice, an endogenous retrotransposon, *Tos17*, has been developed for reverse genetics screens (Hirochika 1997).

The use of knockout mutations is limited, however, as the majority of genes display no obvious phenotype (Burns et al 1994), probably because of functional redundancy, in which one or more other homologous loci can substitute for the same function. Even *Arabidopsis*, with its simple genome, contains large duplicated segments and redundant genes (Lin et al 1999, Mayer et al 1999). Therefore, the sequential disruption of redundant genes in an individual genotype might ultimately reveal a mutant phenotype and uncover the gene function.

Gene detection

Gene detection strategies have been developed to address the function of genes that do not directly reveal a knockout phenotype. One way is expression detection that can make use of inserts containing reporter gene constructs, such as enhancer traps or gene traps, whose expression depends on transcriptional regulatory sequences of the adjacent host gene. Another way is by creating misexpression mutants, such as activation tags that might reveal a gain-of-function phenotype.

Enhancer detection was developed to detect enhancers in the genome, which are capable of orientation-independent transcriptional activation from a distance, and thus identify genes based on their expression pattern (Skarnes 1990) even though they might not display an obvious mutant phenotype. Enhancer detection constructs contain a reporter gene such as ß-glucuronidase (GUS) with a weak or minimal promoter, for example, with a TATA box (transcription initiation signal), situated near the border of the insert and are expressed when integrated adjacent to an enhancer in the genome. Gene trap-type inserts are designed to create fusion transcripts with the target gene. One type used effectively in plants is an exon trap that enables reporter gene fusions to be created at various locations within a gene. By introduction of splice acceptor sites upstream of the reporter gene, transcriptional fusions are created even for insertions in introns, thus increasing the frequency of inserts expressing the reporter gene.

A novel method for efficient selection of stable transpositions, using positive and negative selection markers, has been developed for an *Ac-Ds*-based gene trap and enhancer trap system (Sundaresan et al 1995). Stable transposed *Ds* elements were recovered and revealed GUS

expression activity in about 50% of enhancer trap (*DsE*) and 25% of gene trap (*DsG*) inserts (Sundaresan et al 1995). A similar strategy for selection of stable transposed gene trap *DsG* elements has begun in rice (Chin et al 1999) using greenhouse-selectable marker systems.

Misexpression or gain-of-function mutants can be generated by insertion sequences that carry a strong enhancer element near the border and activate the expression of the adjacent genes. This method of "activation tagging" has been validated by an extensive screen of more than 25,000 T-DNA tags (Weigel et al 2000), revealing a 1/1,000 frequency of dominant mutants that were shown to be caused by the presence of an enhancer between 0.4 and 4 kb from the overexpressed gene. A transposon construct variant to isolate dominant gain-of-function alleles employed the CaMV 35S promoter transcribing outward from the *Ds* transposon end and helped identify a semidominant overexpression mutant (Wilson et al 1996). These examples demonstrate the use of activation tagging to generate dominant mutations by over-/misexpressing genes and generate phenotypes for processes or genes not uncovered by knockout mutants.

Generation of rice transposon insertion populations

To develop efficient transposon mutagenesis strategies in rice, we tested several parameters in order to select components for transposon constructs to be transformed. Constructs were made with the aim of generating populations of transposon inserts for knockout mutagenesis as well as gene detection using enhancer traps and activation tags. Both the maize *Ac-Ds* and *En-I* transposon systems were used and tested for their characteristic advantages. For transformation, primarily *Agrobacterium*-mediated transformation (Hiei et al 1994) was used for japonica cultivars and particle bombardment (Christou et al 1991) for indica cultivars. The analyses of the transformants with different transposon constructs suggest strategies that could be employed to make an efficient mutation machine in rice.

Development of a GFP transposon excision assay

Several promoter-GFP (green fluorescent protein) gene constructs were bombarded (particle gun) into suspension-cultured rice cells and the cells subsequently assayed for GFP expression by fluorescence microscopy. The gene constructs used differed in type of GFP (codon usage, protein solubility/folding, protein chromophore, and presence of an intron in the coding region) as well as in type of promoter. From the detected transient GFP expression levels (Table 1), it was concluded that the GFP types sGFP(S65T) and pGFPint(S65T), in combination with the promoter types 35S, d35S, ACT, GOS2, or UBI, would be best suited to serve as a transposon excision marker in rice. Transgenic plants were also generated and tested with the sGFP(S65T)

under control of the d35S, ACT, UBI, and GOS2 promoters in comparison with the MON30063-GFP construct and confirmed the bombardment experiments.

GFP	Promoter	Excitation		Emission
		UV	Blue	green (blue/yellow)
smGFP	35S	Max	Min	–
smGFP	UBI	Max	Min	+/–
smBFP	35S	Max	0	– (blue)
smRS-GFP (S65T)	35S	Min	Max	+/–
mGFP5	35S	Equal		–
mGFP5	UBI	Equal		–
mGFP6 (S65T)	35S	0	Max	++
sGFP (S65T)	35S	0	Max	++++
sGFP (S65T)	d35S	0	Max	++++
sGFP (S65T)	ACT	0	Max	++++
sGFP (S65T)	UBI	0	Max	+++
sGFP (S65T)	GOS2	0	Max	++++
pGFP-int S65T(MON30063)	d35S	0	Max	+++++
pGFP-int S65T(MON30049)	d35S	0	Max	++++
EGFP-C3	35S	0	Max	++
EGFP-C3	CMV	0	Max	+
EYFP	d35	0	Max	+++ (yellow)

Table 1. Transient expression of promoter-GFP (green fluorescent protein) gene constructs in bombarded (particle gun) suspension-cultured rice cells. Indicated is the type of light (UV, ultraviolet, or blue) required for optimal (max) or suboptimal (min) excitation. The relative expression (emission of green, blue, or yellow light) of the constructs is indicated with minus or plus signs. Promoter abbreviations are 35S = cauliflower mosaic virus 35S; d35S = 35S with double enhancer; UBI = maize ubiquitin; ACT = rice actin; GOS2 = rice constitutive; CMV = (animal) cyto megalo virus. The origin of the GFPs is smG(B)FP (Davis and Vierstra 1998), sGFP(S65T) (Chiu et al 1996), mGFP5 (pCAMBIA vectors), EG(Y)FP (Clontech), and MON30063/MON30049 (Phang et al 1996).

A construct termed d35S-sGFP:Ac (Table 2), harboring an autonomous *Ac* element between the d35S promoter and the sGFP(S65T) gene, was introduced into rice via *Agrobacterium*. The obtained transgenic calli displayed easily detectable fluorescent GFP sectors (Greco et al, submitted), indicative of *Ac* excision. The GFP excision assay allowed the identification of excision at various stages during the transformation process and revealed a high excision rate. As GFP expression is easily visible in the seed, this phenotypic excision assay is potentially suitable for selecting independent germinal excision events before planting.

Table 2. Molecular analysis of transposon construct transformants.

Transposon	Construct	Rice variety	Indep. lines analyzed[a]	Lines with entire T-DNA[a]	Indep. active lines[a]	% active lines[b]
Autonomous Ac	35S-smGFP:Ac	Taipei 309	1 (9)	1 (9)	1 (9)	100
	d35S-sGFP:Ac	Taipei 309	11 (45)	11 (45)	11 (33)	100
		Nipponbare	8 (52)	7 (45)	6 (32)	86
Total			20 (106)	19 (99)	18 (74)	-
Direct DNA transfer	UBI-smGFP:Ac	Indica cvs.	348	-	278	80
Ac total			368 (454)	-	296 (352)	-
2-component Ac-Ds	2-comp. Ac-Ds	Taipei 309	3 (19)	1 (13)	1 (13)	100
Enhancer trap	Ac-Ds ET	Taipei 309	11 (45)	10 (37)	5 (11)	50
	Ac-Ds ET	Nipponbare	10 (31)	8 (27)	4 (14)	50
	Ac-Ds ET with GFP	Nipponbare	219 (278)	104 (119)	63 (72)	60
ET total			240 (354)	122 (183)	72 (97)	-
Activation tag	Ac-Ds AT	Nipponbare	9 (15)	5 (7)	3 (3)	60
	Ac-Ds AT with GFP	Nipponbare	42 (75)	14 (17)	4 (4)	29
AT total			51 (90)	19 (24)	7 (7)	-
Ac transposase	35s-AcTPase+SU1	Taipei 309	15 (15)	12 (12)	-	-
		Nipponbare	12 (15)	11 (13)	-	-

[a] Total number of plants in parentheses.

[b] The percentage is calculated on the number of independent lines that were not rearranged.

Ac transposition behavior in rice

The transformants with the autonomous *Ac* transposon were analyzed by PCR and DNA blot hybridization for molecular evidence of excision and reinsertion. Table 2 summarizes the results of the transformation experiments and molecular analyses of individual regenerants.

The most remarkable discovery was that, in transformants with the d35S-sGFP:*Ac* construct with an enhancer adjacent to the *Ac* promoter (Fig. 1), *Ac* transposed directly after transformation in the plant cell in every transformant. The evidence was obtained by DNA sequence analysis of excision and reinsertion alleles from about 40 regenerated plants (Greco et al, submitted). This has never been shown before and probably occurs because the strong CaMV 35S enhancer adjacent to the *Ac* promoter influences transposase expression. The *cis* effect of the CaMV enhancer on the *Ac* promoter and transposase expression was observed earlier in *Arabidopsis* (Balcells and Coupland 1994), leading to earlier excision of *Ds.*

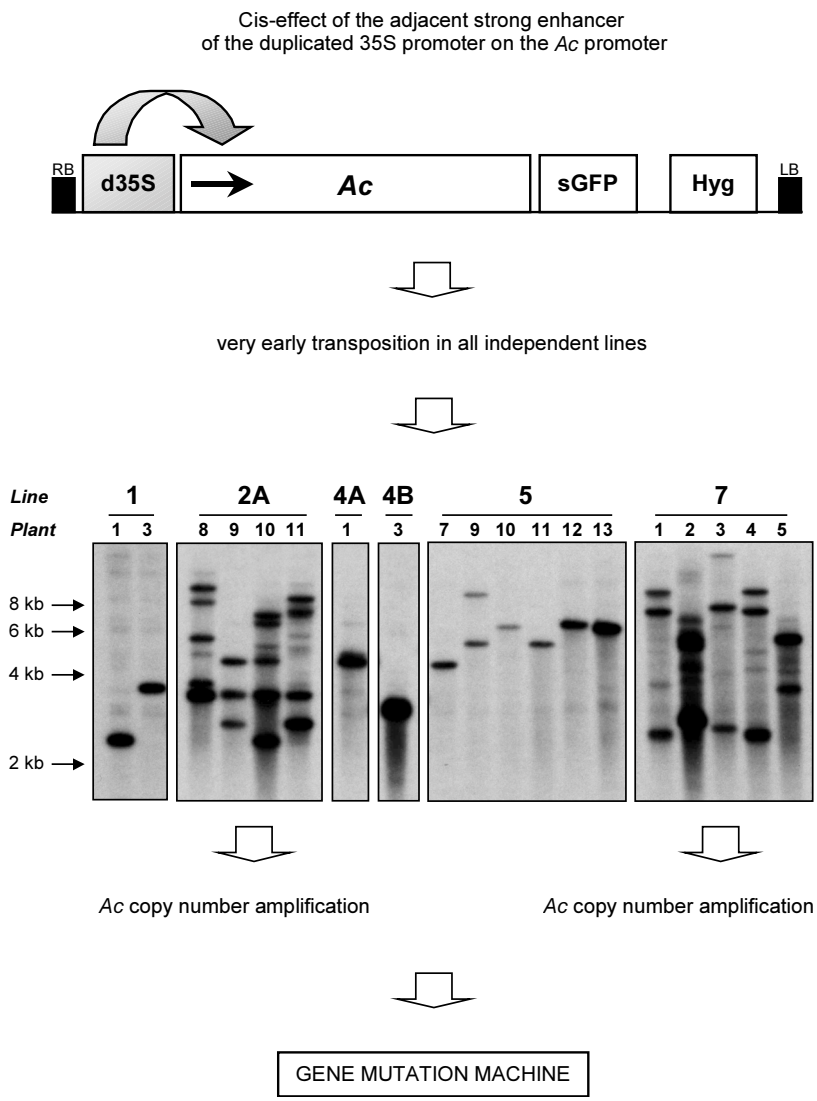


Figure 1. Early transposition and amplification of autonomous *Ac*. The d35S-sGFP::*Ac* construct with a duplicate CaMV 35S enhancer-promoter adjacent to the *Ac* promoter (shown by arrow). On excision, the sGFP expression can be visualized by green fluorescent protein (GFP) fluorescence. Shown below is DNA blot hybridization of individual regenerants from six different transformed lines (as labeled), hybridized with an *Ac*-specific probe and fragment sizes given on left. All lines showed very early transposition and the individual regenerants (e.g., 8, 9, 10, 11) from line 2A (single T-DNA insert) display active transposition and amplification. Multiple *Ac* lines displaying independent transposition are used to make a mutation machine for reverse genetics. RB = right border, LB = left border.

Ac reinsertion was evident in about 70% of the lines and one-third of these also showed an amplification in *Ac* copy number as demonstrated by the number of *Ac*-hybridizing fragments in the one- and two-copy *Ac* lines (lines 2A and 7 in Fig. 1). The process of transposition during replication has been described in maize (Greenblatt 1984), in which *Ac* transposes from replicated DNA to unreplicated DNA, resulting in a moderate increase in copy number. In the amplifications we observed, the presence of a single excision footprint in all the regenerants from a single T-DNA insertion (Greco et al, submitted) suggests that transposition must have begun from an unreplicated site. The mechanism of *Ac* amplification we propose is outlined in Figure 2 and is potentially useful for generating genotypes with multiple transposons (Fig. 1) that can reach genome saturation quickly.

To compare the transpositional activity between transformants, a parameter termed "independent transposition frequency" (ITF) given by the frequency of unique transposition events among the total inserts in a set of related plants (regenerants or seed progeny) was estimated. The somatic ITF calculated over three populations of regenerants varies between 50% and 65%. Transposed *Ac* insertions that were present in only a single plant were considered as unique late events that were useful for producing different insertions in the progeny.

In contrast to the active transposition in regenerating calli, leaf samples of mature plants revealed a stable transposon insertion pattern. The individual progeny again revealed new insertions with an ITF between 15% and 50% in different lines, suggesting that transposition took place prior to gamete formation. This indicates a biphasic transposition pattern: an active phase during regeneration from the callus as well as the pregametic cells and a stable phase in the mature plant.

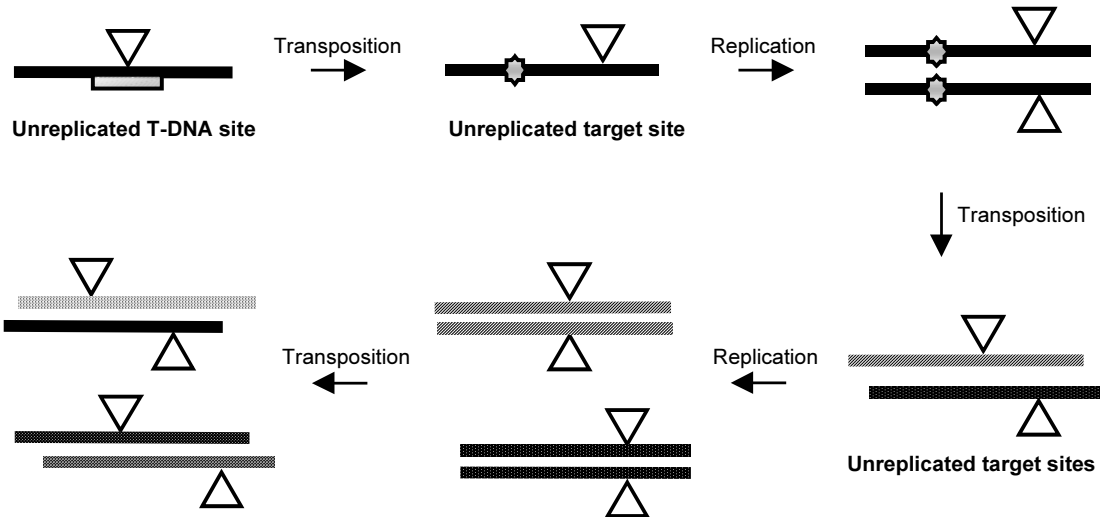


Figure 2. *Ac* amplification during replication in one cell. The mechanism of *Ac* amplification proposed to explain the results of *Ac* (*Ac* designated as triangle) transposition observed. The high transposase activity mediated by the enhancer adjacent to the *Ac* promoter in the initial cell after transformation and before further cell division triggers an early excision event from the unreplicated T-DNA site (shaded box) that is revealed by the occurrence of the same excision footprint (open star on chromosome site) in every regenerant. In the same cell cycle, the excised *Ac* reinserts into a second site, which then replicates to generate a second *Ac* copy in the newly replicated strand. In the continued presence of the high *Ac* transposase activity in the same cell, secondary transposition events of both *Ac* copies result in two new unreplicated positions that get duplicated following replication, giving rise to four copies at the two chromosomal positions. The high transposase activity mediates a third cycle of transposition and replication that produces four new *Ac* chromosomal positions (denoted by different shaded lines) in the cell.

Ac knockout mutagenesis

The *Ac* lines we characterized in rice revealed multiple transpositions due to amplifications with more than four inserts per plant that could generate an average of one to two new inserts per progeny. The propagation of these genotypes for three to four generations can generate a population of plants containing four or more *Ac* inserts at different positions in the genome. Using 25,000 lines for three to four generations would generate about 100,000 insertions that are suitable for identifying knockouts for forward as well as reverse genetic strategies.

To analyze the transposon insertion sites in these lines, we isolated and sequenced DNA flanking transposed *Ac* elements and compared these *Ac* insertion tagged site (ITS) sequences with those in public databases. We could position six ITSs in a 70-kb interval of sequenced DNA on chromosome 6 (Fig. 3), demonstrating linked transposition that is useful for targeted tagging. The majority of the *Ac* inserts are oriented in the same way (thin arrows in Fig. 3), suggesting that the transposition process is not random. Most significantly, as this chromosome is being systematically sequenced, the *Ac* lines we have generated will be very useful for obtaining knockout mutants of genes identified in this region.

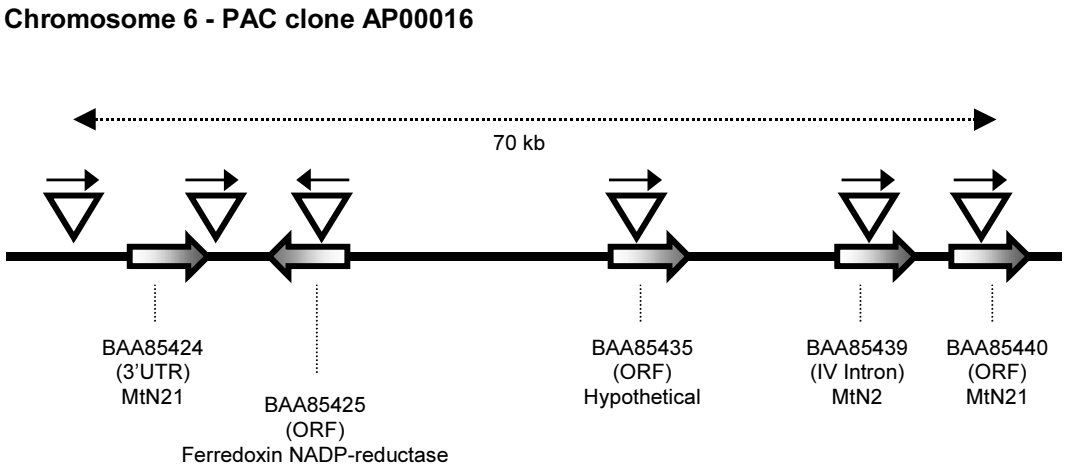


Figure 3. Local transposition of *Ac* on chromosome 6. The isolation and sequencing of DNA flanking *Ac* inserts in a rice *Ac* line enabled the identification of several tagged sites in a sequenced P1-derived artificial chromosome (PAC) clone from rice chromosome 6. The *Ac* insertions are shown as triangles and the orientation of the *Ac* element in the genes is shown by the thin arrows above the *Ac* inserts. The predicted genes containing insertions are denoted by thick solid arrows and are labeled underneath by accession number, position of insert in the gene, and similarity to known/predicted proteins. ORF = open reading frame, UTR = untranslated region.

The ITS sequences revealed a high proportion with homology to predicted genes in the databases. About 29% of the ITSs reveal homology to known proteins and 10% to sequenced ESTs (Greco et al, submitted). To calculate the genome sequence predicted to code for proteins, we assume 30,000 rice genes with an average 2.5-kb coding sequence and 55% predicted genes with similarity to proteins of known function based on *Arabidopsis* estimates (Lin et al 1999, Mayer et al 1999). This transforms to 41.5 Mb (30,000 × 2.5 × 55/100) or 10% of the genome that should show homology to known proteins. The transposon inserts are therefore biased about 3 times higher than expected on a random basis to insert in sequences predicted to code for proteins. To calculate the genome sequence represented by ESTs, with 20,000 different ESTs based on the gene index calculation (Quackenbush et al 2000) and an average size of 400-bp sequence, 8 Mb of unique EST sequence or 1.86% (8/430) of the genome sequence is present in the public databases. We observe 9.6% of the total ITSs with identity to ESTs, meaning that the ITSs are biased for insertions in transcribed genes five times more than randomly expected.

These two calculations reveal a bias of three to five times more insertions in genes and strongly suggest that *Ac* inserts preferentially into genes in rice. This confirms the earlier results of *Ac* insertional preference in rice, in which 4% of the inserts in ESTs were observed (Enoki et al 1999), and suggests that the preferential transposon insertions in genes could be a valuable asset for generating mutants in rice.

Gene detection strategies in rice

Several constructs for gene detection were developed in rice. The general structure of the enhancer trap (ET) and activation tag (AT) constructs is outlined in Figure 4 and Table 2 summarizes the transformation experiments. The *Ac-Ds* construct contains the mobile transposon that can be a *Ds* or an *I* transposon and the corresponding immobile transposase source *Ac* or *En* under control of a strong promoter. To observe transposition early after transformation and help make a choice among transformants, the GFP excision marker can be used to monitor excision. Independent excision events can be selected at the seed or germinated seedling level so that germinal transposants are recovered and grown. To be able to select stable transposed elements, we employed a BAR gene on the *Ds* element conferring resistance to the herbicide Basta ($Basta^R$) and a negative selection marker SU1 (O'Keefe et al 1994) that converts the proherbicide R7402 into the active herbicide and results in shorter plants (Koprek et al 1999). Using a combination of these greenhouse/field selectable markers, progeny of single T-DNA locus transformants can be used to identify stable transposants (BAR^+ $SU1^-$) (Fig. 5), where the *Ds*-BAR transposes from the T-DNA and is segregated away (to another chromosome or after recombination from a linked site).

The enhancer trap construct contains a minimal promoter that can initiate transcription upstream of the GUS marker gene. On insertion near enhancers of host genes in the genome, the GUS gene detector can display the expression pattern of the adjacent gene. This will help identify the adjacent plant gene on the basis of its expression. In plants such as *Arabidopsis*, about 50% of the inserts display expression that is similarly observed in rice. The activation tag construct has a multiple enhancer of the CaMV 35S promoter inserted near the transposon end and can activate adjacent genes. Enhancer trap and activation tag constructs have been transformed and molecularly analyzed in rice (Table 1). They yield about 50% of the active lines that are being propagated for developing tagging strategies. Some lines also contain multiple *Ds* elements with active transpositions that are useful for generating multiple transposon lines.

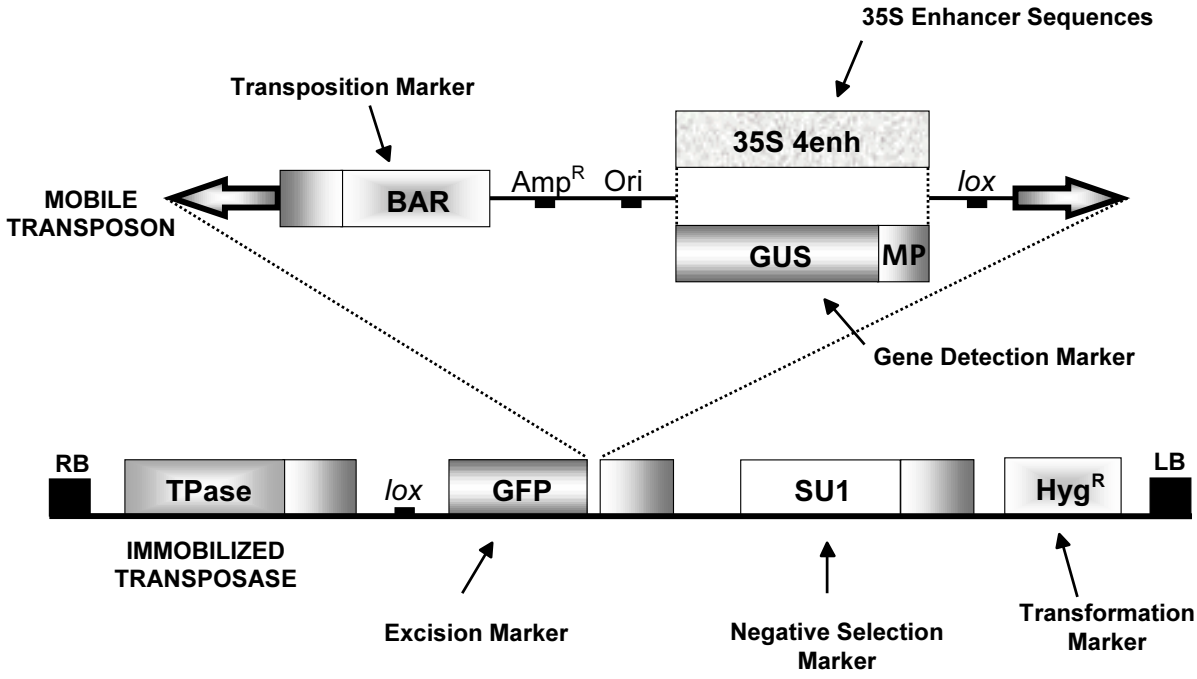


Figure 4. Model gene detection rice construct. The construct has two major parts: the mobile transposon (*Ds* or *I*) drawn above and inserted in the resident T-DNA drawn below. The mobile transposon, with transposon borders denoted by outward pointing thick solid arrows, contains a transposition marker (BAR gene) to select transposants as Basta^R (resistant). Plasmid origin and bacterial selection markers help in the recovery of large fragments of genomic-flanking DNA in *E. coli*. The gene detection marker for an enhancer trap consists of a GUS gene with a CaMV 35S minimal promoter (MP) located very near the transposon border. This is replaced by a multiple 35S enhancer (4enh) sequence to create an activation tag construct. To create chromosomal deletions, *lox* sites (*Cre-lox* system) are located in the mobile transposon as well as in the resident T-DNA. The T-DNA contains the immobilized transposase (*Ac* or *En*) under control of a strong promoter. The mobile transposon is inserted in the excision marker green fluorescent protein (GFP) gene to select transpositions. The transformation marker Hyg^R is used for selection of transformants using hygromycin and a negative selection marker, the SU1 gene, used to select for stable transposants. LB = left border, RB = right border, Ori = origin, Amp^R = ampicillin resistance.

Insertional mutagenesis in rice genomics

The rice genome sequence will probably uncover about 30,000 genes, half of which will have no known function. Transposon mutagenesis, using knockout and gene detection insertions, is an important tool for discovering these gene functions by reverse genetics strategies. In *Arabidopsis*, about 100,000 random inserts (Krysan et al 1999) are required for genome saturation and for rice about four times that number. Multiple independent inserts per plant, averaging four in many of the *Ac* and *Ds* lines, will decrease the required number of plants. The insertional preference of *Ac* for genes as described here can reduce the required number further by a factor of 3–5. A population of about 25,000 *Ac* or *Ds* plants, with multiple inserts, would therefore be sufficient to recover knockout or gene detection inserts for most genes.

Transposon populations for single stable *Ds* insertions can also be generated from a minimum of 10 active lines, as shown in the strategy outlined in Figure 5, in four generations after transformation and seed multiplication. This could produce a population of around 150,000 inserts, sufficient for genome saturation assuming preferential transposition of *Ds* in genes. By a concerted international effort, this transposon library is being produced and will be made available to rice researchers worldwide.

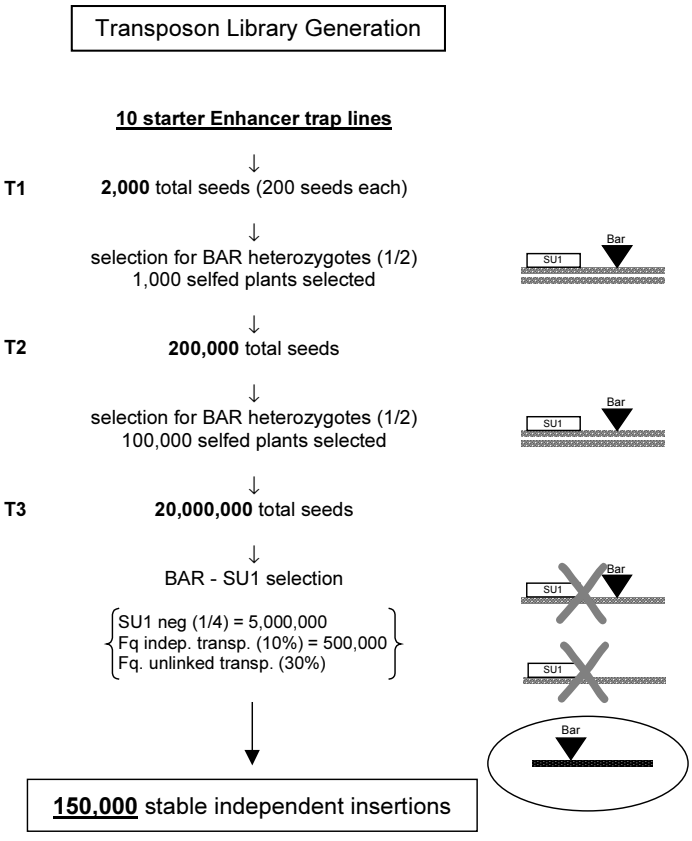


Figure 5. Transposon library generation. Using a minimum of 10 active single-locus starter enhancer trap lines, with a multiplication rate of 200 seeds per plant (Nipponbare), a total of 200,000 T_2 seeds can be obtained from heterozygous plants that are identified on the basis of progeny-test segregation for Basta^R. Similarly, from the 200,000 plants, half are heterozygotes and are identified by progeny tests in the T_3 generation, in which 20 million seeds are produced from the heterozygous parents and can be used for selection of the population of stable transposants. A sketch of the T-DNA carrying the SU1 and BAR (on transposon) markers is shown alongside. Using Basta^R and SU1 selection, only the plants bearing a Basta^R transposed element (BAR$^+$ SU1$^-$) are selected, whereas all SU1-bearing plants (with transposase) are rejected. The frequency of independent transposition was observed to be 10% and that of unlinked transposition about 30%, which will generate about 150,000 stable independent insertions.

LITERATURE CITED

Baldwin D, Crane V, Rice D. (1999). A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr Opin Plant Biol 2:96-103

Baker B, Schell J, Lörz H, Fedoroff N. (1986). Transposition of the maize controlling element 'Activator' in tobacco. Proc Natl Acad Sci 83:4844-4848

Baker B, Coupland G, Fedoroff N, Starlinger P, Schell J. (1987). Phenotypic assay for excision of the maize controlling element *Ac* in tobacco. EMBO J 6:1547-1554

Balcells L, Coupland G. (1994). The presence of enhancers adjacent to the *Ac* promoter increases the abundance of transposase mRNA and alters the timing of *Ds* excision in *Arabidopsis*. Plant Mol Biol 24:789-798

Burns N, Grimwade R, Macdonald PB, Choi E-Y, Finberg K, Roeder GS, Snyder M. (1994). Large-scale analysis of gene expression, protein localization, and gene disruption in *Saccharomyces cerevisiae*. Genes Dev 8:1087-1105

Chin HG, Choe MS, Lee S-H, Park SH, Koo JC, Kim NY, Lee JJ, Oh BG, Yi GH, Kim SC, Choi HC, Cho MJ, Han CD (1999). Molecular analysis of rice plants harboring an *Ac*/*Ds* transposable element-mediated gene trapping system. Plant J 19:615-623

Chiu W, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J. (1996). Engineered GFP as a vital reporter in plants. Curr Biol 6:325-30

Christou O, Ford T, Kofron M. (1991). Production of transgenic rice (*Oryza sativa* L.) from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into immature zygotic embryos. Biotechnology 9:957-962

Davis SJ, Vierstra RD. (1998). Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol Biol 36:521-528

Dove A. (1999). Proteomics: translating genomics into products. Nat Biotechnol 17:233-236

Enoki H, Izawa T, Kawahara M, Komatsu M, Koh S, Kyozuka J, Shimamoto K. (1999). *Ac* as a tool for the functional genomics of rice. Plant J 19:605-613

Gale MD, Devos KM. (1998). Plant comparative genetics after 10 years. Science 282:656-659

Greenblatt IM. (1984). A chromosome replication pattern deduced from pericarp phenotypes resulting from movements of the transposable element Modulator, in maize. Genetics 108:471-485

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

Hirochika H. (1997). Retrotransposons of rice: their regulation and use for genome analysis. Plant Mol Biol 35:231-240

Izawa T, Miyazaki C, Yamamoto M, Terada R, Iida S, Shimamoto K. (1991). Introduction and transposition of the maize transposable element *Ac* in rice (*Oryza sativa* L.). Mol Gen Genet 227:391-396

Izawa T, Ohnishi T, Nakano T, Ishida N, Enoki H, Hashimoto H, Itoh K, Terada R, Wu C, Miyazaki C, Endo T, Iida S, Shimamoto K (1997). Transposon tagging in rice. Plant Mol Biol 35:219-229

Jones JDG, Carland FM, Lin E, Ralston E, Dooner HK. (1990). Preferential transposition of the maize element *Activator* to linked chromosomal locations in tobacco. Plant Cell 2:701-707

Koprek T, McElroy D, Louwerse J, Williams-Carrier R, Lemaux PG. (1999). Negative selection systems for transgenic barley (*Hordeum vulgare* L.): comparison of bacterial *codA* and cytochrome P450 gene-mediated selection. Plant J 19:719-726

Krysan PJ, Young JC, Sussman MR. (1999). T-DNA as an insertional mutagen in *Arabidopsis*. Plant Cell 11:2283-2290

Lin X, Kaul S, Rounsley S, Shea TP, Benito MI et al. (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402:761-768

Matsumura H, Nirasawa S, Terauchi R. (1999). Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). Plant J 20:719-726

Mayer K, Schuller C, Wambutt R, Murphy G, Volckaert G et al. (1999). Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. Nature 402:769-777

Murai N, Li ZJ, Kawagoe Y, Hayashimoto A. (1991). Transposition of the maize *Activator* element in transgenic rice plants. Nucleic Acids Res 19:617-622

O'Keefe DP, Tepperman JM, Dean C, Leto KJ, Erbes DL, Odell JT. (1994). Plant expression of a bacterial cytochrome P450 that catalyzes activation of a sulfonylurea pro-herbicide. Plant Physiol 105:473-482

Pereira A. (1998). Heterologous transposon tagging systems. In: Lindsey K, editor. Transgenic plant research. Harwood Academic Publishers, Amsterdam, The Netherlands. pp 91-108

Pereira A, Saedler H. (1989). Transpositional behaviour of the maize *En/Spm* element in transgenic tobacco. EMBO J 8:1315-1321

Phang S-Z, DeBoer DL, Wan Y, Ye G, Layton JG, Neher MK, Armstrong CL, Fry JE, Hinchee MAW, Fromm ME. (1996). An improved green fluorescent protein gene as a vital marker in plants. Plant Physiol 112:893-900

Quackenbush J, Liang F, Holt I, Pertea G, Upton J. (2000). The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res 28:141-145

Shimamoto K, Miyazaki C, Hashimoto H, Izawa T, Itoh K, Terada R, Inagaki Y, Iida S. (1993). Trans-activation and stable integration of the maize transposable element *Ds* cotransfected with the *Ac* transposase gene in transgenic rice plants. Mol Gen Genet 239:354-360

Skarnes WC. (1990). Entrapment vectors: a new tool for mammalian genetics. Biotechnology 8:827-831

Sundaresan V, Springer P, Volpe T, Haward S, Jones JDG, Dean C, Ma H, Martienssen R. (1995). Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev 9:1797-1810.

Swinburne J, Balcells L, Scofield SR, Jones JDG, Coupland G. (1992). Elevated levels of *Activator* transposase mRNA are associated with high frequencies of *Dissociation* excision in *Arabidopsis*. Plant Cell 4:583-595

Trethewey RN, Krotzky AJ, Willmitzer L. (1999). Metabolic profiling: a Rosetta Stone for genomics. Curr Opin Plant Biol 2:83-85

Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, Nguyen JT, Sato S, Wang ZY, Xia Y, Dixon RA, Harrison MJ, Lamb CJ, Yanofsky MF, Chory J. (2000). Activation tagging in *Arabidopsis*. Plant Physiol 122:1003-1013

Wilson K, Long D, Swinburne J, Coupland G. (1996). A *Dissociation* insertion causes a semidominant mutation that increases expression of *TINY*, an *Arabidopsis* gene related to *APETALA2*. Plant Cell 8:659-671

Chapter 3

Early and multiple *Ac* transposition in rice suitable for efficient insertional mutagenesis.

Raffaella Greco[1], Pieter B.F. Ouwerkerk[2], Anke J.C. Taal[2], Cristina Favalli[3], Thierry Beguiristain[4], Pere Puigdomènech[4], Lucia Colombo[3], J. Harry C. Hoge[2†] and Andy Pereira[1]

[1] *Plant Research International, Wageningen, The Netherlands*
[2] *Institute of Biology, Leiden University, Leiden, The Netherlands*
[3] *Dept. of Biology, University of Milan, Milano, Italy*
[4] *CID-CSIC, Barcelona, Spain*

Plant Molecular Biology 46: 215-227 (2001)

ABSTRACT

A GFP excision assay was developed to monitor the excision of *Ac* introduced into rice by Agrobacterium mediated transformation. The presence of a strong double enhancer element of the CaMV 35S promoter adjacent to the *Ac* promoter induced very early excision, directly after transformation into the plant cell, exemplified by the absence of *Ac* in the T-DNA loci. Excision fingerprint analysis and characterisation of transposition events from related regenerants, revealed an inverse correlation between the number of excision events and transposed *Ac* copies, with single early excisions after transformation generating *Ac* amplification. New transpositions were generated at a frequency of 15-50% in different lines, yielding genotypes bearing multiple insertions, many of which were inherited in the progeny. The sequence of DNA flanking *Ac* in three representative lines provided a database of insertion tagged sites suitable for the identification of mutants of sequenced genes that can be examined for phenotypes in a reverse genetics strategy to elucidate gene function. Remarkably, two-thirds of *Ac* tagged sites showing homology to sequences in public databases were in predicted genes. A clear preference of transposon insertions in genes that are either predicted by protein coding capacity or by similarity to ESTs suggests that the efficiency of recovering knockout mutants of genes could be about three times higher than random. Linked *Ac* transposition, suitable for targeted tagging was documented by segregation analysis of a crippled *Ac* element and by recovery of a set of six insertions in a contiguous sequence of 70 kb from chromosome 6 of rice.

***Key words**: Ac transposon, CaMV 35S enhancer, functional genomics, insertion specificity, rice, transposon amplification.*

The nucleotide sequences data reported are deposited in GenBank under accession numbers from AZ923008 to AZ923053.

Dedication. We dedicate this report to the loving memory of our colleague Dr J. Harry C. Hoge who passed away during the course of this research.

INTRODUCTION

Transposons are versatile molecular tools for the isolation and identification of genes that display a mutant phenotype when inactivated. Mutant alleles which are "tagged" with the transposon, can be molecularly isolated by homology to the cloned transposon-tag sequences. This forward genetics technique of screening for mutants and identification of the corresponding gene has been termed transposon tagging (Bingham *et al*., 1981). In the contemporary genomics revolution, reverse genetics strategies are being developed using insertion sequences to elucidate the function of genes discovered by sequencing. The available sequence of the rice genome, a model for cereal crops, reveals many new genes whose functions can now be elucidated using appropriate insertional mutagenesis tools.

The maize autonomous *Activator* (*Ac*) transposon can transpose by itself and also induce the transposition of non-autonomous *Dissociation* (*Ds*) transposons (McClintock, 1947). These transposons have been shown to transpose in a wide variety of heterologous plants (reviewed in Haring *et al*., 1991; Sundaresan, 1996), where the behaviour of *Ac* has been rather unpredictable and subsequently revealed new insights in transposon biology. By virtue of its simplicity, *Ac* has been successfully used for transposon tagging in heterologous plants like petunia, tobacco, Arabidopsis and flax (Chuck *et al*., 1993; Whitham *et al*., 1994; James *et al*., 1995; Lawrence *et al*., 1995). Certain features of *Ac-Ds* transposition have been conserved between maize and heterologous hosts like preferential transposition to linked positions (Jones *et al*., 1990) that can be utilised to efficiently tag linked genes (Jones *et al*., 1994, James *et al*., 1995). In maize the transposition rate is inversely proportional to *Ac* dosage, while in heterologous systems higher copy numbers can increase the frequency of transposition (Jones *et al*., 1989). The transposition of *Ac* has been shown in maize to be related to DNA replication (Greenblatt, 1984), which tends to increase the overall copy number. Amplification of *Ac* has been also observed in tomato (Yoder, 1990), a phenomenon that provides a means of obtaining multiple transposons for mutagenesis.

An assortment of phenotypic excision assays have been developed using marker genes containing transposon inserts, where excision can be monitored by restoration of marker gene activity (reviewed in Pereira, 1998). Transposon excision can be selected for by resistance to an antibiotic such as kanamycin (Baker *et al*., 1987) or visualised through markers like GUS (Finnegan *et al*., 1989), luciferase (Charng *et al*., 1995) and streptomycin (SPT) resistance (Jones *et al*., 1989) that help distinguish between somatic and germinal excision events. We describe here the use of the green fluorescent protein (GFP) as an excision marker in rice, convenient for use in many laboratories around the world where rice is grown and also with potential applications in many other plants.

Ac-Ds transposons were first introduced by electroporation into rice (Izawa *et al*., 1991; Izawa *et al*., 1997; Murai *et al*., 1991; Shimamoto *et al*.,1993) and shown to transpose. It was however noted that the two-component *Ac-Ds* system can get inactivated in later generations (Izawa *et al*., 1997), justifying a systematic investigation of *Ac-Ds* biology in rice. Recently the behaviour of *Ac* and *Ac-Ds* have been followed through three successive generations and revealed characteristics suitable for functional genomics strategies (Enoki *et al*., 1999; Nakagawa *et al*., 2000). A more advanced gene trapping system has also been described (Chin *et al*., 1999), where *Ds* elements can be employed to trap expressed genes. The development of a variety of efficient transposon mutagenesis systems in rice will help functional genomics strategies in this model plant. Here we address the phenomena of *Ac* amplification and insertional specificity, identifying rice genotypes containing actively transposing multiple transposons. These selected genotypes can be used to develop populations of rice plants with enough insertions to efficiently saturate the genome and identify mutations in genes of interest by reverse genetics strategies.

MATERIAL AND METHODS

Construction of Ac:GFP vectors

The *Ac* constructs (Figure 1) were assembled by multi-point ligations, in which the individual fragments (promoter, GFP gene, *Ac* element) with appropriate compatible cohesive ends were ligated together to the binary vector in one reaction. A CaMV35S promoter fragment extending from –526 to the transcription start site, was obtained as a 0.55 kb *Hin*dIII-*Sal*I fragment from a pBR322 derivative of pDH51 (Pietrzak *et al*., 1986). The CaMV35S promoter and double enhancer (CaMV35Sde) including a synthetic untranslated leader sequence from Alfalfa Mosaic Virus RNA4 (AMV) was derived, after removal of the ATG initiation codon, as a 0.85 kb *Sst*I-*Sal*I fragment from pMOG18 (Sijmons *et al*., 1990). The entire autonomous *Ac* element was cloned as two fragments from a derivative of pKU2 (Baker *et al*., 1987): a 3.5 kb *Sal*I-*Xba*I fragment containing the 5' region till position 3426 and a 1.1 kb *Xba*I-*Bam*HI fragment extending from position 3426 to the end of the element. The two GFP variants employed, the soluble–modified GFP (sm-GFP; Davis and Vierstra, 1998) and the humanised red-shifted GFP (sGFPS65T; Chiu *et al*., 1996), were derived as 1 kb *Bam*HI-*Eco*RI GFP-tNos fusions. The constructs were made in the binary vector pMOG22 (ZENECA-MOGEN, NL) which contains a chimaeric CaMV 35S-hygromycin phosphotranferase-tNos for selection during transformation.

The 35S-smGFP:*Ac* construct was assembled by ligating the CaMV35S promoter (*Hin*dIII-*Sal*I fragment), the *Ac* element (*Sal*I-*Xba*I and *Xba*I-*Bam*HI fragments) and the smGFP-tNos cassette (*Bam*HI-*Eco*RI fragment) in between the *Hin*dIII and *Eco*RI sites of pMOG22. The 35Sde-sGFP:*Ac* construct was similarly assembled by ligating the CaMV35Sde (*Sst*I-*Sal*I fragment), the *Ac* element (see above) and the sGFP-tNos cassette (*Bam*HI-*Eco*RI fragment) in pMOG22 as above.

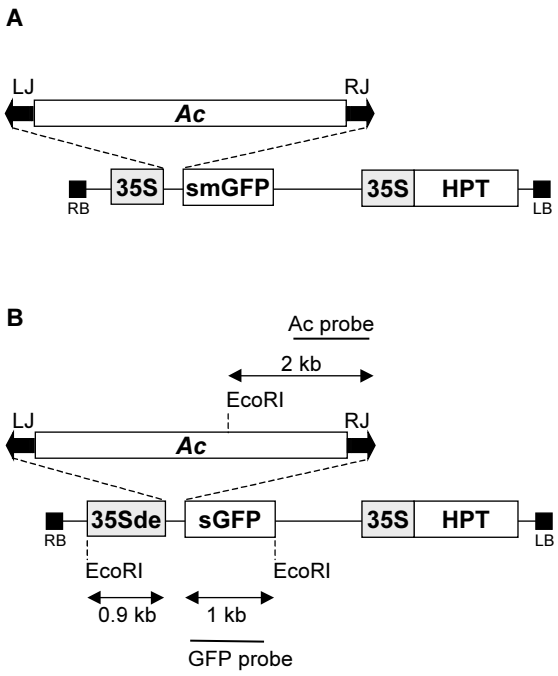


Figure 1. Schematic representation of 35S-smGFP:*Ac* (A) and 35Sde-sGFP:*Ac* (B). In both constructs the autonomous *Ac* element is inserted between the CaMV35S promoter and the GFP gene used as excision markerso that expression of GFP is restored upon *Ac* excision. The orientation of *Ac* is such that the endogenous promoter is located near the CaMV35S promoter and drives transcription of the transposase gene in the same direction. The hygromycin gene is used to select for transformation. In B are also shown the probes used for Southern analysis to detect excision (GFP probe) and transposition (*Ac* probe), and the size of the expected hybridizing fragments on genomic DNA restricted with *Eco*RI. RB, LB = T-DNA right and left border; HPT = hygromycin phosphotranferase gene; 35S = CaMV35S promoter; 35Sde = doubly enhanced CaMV35S promoter; Rj, Lj = *Ac* right and left junction.

Agrobacterium transformation of rice

Embryogenic calli were induced on scutella from germinated seeds (Rueb *et al*., 1994) and transformed with *Agrobacterium tumefaciens* LBA4404 harbouring the desired binary vector essentially as described by Hiei *et al*. (1994). In the first transformation experiment, transgenic calli were selected on 50 mg/l hygromycin before transfer on embryo-induction medium containing 75 mg/l hygromycin. The regenerated plantlets were transferred to the greenhouse and grown in hydroponic culture with a regime of 12 hours light, 28ºC, 85% relative humidity (RH) and 12 hours dark, 21ºC, 60% RH. The concentration of hygromycin during embryo-induction was increased to 100 mg/l in the next transformation experiments, and regenerated plantlets were additionally screened on 25 mg/l hygromycin.

Two transformation experiments were carried out with *Oryza sativa* ssp *japonica* cv. Taipei 309, using 35S-smGFP:*Ac* and 35Sde-sGFP:*Ac*. In a third transformation experiment, 35Sde-sGFP:*Ac* was introduced in *O.sativa* ssp *japonica* cv. Nipponbare.

GFP expression analysis

GFP expression in calli or transgenic plants was analysed using a binocular (Leica MZ FluoIII) with a fluorescent light source and a Leica GFP-plus (λ excitation=480/40, λ emission=510 LP barrier filter) and plant-GFP filters sets (λ excitation=470/40, λ emission=525/50 LP barrier filter). Images of plant tissues were taken using a colour video camera (Sony-iris with integration unit, Sony DKR700) and processed using Adobe Photoshop.

Molecular analysis of Ac transposition

Leaf samples (1-2 per plant) were collected either from *in vitro* young plantlets after regeneration and before transfer to the greenhouse, or from adult greenhouse plants. Genomic DNA was isolated by a miniprep procedure (Pereira and Aarts, 1998).

Regenerated transformants were first checked by PCR using hygromycin phosphotransferase (HPT) specific primers (HPT-forward: 5'-AAAAGTTCGACAGCGTCTCCGACC and HPT-reverse: 5'-TCTACACAGCCATCGGTCCAGACG) to confirm the presence of the T-DNA. Excision events were monitored by amplification of the empty donor site (EDS) using a primer annealing to the 35S promoter (35S-forward: 5'-ATCCCACTATCCTTCGCAAGACCC) in combination with a primer annealing to the GFP gene (smGFP-reverse: 5'-GAAAGGGCAGATTGTGTGGACAGG or sGFP-reverse: 5'-GCTTGTCGGCCATGATATAGACG). An amplification product is detected only if *Ac* excised from the T-DNA. To confirm excision, the EDS fragments were cloned in pGEM-T Easy (PROMEGA) and sequenced.

The full donor site (FDS) corresponding to *Ac* at its original position in the T-DNA was also examined. The FDS-left junction was amplified using primers from the 35S promoter (see above) and the 5' end of *Ac* (*Ac*282-reverse: 5'-CTCAGTGGTTATGGATGGGAGTTG), while for the FDS-right junction primers from the 3'region of *Ac* (*Ac*4374-forward: 5'-GAACAAAAATACCGGTTCCCGTCC) and the GFP gene (smGFP or sGFP, see above) were used. The presence of *Ac* in the genome was determined by PCR with internal *Ac* primers (*Ac*971-forward: 5'-ACGACTCCATTCCTCAGATGACG and *Ac*1395-reverse: 5'-CTTGACTCGGATCTGTAGCTGTACC). All PCR reactions were performed using 50 ng of genomic DNA and standard amplification conditions.

DNA blot hybridisation

3 µg of genomic DNA were restricted with *Eco*RI, separated overnight on a 0.8% agarose gel in 1X TAE at 4°C and vacuum-blotted onto Hybond-N^+ membranes. Hybridisation was performed as described previously (Aarts *et al*., 1995). To detect excision, a 0.75 kb fragment of the GFP gene was used as probe. A 1.1 kb *Xba*I-*Bam*HI fragment from the 3' region of *Ac*, was used as probe to reveal *Ac* positions in the genome. The T-DNA left border specific probe was a 0.9 kb internal fragment from the HPT gene. As a T-DNA right border specific probe, a CaMV 35S promoter fragment extending from -526 till the *Eco*RV site at position –93, was used on genomic DNA digested with *Eco*RV.

Isolation of Ac insertion tagged sites

Genomic DNA fragments flanking the left junction of *Ac* were isolated by inverse-PCR (Earp *et al*., 1990). 200-400 ng of genomic DNA was restricted separately with *Hae*III or *Nla*III, self- ligated and linearized again using *Bam*HI. Flanking DNA was then recovered by PCR amplification using two *Ac* specific primers (*Ac*44-reverse: 5'- GATAACGGTCGGTACGGGAT and *Ac*197-forward: 5'- CGGGATGATCCCGTTTCGTT). The inverse-PCR products were cloned in pGEM-T Easy (PROMEGA) and clones of different sizes sequenced. Similarity to known sequences in public databases was determined using the BLAST algorithm (Altschul *et al*., 1997).

RESULTS

Development of a GFP excision assay

To develop a transposon excision assay using GFP, two constructs were made in which the maize *Ac* transposon was cloned in the transcriptional leader of the GFP gene. Figure 1A displays a CaMV 35S promoter-smGFP fusion (35S-smGFP:*Ac*) and Figure 1B shows a duplicated enhancer CaMV 35S promoter-sGFP fusion (35Sde-sGFP:*Ac*).

Both constructs were introduced into rice by *Agrobacterium* mediated transformation and the regenerating calli examined for GFP fluorescence. In the first transformation experiment, using the 35S-smGFP:*Ac* construct, no GFP activity was visible among the hygromycin-selected calli. Molecular analysis of regenerated plants revealed that only the regenerants from one callus were transformed.

A second transformation experiment was performed using the 35Sde-sGFP:*Ac* construct, with an improved GFP variant and modified transformation selection conditions. A large number of hygromycin resistant calli were obtained, which were selected for GFP activity (Figure 2A).

Plants were regenerated from these calli and showed GFP expression in most tissues examined. Due to the selection procedure, we did not expect to see variegation for GFP activity. To visualise variegation for GFP expression due to transposon excision, we performed a third transformation experiment using the same construct, in which no pre-selection for GFP expression was applied. The hygromycin resistant calli yielded multiple regenerants that were grown in the greenhouse till maturity and displayed GFP activity at various stages (Figure 2B-F). Uniform high expression was observed generally in the roots and a few high-expressing lines displayed activity in leaves, mostly in the midribs. Seeds displayed light GFP expression in the seed coat and high activity in the embryo. No clear evidence of GFP variegation was obtained, suggesting that early excision had occurred.

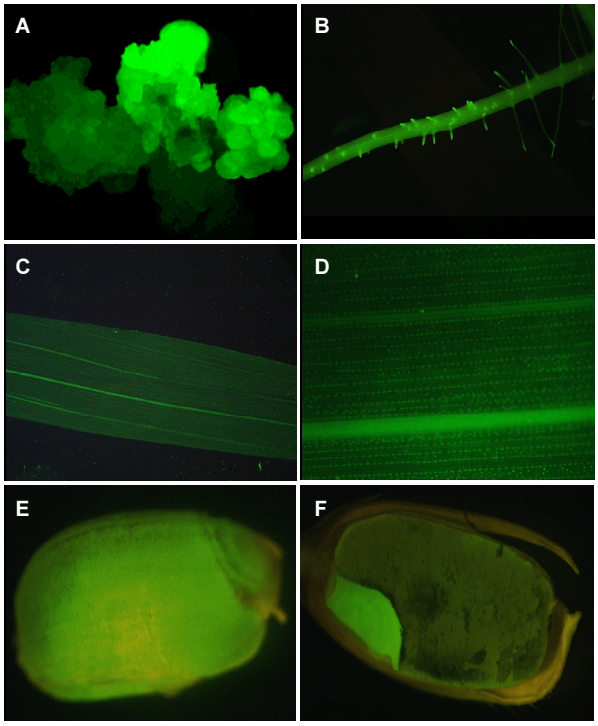


Figure 2. sGFP expression after *Ac* excision. GFP fluorescence was detected in regenerating calli (A), roots (B) and leaves of regenerated plants (C, D), T1 seeds (E) and embryos (F).

Ac excision in primary transformants is construct dependent

Multiple shoots were regenerated from each transformed callus and the T_0 regenerants tested by PCR and DNA blot hybridisation for the presence of all T-DNA components. In the first transformation experiment (35S-smGFP:*Ac* construct), only one transformed callus with nine regenerants was obtained. All regenerants displayed a full donor site (FDS) and an empty donor site (EDS) by PCR, while only some plants revealed a clear EDS by genomic DNA blot hybridisation (data not shown), suggesting variable chimeric nature of excision events. The EDS

fragments from each regenerant were isolated by PCR, cloned and sequenced. The six different footprints after *Ac* excision, displayed in Figure 3, indicate multiple independent excision events that occurred late during the development of the individual plants. Some regenerants indeed revealed identical footprints, indicating that they were probably regenerated from clonal cells arising from the same excision event.

waxy (maize)			GCGTGACC
Ac at *wx-m7*			GCGTGACC - *Ac* - GCGTGACC
		Construct 35S-smGFP:*Ac*	
Line	Ac copies	# Plants	Footprint
10	single	1	GCGTGACC ––––––––
		1	GCGTGAC – – –TGACC
		3	GCGTGACG CCGTGACC
		1	GCGTGACC ––GTGACC
		1	GCGTGA–– ––––GACC
		2	GCGTGACG GCGTGACC
		Construct 35Sde-sGFP:*Ac*	
Line	Ac copies	# Plants	Footprint
1	single	1	GCGTG ––– CCGT GACC
		1	GCGTGACG CCGTGACC
		1	GCGTGAC – – – GTGACC
2A	multiple	all 10	GCGTGACG CCGTGACC
2B	lost	all 7	GCGTGACG GCGTGACC
4A	single	1	GCGTGACG GTGTGACC
4B	single	1	GCGTGACG – – GTGACC
5	single	4	GCGTGACG CCGTGACC
		2	GCGTGACG GCGTGACC
		2	GCGTGAC – – – GTGACC
		1	GCGTGACG G CCGTGACC
7	multiple	all 9	GCGTGACG CCGTGACC

Figure 3. EDS footprints. Empty donor sites in primary regenerants transformed with the 35S-smGFP and 35Sde-sGFP constructs were amplified and sequenced as described in Material and Methods. Molecular footprints left after *Ac* excision are indicated as alterations of the target site duplication sequence of the maize *waxy-m7* allele, from which *Ac* was originally cloned. Nucleotide substitutions or insertions are underlined, while dashes indicate nucleotide deletions. The number of *Ac* copies as identified by Southern hybridization and the number of plants bearing the same footprint are shown for each independent transformed line.

The second transformation experiment (using the 35Sde-sGFP:*Ac* construct) yielded 11 transformed lines with 45 plants regenerated after selection for GFP activity. All the regenerants revealed only EDSs and no FDSs (100% excision), with 73% still showing presence of an *Ac* element by PCR with internal primers. The absence of an FDS in all regenerants could be accounted due to the calli being selected for GFP activity.

A more detailed analysis was performed on the regenerants unselected for GFP activity from the third transformation experiment. DNA blot hybridisation using probes for both sides of

the T-DNA allowed classification of the regenerants into 7 independent transformed lines and identification of single or multiple copy transformants. The transformants with an incomplete or rearranged T-DNA insert were removed from the analysis. Most surprisingly, almost every regenerant (41 out of 42 tested) displayed only EDSs and no FDSs, indicating very early excision in 97.6 % of the plants. DNA blot hybridisation shown in Figure 4, revealed only the 1.9 kb EDS expected fragment and no FDS. *Ac* was still present in 70.7% (29/41) of the regenerants demonstrating the complete transposition process.

To characterise the nature of the early excision events in these transformants, EDSs from the 41 regenerants were isolated and sequenced (Figure 3). One group of transformed lines displayed a few different EDSs, as exemplified by line 5 with four different footprints among the ten regenerants. The presence of a complete EDS allele and no FDS in these plants suggests that the few excisions took place in the first few cells after transformation. Most surprisingly another group of transformed lines displayed only a single footprint amongst all regenerants, as illustrated by the ten regenerants of lines 2A and 7. This indicates that in these lines only one early *Ac* excision event took place from the T-DNA in the primary cell after transformation and before further division.

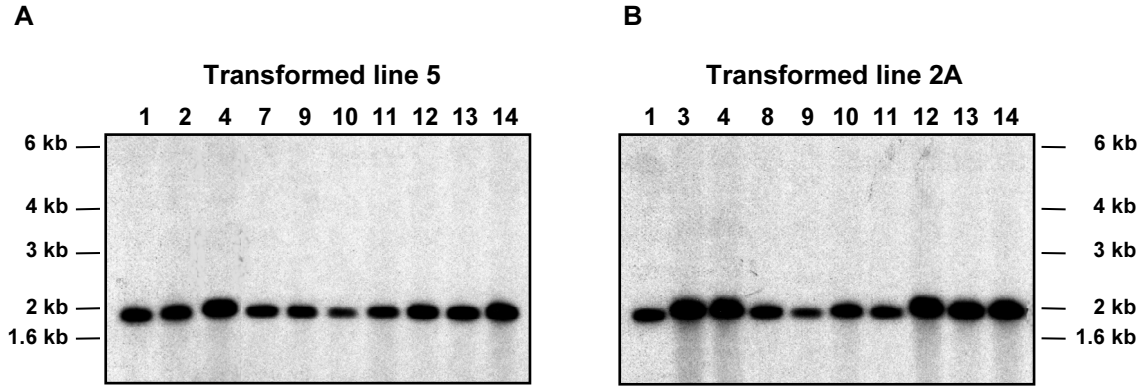


Figure 4. Excision analysis in regenerants from construct 35Sde-sGFP:*Ac*. DNA blot hybridisation of 10 regenerants of line 5 (A) and line 2A (B) digested with *Eco*RI and probed with the sGFP gene. *Ac* excision reveals an EDS fragment of 1.9 kb, between *Eco*RI sites upstream of the 35S promoter and downstream of the GFP-tNos cassette (Figure 1). The FDS (absent) with *Ac* in the T-DNA would generate a 3 kb band, due to the *Eco*RI site at position 2487 in *Ac*.

Multiple transpositions in regenerants

The regenerants from the third transformation experiment, unselected for GFP activity, were analysed further by DNA blot hybridisation with an *Ac* specific probe. Common early transpositions and unique late transpositions were evident, with some lines displaying multiple *Ac*

insertions and others single or no *Ac* re-insertions. Figure 5 shows the results of two single-copy T-DNA lines, containing single and multiple transposed *Ac* inserts (line 5 and 2A respectively). A remarkable inverse correlation is evident between the number of *Ac* copies in a line and the number of footprints revealed (outlined in Figure 3). The two lines with multiple *Ac* elements (Figure 5 and data not shown) contain a single footprint each, while the lines with more EDS footprints contain a single transposing *Ac*.

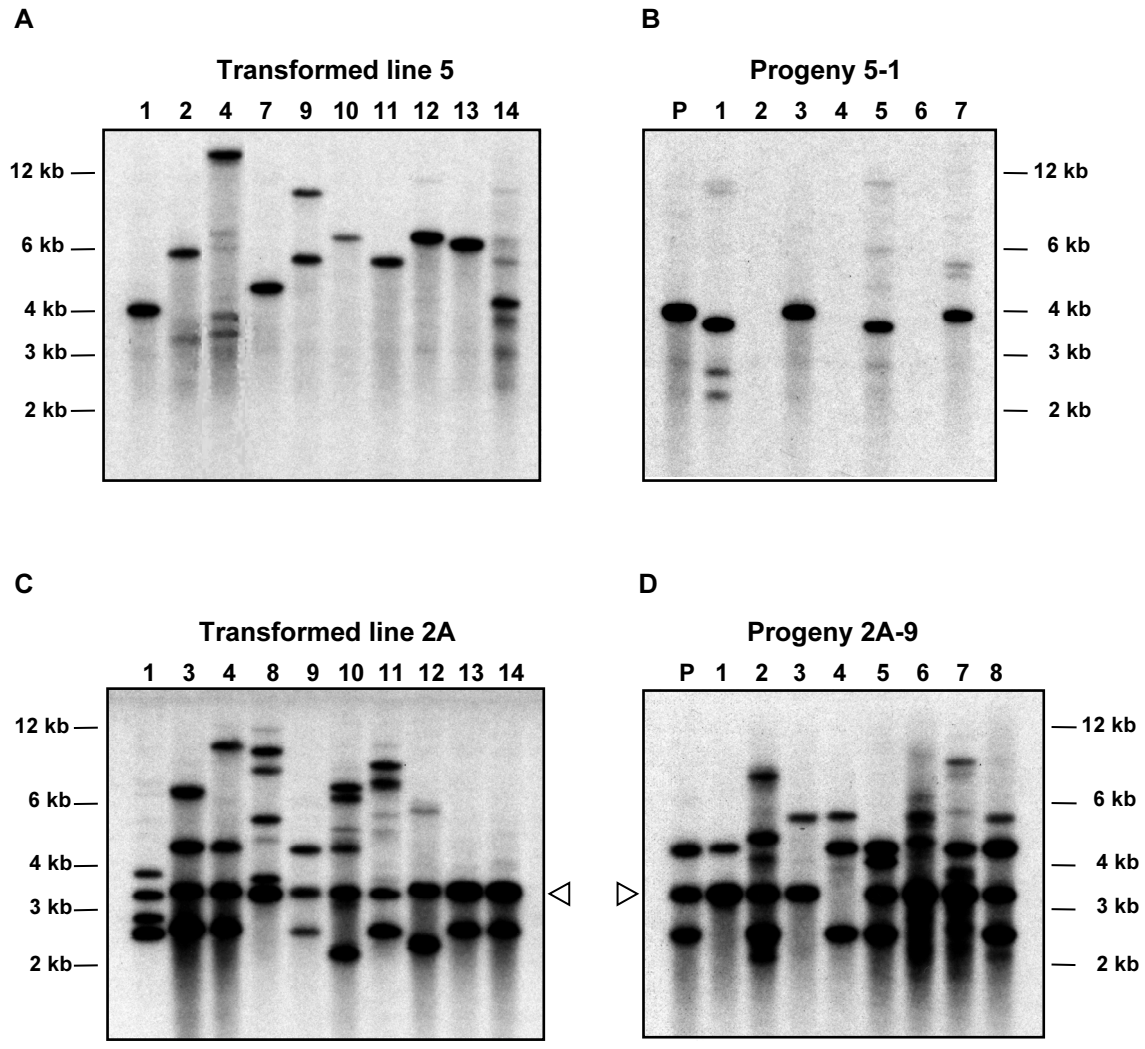


Figure 5. Transposition analysis in different regenerants and progeny plants transformed with construct 35Sde-sGFP:*Ac*. (A, C) Blots from regenerants of line 5 and 2A shown in Figure 4 were hybridised with a 1.1 kb *Ac* 3' region probe. A predicted FDS of 3 kb is not visible but other fragments (larger than 2 kb) originate from reinsertion of *Ac* at new positions in the genome. (B, D) DNA blot from progeny plants derived from regenerant 5-1 and 2A-9 respectively, hybridised with the *Ac* probe as above. The 3.4 kb hybridising fragment corresponding to the crippled *Ac* in C and D is marked by an arrow head.

To estimate the transpositional activity in the transformants, the frequency of unique transposition events among the total inserts in a set of related plants (regenerants or seed progeny) was estimated, termed as ”independent transposition frequency” (ITF; Aarts *et al*., 1995). Transposed *Ac* insertions that were present in only a single plant were considered as unique late events that were useful to produce different insertions in the progeny. The somatic ITF calculated over three populations of regenerants is given in Table 1, and varies between 49% to 64%. The regenerants from line 5 with a single *Ac* element reveal an ITF of 64 %. The plants from line 2A and line 7, although from a single copy T-DNA, displayed multiple *Ac* elements with ITFs of 49% and 61% respectively.

The evidence of active transposition in regenerating calli led us to examine somatic transpositions within a mature plant. From two lines with multiple inserts, six leaf samples of greenhouse grown plants were analysed. DNA blot hybridisation with an *Ac* probe, revealed that the transposon insertion pattern was very uniform (data not shown). This indicates that once transposition occurred during regeneration from the callus, the mature plant did not display significant further transposition.

Table 1. Independent transposition frequency (ITF) of *Ac* in different regenerants and progeny plants of construct 35Sde-sGFP:*Ac*.

	Regenerants			Progeny		
Transformed line	2A	5	7	2A-9	5-1	7A-1
T-DNA copies	1	1	2	1	1	2
Ac copies[a]	2-8	1	3-7	2-5 (3)	1 (1)	3-7 (4)
Total number of plants	10	10	10	8	7	12
Plants with unique inserts	7	7	5	6	2	5
Total number of inserts	45	11	23	32	6	47
Unique inserts	22	7	14	11	3	7
ITF[b]	49%	64%	61%	34%	50%	15%

[a] The minimal and the maximal number of *Ac* hybridizing fragments in each set of plants. Within brackets the number present in the parental plant.
[b] The ITF is calculated as the percentage of unique inserts amongst the total number of inserts.

Analysis of transposition in progeny

To study the inheritance of transposition events to the next generation and estimate the frequency of new insertions, selfed seed progeny of the three lines examined for somatic transpositions were analysed by blot hybridisation with an *Ac* probe (Figure 5). Hybridisation of the same blot with a T-DNA probe revealed plants carrying the T-DNA and enabled cosegregation analysis with different inserts.

The progeny of line 5-1 carried a single insert each (Figure 5D). Three new inserts were unique (Table 1) representing an ITF of 50% and indicating that each plant had a 50% chance of

generating a new insert. Progeny of line 2A-9 with multiple *Ac* inserts (Figure 5B) revealed an ITF of 34 % considering all *Ac* hybridising fragments and 27 % with only the strongly hybridising *Ac* fragments, potentially inherited from the parent as germinal inserts. A 3.4 kb *Ac* homologous fragment cosegregated with the T-DNA, while two other inherited *Ac* elements were present in most progeny and some plants carried extra 1-3 inserts due to new transpositions. Line 7-1 with multiple *Ac* elements, displayed a lower ITF of 15% with one insert completely linked to the T-DNA (data not shown).

Database of sequenced insertion tagged sites

The active *Ac* lines with multiple independent transpositions offered a means to develop a population for insertional mutagenesis. DNA from representative regenerants analysed above (plant numbers listed in Table 2) carrying a variety of insertions, was used in inverse PCR (IPCR) reactions to amplify rice genomic DNA flanking the *Ac* insertions. The different sequences obtained were compared to public databases for similarity to known proteins, genomic sequences and ESTs. A summary of the sequences, their origin and homologues in databases is given in Table 2. This catalogue of *Ac* flanking DNA sequences was used to make an insertion tagged site (ITS) database, suitable for reverse genetics.

Some ITSs were recovered in more plants originating from a transformed line. A common insert from line 2A with homology to a BAC clone (accession AQ325528), turned out to have an *Ac* element with a 4-bp deletion in the left junction and should therefore be a stable insert. The stability of this crippled *Ac* is indeed evident by its presence in all the regenerants and the seed progeny of line 2A-9 where it is linked to the T-DNA as a 3.4 kb fragment, described in the section above (Figure 5C ,D).

Linkage between a number of transposed *Ac* inserts was revealed by a set of ITSs homologous to a contiguous genomic sequence in a PAC clone (accession AP000616) from chromosome 6 of rice. This comprised four ITSs from regenerant 2A-9 and two from regenerant 2A-11, clustered in a 70-kb interval. Remarkably, five of the six insertions were in predicted genes, either homologues of known proteins or ESTs.

The total frequency of ITSs in predicted genes, summarised in Table 3, was calculated to be 34.5% among all ITSs or 38.5% considering only that above 40-bp length, as smaller ITSs never displayed significant homology to databases and were therefore considered separately in calculating frequencies. This suggests that more than one third of all the *Ac* insertions are in genes that can be predicted on basis of homology to proteins or ESTs. ITSs in predicted genes like S-adenosyl-L-methionine:salicylic methyltransferase or Cyt-5 DNA methyltransferase can

Table 2. Insertion Tagged Sites with similarity to sequences in public databases.

Plant	ITSs[a]	Blast homology	Organism	Accession	Insertion	Identity (bp) / Similarity (aa)	%
2A-8	8	Genomic clone [b]	O.sativa	AQ325528	-	46/51 bp	[90%]
		Genomic clone	O.sativa	AQ796410	-	304/319 bp	[95%]
2A-9	8	Ferredoxin NADP-reductase[c]	O.sativa	BAA85425	ORF	13/13 aa	[100%]
		Nodulin N21-like protein[c, e]	O.sativa	BAA85439[e]	5th intron	46/46 bp	[100%]
		Nodulin N21-like protein[c]	O.sativa	BAA85440	ORF	51/57 aa	[89%]
		Hypothetical protein[c]	O.sativa	BAA85435	ORF	25/25 aa	[100%]
		Genomic clone[b]	O.sativa	AQ325528	-	46/51 bp	[90%]
2A-10	4	Genomic clone[b]	O.sativa	AQ325528	-	46/51 bp	[90%]
		Genomic clone	O.sativa	AQ861699	-	217/220 bp	[98%]
		Genomic clone	O.sativa	AP002883	-	128/139 bp	[92%]
2A-11	7	EST / Nodulin N21-like protein[c, e]	O.sativa	AU078645 / BAA85424[e]	3' UTR	112/112 bp	[99%]
		EST / Putative harpin-induced protein[f]	O.sativa / A.thaliana	BF430546 / AAD21461[f]	5' UTR	186/187 bp	[99%]
		EST	O.sativa	AU69979	-	42/43 bp	[97%]
		Genomic clone[b]	O.sativa	AQ325528	-	46/51 bp	[90%]
		Genomic clone[c]	O.sativa	AP000616	-	166/181 bp	[91%]
		Genomic clone	O.sativa	AP002069	-	356/373 bp	[95%]
5-1	6	RPR1 / EST	O.sativa	BAA75812 / AW070066	ORF	33/44 aa 112/134 bp	[74%] [83%]
		Genomic clone / Hypothetical protein[g]	O.sativa	AP002537 / BAB16867[g]	5' UTR	245/270 bp	[94%]
5-7	3	Unknown protein / Genomic clone	A.thaliana / O.sativa	AAC77868 / AQ074191	ORF	66/97 aa 291/303 bp	[67%] [96%]
		S-adenosyl-L-methionine:salicylic methyltransferase	A.thaliana	BAB08594	ORF	71/90 aa	[78%]
5-11	3	Genomic clone / Ribosome apurinic lyase[h]	O.sativa / T.aestivum	AQ691029 / BAA87875[h]	3' UTR	268/280 bp	[95%]
5-12	3	EF-hand cont. protein / EST	A.thaliana / O.sativa	BAB02809 / C98851	ORF	122/140 aa 210/226 bp	[88%] [92%]
		EST	O.sativa	C99498	-	171/176 bp	[97%]
7-1	6	Cyt-5 DNA methyltransferase (MET1) / EST	Z.mays / T.aestivum	AAG15406 / BE400695	ORF	42/42 aa 110/126 bp	[100%] [87%]
		EST / Ribosomal protein[d, i]	O.sativa / A.thaliana	AU031084 / AAB71459[i]	3' UTR	62/63 bp	[98%]
		Hypothetical protein	O.sativa	BAA81761	ORF	47/59 aa	[78%]
		Genomic clone	O.sativa	AP002901	-	34/36 bp	[94%]
7-2	4	EST / Ribosomal protein[d, i]	O.sativa / A.thaliana	AU031084 / AAB71459[i]	3' UTR	62/63 bp	[98%]
		Genomic clone / Hypothetical protein[l]	O.sativa / A.thaliana	AQ688452 / AAD32789[l]	ORF	59/59 bp	[100%]
		Genomic clone / Hypothetical	O.sativa	AP002843 /	ORF	52/52 bp	[100%]
7-3	6	Hypothetical protein / Genomic clone	O.sativa	BAB17166 / AP002843	ORF	14/16 aa 65 / 68 bp	[87%] [95%]
7-5	4	Genomic clone	O.sativa	AZ046446	-	229/266 bp	[86%]

[a] Number of different ITSs.
[b, d] Same *Ac* insertion.
[c] Ac insertions located on the same PAC clone P0514G12 on chromosome 6 (accession AP000616).
[e, g, m] Homology derived from AP000616 (chr. 6), AP002537 (chr. 1), AP002843 (chr. 1) respectively.
[f] Homology derived from BlastX of BF430546 (79/161 aa – 48%).
[h] Homology derived from BlastX of AQ691029 (67/109 aa – 60%).
[i] Homology derived from BlastX of AU031084 (32/33 aa – 96%).
[l] Homology derived from BlastX of AQ688452 (40/76 aa – 52%).

provide information on the unique function of these genes in rice. Interesting mutants are being followed up in the next generation for phenotypes to validate this reverse genetics strategy.

Table 3. Frequency of *Ac* insertions in predicted genes.

Insertion Tagged Sites	Line 2A	Line 5	Line 7	Total	Frequency[a]
Total ITS[b]	24 [23]	15 [13]	19 [16]	58 [52]	-
ITSs with homology to codogenic regions	7	7	6	20	34.5% [38.5%]
ITSs with homology to predicted proteins	6	6	6	18	31.0% [34.6%]
ITSs with homology to ESTs[c]	3	1	1	5	8.6% [9.6%]

[a] The frequency is calculated on basis of the total number of ITSs and on the number of ITSs larger than 40 bp (in brackets).
[b] Within brackets is indicated the number of ITSs longer than 40 bp.
[c] Only the number of identical ESTs is reported (bp identity>95%).

DISCUSSION

GFP Excision marker

A phenotypic excision assay using GFP was developed to monitor *Ac* excision in rice. The red-shifted sGFP gene under control of the CaMV 35S promoter with a double enhancer displayed expression in most tissues after *Ac* excision, demonstrating suitability for this use. In our research the extremely early *Ac* excision was first observed using the GFP assay and showed the prevalence of this phenotype in every transformant. However, variegation for *Ac* activity was not observed with these constructs. The GFP excision assay used directly after transformation offers a convenient way to quickly test effectivity of constructs at the regenerating callus level. Moreover, as GFP expression is easily visible in the seed, it is potentially suitable to select independent excision events. This would be applicable to many plants where seeds or small seedlings can be screened for GFP activity and independent germinal excisions selected.

Early excision behaviour

The presence of the strong enhancer adjacent to the *Ac* transposon conferred very early excision in almost every transformant, probably by influencing autonomous *Ac* transposase expression. This is exemplified by the remarkable absence of a FDS in almost every transformant. The enhancer effect we show is due to the duplicated domain B of the CaMV 35S promoter (Benfey and Chua, 1990). The effect of the CaMV enhancer on the *Ac* promoter and transposase expression has been reported earlier for a two-component *Ac-Ds* system in Arabidopsis (Balcells and Coupland, 1994), where 4-8% of seedlings displayed excision. In rice a higher level of excision conferred by a CaMV promoter-*Ac* transposase fusion has been

documented (Chin *et al*., 1999), however a FDS was always observed indicating chimeric tissue and later excision. Here we demonstrate that a strong enhancer cis to the autonomous *Ac* element always confers extreme instability displayed by very early excision after transformation. This observation is very relevant as it suggests that if *Ac* inserts in the genome near a strong constitutive enhancer, it would probably excise immediately leading to a very unstable mutant or non-recovery of such mutants. But as not all genes contain such strong enhancers, less unstable *Ac* transposition events will be predominantly obtained.

The prevalence of very early excision manifested by complete EDSs, raised the question whether *Ac* transposed once very early or later in every cell. *Ac* excision leaves behind unique footprints that are random nucleotide changes like deletions, additions or conversions, which can be used to determine independence of excision events. The footprints obtained by sequencing the EDSs from all the regenerants established a group of transformants with a single excision event and another group with a few early excisions (Figure 3). Remarkably, an inverse correlation is revealed between the number of excision events and the number of transposed *Ac* copies in the lines. Transformants with a single early excision contained multiple *Ac* insertions, while lines with more early excisions had single re-insertions.

Ac amplification

The generation of multiple *Ac* copies from a single early excision event led us to examine the nature of these *Ac* transposition events. In the regenerants of the single T-DNA copy line 2A, multiple insertions and uniform intensity of the 3-4 inserts suggested that they all arose as a result of early amplification. One insert is a crippled *Ac*, with a deleted left terminal repeat, that is present in every regenerant and selfed progeny plant and is closely linked to the T-DNA. A similar crippled *Ac* has been described in tobacco to be stable (Hehl and Baker, 1989). The prevalence of this crippled *Ac* in all regenerants and progeny suggests that it transposed in the first cell after transformation and is present in every cell of the transformed plants. Because it is transposition defective, this *Ac* could not have given rise to the other *Ac* transpositions that must have occurred before this event, but also in the same cell following transformation. Furthermore, the crippled *Ac* and two other inserts are inherited in individual progeny, implying that at least two of the *Ac* elements were present in the same premeiotic cell and transmitted through the same gamete. The incidence of multiple *Ac* inserts is therefore not only a result of chimaerism of different clonal insertion events but is due to presence of multiple insertions in the same cell, which might have originated from amplification occurring directly after transformation.

The process of transposition during replication has been described in maize (Greenblatt, 1984), in which *Ac* transposes from replicated DNA to unreplicated DNA, resulting in a moderate

increase in copy number. In the amplifications we observed, the presence of a single excision footprint suggests that transposition must have initiated from an unreplicated site.

An excision event occurring after transformation and before cell division from the unreplicated T-DNA site, would give the same excision footprint in every cell. Reinsertion of the excised *Ac* into a second site that subsequently gets replicated would generate a second *Ac* copy in the newly replicated strand. Simultaneous secondary transposition events in which both *Ac* copies move to two new unreplicated positions and get duplicated following replication, would give rise to four copies at the two chromosomal positions. A third cycle of transposition and replication would then produce four new *Ac* positions in the cell, as observed here.

Amplification has also been documented in tomato (Yoder, 1990) where multiple *Ac* transposition and replication cycles might also have occurred in the same cell. A correlation between high *Ac* transposition activity and rapid amplification has been proposed (Peterson and Yoder, 1995) which suggests that, in the case described here, amplification of *Ac* directly following transformation might have been induced by the adjacent strong enhancer. This has not been documented before in any transgenic transposon situation. Transcriptional control or induction of transposition has been shown in a number of cases (Swinburne *et al*., 1992; Scofield *et al*., 1993; Balcells *et al*., 1994), suggesting that high transposase expression could induce multiple transpositions in a cell.

Reverse genetics

The behaviour of *Ac* in rice is characterised by high transposition in regenerating tissue and in cells giving rise to the next generation and relatively low transposition in older plant tissue. The frequency of independent transpositions estimated by the ITF ranged from 15-50% in different progenies, extrapolating to a new transposition for every three inserts. Multiple independent transpositions can generate a population of insertion mutants that are accessible by forward as well as reverse genetic strategies.

Whole genome and EST sequence information uncovers many genes with no known function that can be addressed by examining the phenotype of loss of function mutations. We generated an insertion tagged site database containing sequences of DNA flanking *Ac* inserts in the genome. On comparison to genome databases, the position of inserts in sequenced genes can be precisely determined. This will become an efficient way to identify inserts in genes of interest once the genome is sequenced, as already the case in Arabidopsis (Parinov *et al*., 1999). We placed six ITSs within 70 kb of sequenced DNA on chromosome 6, exemplifying linked transpositions and demonstrating the power of this method to identify inserts in genes. As this chromosome will be systematically sequenced soon, the *Ac* lines we have generated will be very pertinent to obtain knockout mutants of genes in this region.

Insertion specificity

In the analysis of our insertion sequence tagged site database, a remarkably high frequency of inserts displayed similarity to predicted genes or proteins in the public databases. There are twenty ITSs homologous to predicted proteins or ESTs using Blast searches, amounting to about 34.5% of the total (or 38.5% considering only larger ITSs). Assuming that there are about 30,000 rice genes with an average 2.6 kb coding sequence (Jeon *et al*., 2000), 78 Mb or 18% of the genome codes for proteins. This means that 18% of random rice DNA fragments would be expected to show homology to predicted genes in the databases. The ITS homologies of 35-39% to known proteins indicates that about two times more ITSs are in genes than expected by a random sequence. However, not all rice genes would be predictable based on homology searches or available gene prediction programs, suggesting that our calculations are a minimal estimate.

Considering the ITSs homologous to ESTs, five are identical with more than 95% identity using a similar cut-off as taken for calculation of the Gene Index developed by TIGR (Quackenbush *et al*., 2000). The rest of the ESTs probably indicate homology to a family member, a redundancy calculated for Arabidopsis to be about two-third on basis of genome sequence information (The Arabidopsis Genome Initiative, 2000). From the about. 70,000 rice ESTs in the database, about 26,000 are different based on the Gene Index calculation (Quackenbush *et al*., 2000) and taking into account an average size of 400 bp sequence, an estimated 10.4 Mb of unique EST sequence is present in the public databases. In the rice genome with size 430 Mb, this EST sequence information amounts to 2.5% of the genome, meaning that random rice DNA sequences would show a 2.5% chance of being homologous to ESTs in the database. In contrast, the ITSs represented by identical homologous ESTs are 8.6-9.6% of the total ITSs and are therefore biased with about 3-4 times more insertions in transcription units of genes.

These simple calculations using the two methods above, reveal a consistent bias of about three times more insertions in genes and affirm that *Ac* inserts preferentially into genes in rice. These studies confirm the earlier results (Enoki *et al*., 1999) of *Ac* insertional preference in rice where 4% inserts in ESTs was observed, although a different number of available ESTs was taken into account compared to our estimate based on the TIGR Gene Index (Quackenbush *et al*., 2000). However, on considering only the 28 ITSs that show some similarity to public databases, a remarkable 20 or about two-third show homology to either predicted proteins or ESTs. With a completed rice genome sequence all ITSs can be positioned and will probably confirm the remarkable bias of insertions in genes. Due to the available DNA sequence information for rice, this genome is a good model to assess the phenomenon of transposon insertional specificity that has been often alluded to in complex genomes but never proven

conclusively. The implication of this transpositional bias means that about three times less number of insertions need to be generated to be able to saturate the genome with insertions.

Ac in rice genomics

Rice is a model plant for the discovery of gene functions in cereals. The rice genome sequence will uncover about 30,000 genes, half of which will have no known function. Transposon mutagenesis using knockout and gene-detection insertions will be a very important tool to discover these gene functions by reverse genetics strategies. Even with a smaller genome such as Arabidopsis, knocking out every gene would entail about 100,000 random inserts (Krysan *et al*., 1999) and for rice about four times that number. Multiple independent inserts per plant, averaging four in many of the *Ac* lines, will bring down the required number of plants. The insertional preference of *Ac* for genes as described here, will reduce the required number further by a factor of 3. A population of about 30,000 *Ac* plants would therefore be sufficient to recover knockout mutants in most genes. These populations can be produced with minimal effort and available to research programs worldwide involved in rice functional genomics.

Acknowledgements

We are grateful to Pankaj Dhonukshe for help in some DNA analysis. This research was funded in the participating labs by a European Union project BIO4 CT 972132 on “Transposon mutagenesis in rice”.

LITERATURE CITED

Aarts, M.G.M., Corzaan, P., Stiekema, W.J. and Pereira, A. (1995) A two-element *Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana*. Mol. Gen. Genet. 247: 555-564

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402

Baker, B., Coupland, G., Fedoroff, N., Starlinger, P. and Schell, J. (1987) Phenotypic assay for excision of the maize controlling element *Ac* in tobacco. EMBO J. 6: 1547-1554

Balcells, L. and Coupland, G. (1994) The presence of enhancers adjacent to the *Ac* promoter increases the abundance of transposase mRNA and alters the timing of *Ds* excision in *Arabidopsis*. Plant Mol. Biol. 24: 789-798

Balcells, L., Sundberg, E. and Coupland, G. (1994) A heat-shock promoter fusion to the *Ac* transposase gene drives inducible transposition of a *Ds* element during *Arabidopsis* embryo development. Plant J. 5: 755-764

Benfey, P.N. and Chua, N.-H. (1990) The Cauliflower Mosaic Virus 35S promoter: combinatorial regulation of transcription in plants. Science 250: 959-966

Bingham, P.M., Levis, R. and Rubin, G.M. (1981) Cloning of DNA sequences from the *white* locus of *D. melanogaster* by a novel and general method. Cell 25: 693-704

Charng, Y.-C., Pfitzner, U.M. and Pfitzner, A.J.P. (1995) Fusion of the inducible promoter of the PR-1a gene to the Activator transposase gene can transactivate excision of a non-autonomous transposable element by external and by internal stimuli. Plant Sci. 106: 141

Chin, H.G., Choe, M.S., Lee, S.H. *et al*. (1999) Molecular analysis of rice plants harboring an *Ac/Ds* transposable element-mediated gene trapping system. Plant J. 19: 615-623

Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. and Sheen J. (1996) Engineered GFP as a vital reporter in plants. Curr. Biol. 6: 325-30

Chuck, G., Robbins, T., Nijjar, C., Ralston, E., Courtney-Gutterson, N. and Dooner, H.K. (1993) Tagging and cloning of a petunia flower color gene with the maize transposable element *Activator*. Plant Cell 5: 371-378

Davis, S.J. and Vierstra, R.D. (1998) Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol. Biol. 36: 521-528

Earp, D.J., Lowe B. and Baker B. (1990) Amplification of genomic sequences flanking transposable elements in host and heterologous plants: a tool for transposon tagging and genome characterization. Nucleic Acids Res. 18: 3271-3278

Enoki, H., Izawa, T., Kawahara, M., Komatsu, M., Koh, S., Kyozuka, J. and Shimamoto K. (1999) *Ac* as a tool for the functional genomics of rice. Plant J. 19: 605-613

Finnegan, E.J., Taylor, B.H., Craig, S. and Dennis, E.S. (1989) Transposable elements can be used to study cell lineages in transgenic plants. Plant Cell 1: 757-764

Greenblatt, I.M. (1984) A chromosome replication pattern deduced from pericarp phenotypes resulting from movements of the transposable element *Modulator*, in maize. Genetics 108: 471-485

Haring, M.A., Rommens, C.M.T., Nijkamp, H.J.J. and Hille, J. (1991) The use of transgenic plants to understand transposition mechanisms and to develop transposon tagging strategies. Plant Mol. Biol. 16: 449-461

Hehl, R. and Baker, B. (1989) Induced transposition of *Ds* by a stable *Ac* in crosses of transgenic tobacco plants. Mol. Gen. Genet. 217: 53-59

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

Izawa, T., Miyazaki, C., Yamamoto, M., Terada, R., Iida, S. and Shimamoto, K. (1991) Introduction and transposition of the maize transposable element *Ac* in rice (*Oryza sativa* L.) Mol. Gen. Genet. 227: 391-396

Izawa, T. Ohnishi, T., Nakano. *et al*. (1997) Transposon tagging in rice. Plant Mol. Biol. 35: 219-229

James, D.W.Jr., Lim, E., Keller, J., Plooy, I., Ralston, E. and Dooner, H.K. (1995) Directed tagging of the Arabidopsis *FATTY ACID ELONGATION1* (*FAE1*) gene with the maize transposon *Activator*. Plant Cell 7: 309-319

Jeon, J.S., Lee, S., Jung, K.H. *et al.* (2000) T-DNA insertional mutagenesis for functional genomics in rice. Plant J. 22: 561-70

Jones, J.D.G., Carland, F.M., Maliga, P. and Dooner, H.K. (1989) Visual detection of transposition of the maize element *Activator* (*Ac*) in tobacco seedlings. Science 244: 204-207

Jones, J.D.G., Carland, F.M., Lin, E., Ralston, E. and Dooner, H.K. (1990) Preferential transposition of the maize element *Activator* to linked chromosomal locations in tobacco. Plant Cell 2: 701-707

Jones, D.A., Thomas, C.M., Hammond-Kosack, K.E., Balint-Kurti, P.J. and Jones, J.D.G. (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266: 789-793

Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) T-DNA as an Insertional Mutagen in Arabidopsis. Plant Cell 11: 2283-2290

Lawrence, G.J., Finnegan, E.J., Ayliff, M.A. and Ellis, J.G. (1995) The *L6* gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. Plant Cell 7: 1195-1206

Lin, X., Kaul, S., Rounsley, S. *et al*. (1999) Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. Nature 402: 761-8

Mayer, K., Schuller, C., Wambutt, R. *et al*. (1999) Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana. Nature 402: 769-77

McClintock, B. (1947) Cytogenetic studies of maize and Neurospora. Carnegie Inst. Wash. Yearbook 46: 146-152

Murai, N., Li, Z.J., Kawagoe, Y. and Hayashimoto, A. (1991) Transposition of the maize *activator* element in transgenic rice plants. Nucleic Acids Res. 19: 617-22

Nakagawa, Y., Machida, C., Machida, Y., Toriyama, K. (2000) Frequency and pattern of transposition of the maize transposable element *Ds* in transgenic rice plants. Plant Cell Physiol. 41: 733-42

Parinov, S., Sevugan, M., Ye, D., Yang, W.-C., Kumaran, M. and Sundaresan, V. (1999) Analysis of Flanking Sequences from *Dissociation* Insertion Lines: A Database for Reverse Genetics in Arabidopsis. Plant Cell 11: 2263-2270

Pereira, A. (1998) Heterologous transposon tagging systems. In: K. Lindsey (Ed.), Transgenic Plant Research, Harwood Academic Publishers, UK, pp. 91-108

Pereira, A. and Aarts, M.G.M. (1998) Transposon tagging with the *En-I* system. In: J. Martinez-Zapater and J. Salinas (Eds.), Arabidopsis Protocols, Humana Press Inc, Totowa, pp. 329-338

Peterson, P.W. and Yoder, J.I. (1995) Amplification of *Ac* in tomato is correlated with high *Ac* transposition activity. Genome 38: 265-276

Pietrzak, M., Shillito, R.D., Hohn, T. and Potrykus, I. (1986) Expression in plants of two bacterial antibiotic resistance genes after protoplast transformation with a new plant expression vector. Nucleic Acids Res. 14: 5857-5868

Quackenbush, J., Liang, F., Holt, I., Pertea, G. and Upton, J. (2000) The TIGR Gene Indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res. 28: 141-145

Rueb, S., Leneman, M., Schilperoort, R.A. and Hensgens, L.A.M. (1994) Efficient plant regeneration through somatic embryogenesis from callus induced on mature rice embryos (*Oryza sativa* L.) Plant Cell Tiss Org 36: 259-264

Scofield, S.R., English, J.J. and Jones, J.D.G. (1993) High level expression of the *Activator* transposase gene inhibits the excision of *Dissociation* in tobacco cotyledons. Cell 75: 507-517

Shimamoto, K., Miyazaki, C., Hashimoto, H., Izawa, T., Itoh, K., Terada, R., Inagaki, Y. and Iida, S. (1993) 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: 354-60

Sijmons, P.C., Dekker, B.M., Schrammeijer, B., Verwoerd, T.C., van den Elzen, P.J. and Hoekema, A. (1990) Production of correctly processed human serum albumin in transgenic plants. Biotechnology 8: 217-221

Sundaresan, V. (1996) Horizontal spread of transposon mutagenesis: new uses for old elements. Trends Plant Sci. 1: 184-190

Swinburne, J., Balcells, L., Scofield, S.R., Jones, J.D.G. and Coupland, G. (1992) Elevated levels of *Activator* transposase mRNA are associated with high frequencies of *Dissociation* excision in Arabidopsis. Plant Cell 4: 583-595

The Arabidopsis Genome Initiative, (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815

Whitham, S., Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C. and Baker, B. (1994) The product of the tobacco mosaic virus resistance gene *N*: similarity to Toll and the Interleukin-1 receptor. Cell 78: 1101-1115

Yoder, J.I. (1990) Rapid proliferation of the maize transposable element *Activator* in transgenic tomato. Plant Cell 2: 723-730

Chapter 4

Tagged Transcriptome Display in *indica* rice using *Ac* transposition.

Ajay Kohli[1], Jianhua Xiong[1], Raffaella Greco[2], Paul Christou[1] and Andy Pereira[2]

[1]*John Innes Centre, Norwich, UK*
[2]*Plant Research International, Wageningen, The Netherlands*

Molecular Genetics and Genomics 266: 1-11 (2001)

ABSTRACT

We have developed a transgenic population of *indica* rice lines containing the maize autonomous *Activator* transposon (*Ac*). A transposon excision assay using the green fluorescent protein (GFP) gene driven by the ubiquitin promoter was used to monitor *Ac* excision in various tissues. Our results, concerning *Ac* excision and re-insertion in 289 independent rice transformants, provide an insight into transposon biology in this heterologous model cereal system. Twenty percent of the transformed calli displayed uniform GFP activity, demonstrating very early *Ac* excision, while later excision in another 40% was revealed by mosaic GFP activity. Both phenotypes were confirmed by molecular analysis. Progeny analysis revealed active transposition with some lines displaying transposition to unlinked positions. Amplification of the *Ac* copy number was observed in approximately 30% of the lines, thus tagging multiple sites. We developed a 'transposon insertion display' procedure, involving CpG methylation-sensitive enzymes and a macro-array based approach using cDNA as a complex hybridization probe, to selectively detect transposon insertions in transcribed sequences. Sequencing of tagged sites identified by hybridization to leaf cDNA revealed that all the hybridizing tags were homologous to genes or ESTs in the database. Extrapolation of this data suggests that 12% of all amplified tags display homology to genes or ESTs, a remarkable five times higher than that expected on the basis of random transpositional insertion. This Tagged Transcriptome Display (TTD) technique, using a population of *Ac* insertion-tagged lines, is therefore a convenient approach to identify tagged genes involved in specific processes, as revealed by the expression pattern.

Key Words: *Rice, Functional genomics, Ac Transposon, Transcriptome, GFP, Excision marker, Ubiquitin.*

We dedicate this report to the fond memory of our dear friend and colleague Dr. J Harry C. Hoge of the Institute of Molecular Plant Sciences, Leiden University, The Netherlands.

INTRODUCTION

Transposons were once a classical tool for the identification and isolation of individual genes (reviewed by Balcells et al. 1991). They have been used in both forward genetics approaches to identify genes (Bingham et al. 1981) and in reverse genetics approaches to assign functions to known genes (Hamilton et al. 1991; Koes et al. 1995). However, transposon tagging has now become established as a genomics-based high-throughput technique for the simultaneous characterization of numerous genes (Martin 1998; Parinov and Sundaresan 2000).

The autonomous maize *Activator* (*Ac*) transposon can induce its own transposition and that of non-autonomous *Dissociation* (*Ds*) elements (McClintock 1947). The *Ac*/*Ds* system has been widely used as a molecular tool for gene tagging in a variety of heterologous dicot plants, e.g. tobacco, Arabidopsis, carrot, potato, tomato, soybean, lettuce and flax (Haring et al. 1991; Sundaresan 1996). Cereal crop plants also support *Ac*/*Ds* transposition. Recently Takumi et al. (1999) demonstrated *Ds* transposition in transgenic wheat plants expressing the *Ac* transposase gene. Koprek et al. (2000) described a method for dispersing *Ds* elements in the barley genome and Scholz et al. (2001) also reported transposition of *Ac* in barley. The use of *Ac*/*Ds* transposition in rice was reported ten years ago (Izawa et al. 1991; Shimamoto et al. 1993), and has recently been shown to be efficient for functional genomics in rice (Izawa et al. 1997; Enoki et al. 1999; Chin et al. 1999; Nakagawa et al. 2000; Greco et al. 2001). However, all of the transposon related work carried out in rice has involved model *japonica* rice varieties, which are relatively amenable to genetic, molecular and tissue culture techniques. *Agrobacterium*-mediated transformation of rice is also easier for the *japonica* group but this is not so for the *indica* group. However, the long-grain *indica* rice are commercially very important, since the economies of many South-East Asian countries (e.g. Vietnam, Thailand and Cambodia) depend heavily on its export. It is therefore important to generate an *indica* library of *Ac* tagged sites to complement existing *japonica* rice libraries.

A variety of phenotypic excision assays employing selectable or visual marker genes have been used to monitor transposon activity (reviewed in Pereira 1998). Vain et al. (1998) demonstrated the expression of GFP in rice, paving the way for the utilization of GFP as an excision marker. We have generated a library of tagged sites in *indica* rice variety using the autonomous *Ac* element and the green fluorescent protein (GFP) gene, driven by the ubiquitin promoter, as an excision marker. Transgenic plants were obtained by particle bombardment, and multiple plants regenerated from each transformed callus. Molecular analysis for the presence, excision and re-insertion of *Ac* confirmed its transpositional activity in *indica* rice and also revealed unpredictable transposition behavior. We observed very early *Ac* transposition activity,

transposition to unlinked sites, secondary transpositions, and amplification of *Ac* copy number. This population can be used to isolate and characterize novel rice genes.

Most genes are likely to be conserved between the *indica* and *japonica* groups, but the importance of finding differences cannot be overstated. Several reports suggest that *indica-japonica* polymorphism involves a cumulative process of gene change (Li et al. 2000; Cho et al. 1998). At the single gene level, differences between the *indica* and *japonica* groups have been characterized for important traits such as nitrate utilization (Barlaan and Ichii 1996). RFLP analysis of the glutelin gene subfamily also showed major differences between *indica* and *japonica* rice (Takaiwa et al. 1991). It would therefore be prudent to carry out comparative studies between the *indica* and *japonica* groups to investigate such differences. This would add another dimension to rice as a model genomics system. A wealth of information on evolutionary biology could be generated, defining the essential differences between two closely related groups. Rice is important both as a model cereal system and a staple for a large proportion of the world's population. This unique position can be used to understand the factors that lead to the differences in grain characteristics between the *indica* and *japonica* groups (applied/commercial science prospecting) and the differences in gene architecture and gene expression (basic science prospecting).

The rice genome-sequencing project is yielding new sequence information every day. However, information concerning the function of these sequences is lagging behind. It is therefore important to concentrate on elucidating the functions of these new sequences. We have standardized a method to identify the transcribed genomic sequences flanking transposon insertion sites. The method uses a combination of 'transposon insertion display' described by Yephremov and Saedler (2000), and a macro-array based functional genomics approach using total cDNA as a complex probe. This method efficiently identifies transposon insertion sites in transcribed regions and can therefore be used for generating a 'tagged transcriptome display' (TTD). The TTD method could be useful for the identification of *indica* group-specific gene expression in our rice population. Using this method on a T_1 transgenic population, we identified *Ac* insertions in a number of transcribed sequences when leaf cDNA was used as the probe. The probe was prepared from leaves growing under normal conditions, but various cDNA populations could be used serially to identify tagged genes functional at different developmental stages or under various biotic, abiotic conditions. This would result in the identification of tagged genes involved in specific biological processes and aid in their phenotypic characterization.

MATERIALS AND METHODS

Construction of Ubi:Ac:GFP

A 0.95-kb *Xba*I-*Bgl*II fragment containing the maize ubiquitin promoter, up to the start of the first untranslated exon (Ubi-1), was transferred from pAHC25 (Christensen and Quail 1996) into pBluescript-SK$^+$ and then isolated from this plasmid as an *Sst*I-*Sal*I fragment.

The *Ac* element was obtained from a derivative of pKU2 (Baker et al. 1987), in which a 4.6-kb *Bss*HII fragment containing the entire *Ac* element (originally isolated from the maize *wx-m7* allele) was treated with Klenow polymerase and then subcloned at the *Bam*HI site between the 35S promoter and the *nptII* gene in the inverse orientation. From this plasmid, a 3.5-kb *Pst*I-*Xba*I fragment containing the 5' portion of *Ac* (from position 1 to the internal *Xba*I site at position 3426) and a 1.1-kb *Xba*I-*Bam*HI fragment containing the 3' portion of *Ac* (3426-4565) were obtained and used in the final ligation. The gene encoding the soluble-modified green fluorescent protein (sm-GFP; Davis and Vierstra 1998) containing an ATG start codon and fused to the NOS-terminator, was derived as a 1-kb *Bam*HI-*Eco*RI fragment from the plasmid CDC-326. The final construct was assembled in a one step reaction by ligating the four fragments described above in the binary vector pMOG22 (ZENECA-MOGEN, NL) previously linearized with *Sst*I and *Eco*RI.

Transformation of rice

Particle bombardment-mediated transformation of *indica* rice (*Oryza sativa* L. *indica* cv. Bengal and Pusa Basmati) was carried out as described by Sudhakar et al. (1998). The primary and secondary selection of the putative transgenic calli was made on media containing hygromycin as a selectable agent. Calli were transferred to fresh media after 15 to 18 days and monitored for GFP activity by looking for bright green sectors while they were still enclosed in the petri-plates, so as to avoid contamination.

GFP expression analysis

The expression of GFP in calli and transgenic plants was monitored with a Leica MZ FluoIII with a fluorescent source as described earlier (Vain et al. 1998).

DNA isolation

Genomic DNA was isolated from transgenic plants using a modified version of the method of Edwards et al. (1991). Up to 200 mg of leaf tissue was frozen crisp in liquid nitrogen in a microfuge tube. The tissue was macerated with a T-shaped stainless steel grinder (custom built at the JIC workshop for use in microfuge tubes). Four hundred microliters of extraction buffer (200

mM Tris HCl, pH 7.5; 250 mM NaCl; 25 mM EDTA; 0.5% SDS) was added to the microfuge tube followed by further maceration. Then, 400 µl of phenol:chloroform:isoamyl alcohol (25:24:1) was added and the tube vortexed briefly. The mixture was left on ice for 5 minutes and then centrifuged at 13,000 rpm for 10 minutes in a microfuge. The supernatant was transferred to a fresh microfuge tube and 0.7 volumes of isopropanol was slowly added and mixed by inverting the tube gently. DNA was precipitated by centrifuging this mixture at 13,000 rpm for 5 minutes. The pellet was air dried and dissolved in 50-100 µl TE containing RNase to a final concentration of 10 ng/µl. The DNA solution was incubated at 65°C for 15 minutes and then stored at -20°C. Using this modified protocol and a moderate scale up of the procedure, we obtained a quality and quantity of DNA that was adequate for Southern hybridization and PCR.

PCR analysis for transformation and transposition.

PCR analysis to confirm transformation was carried out by amplification of the selectable marker gene hygromycin phosphotransferase (*hpt*), using *hpt* specific primers in the coding region: HPTF1, 5'-ACTCACCGCGACGTCTGTCG-3' and HPTR1, 5'-GATCTCCAATCTGCGGGATC-3'. All plants were also analyzed for the presence of the transposon using *Ac* primers: Ac971-F, 5'-ACGACTCCATTCCTCAGATGACG-3' and Ac1395-R, 5'-CTTGACTCGGATCTGTAGCTGTACC-3', which produced a 424-bp fragment. *Ac* excision events were detected by amplification of the empty donor site (EDS) using a primer annealing to the ubiquitin promoter (UBI1: 5'-TGGACCCCTCTCGAGAGTTC-3') in combination with a primer annealing to the *gfp* gene (smGFPR1: 5'-GAAAGGGCAGATTGTGTGGACAGG-3'). Following *Ac* transposition, a 1.45-kb fragment was amplified. The presence of a full donor site (FDS), indicating the presence of *Ac* (i.e. no transposition), was also examined. The FDS-left junction was amplified using the UBI1 primer in conjunction with a reverse primer in *Ac* (Ac282R: 5'-CTCAGTGGTTATGGATGGGAGTTG-3'). For the FDS-right junction, *Ac* forward primer (Ac3877F: 5'-CCAGCCGTGTTAGTTCACAGATTC-3') was used in conjunction with the smGFPR1 primer. Amplification of an EDS fragment and the 424-bp *Ac* fragment in the same plant indicated *Ac* excision and reinsertion. Standard PCR profiles were used for 35 cycles, the precise conditions depending upon the T_m of the primers.

Southern blot hybridization

Southern blot hybridization was performed either on genomic DNA or on PCR products. In the case of genomic DNA, 5 µg of DNA was digested with appropriate restriction enzymes. DNA (genomic or PCR products) was fractionated by 0.8% agarose gel electrophoresis in 1X TAE and blotted onto Amersham Hybond N^+ membranes as per the manufacturer's instructions. Probe DNA (25 ng of gene specific PCR fragment) was radiolabeled with [^{32}P]dCTP by the random

primer method (Feinberg and Vogelstein 1983). Hybridization was performed as described earlier (Kohli et al. 1999). The *Ac* and *gfp* probes used on genomic DNA digested with *Hin*fI were generated after double digestion of the right junction FDS PCR fragment with *Hin*fI and *Bam*HI. Results were obtained either as autoradiographic films or as printouts from phophorimager analysis, and the collage was assembled using Adobe Photoshop 5.0 and Microsoft Powerpoint 97.

Identification of transcribed Ac tagged sites

Two complementary 50-mers (ADA1: 5'-CCAGATGAGATTCAGTTGGCACAGGTA CCTGTCGACGTCTCAGCTTCTAG-3' and ADA2: 5'-CTAGAAGCTGAGACGTCG ACAGGTACCTGTGCCAACTGAATCTCATCTGG-3'), engineered to contain *Sal*I (GTCGAC) and *Kpn*I (GGTACC) restriction sites, were mixed at an equimolar concentration. The mixture was heated to 65°C for 5 minutes, left for 20 minutes to anneal at room temperature and finally 200 pMoles of the dsDNA was digested with *Sal*I (Pharmacia) as per manufacturer's instructions. Transgenic rice genomic DNA (500 ng) was also digested with the restriction enzyme *Sal*I, which does not have a site in the *Ac* transposon. The reaction products would therefore comprise genomic sequences flanking either end of the *Ac* element, mixed with numerous other *Sal*I fragments. The two digestion products were mixed (50 pMoles adapter/500 ng genomic DNA) and ligated in the presence of 1 unit T4 DNA ligase, 1X ligation buffer, 1.2 mM ATP and 2.5 mM DTT. The reaction was carried out at 37°C for 3 hours in a 20-µl final volume. 2-µl ligated product was used as the template in PCR amplifications with one primer annealing to the adapter (AP1: 5'-CCAGATGAGATTCAGTTGGCAC-3') and the other forward or reverse primer annealing to the *Ac* element (ACGF1: 5'-CAGCCGTGTTAGTTCACAGATTC-3' or ACGR1: 5'-CAGGTG TTCAGCAATCAGCTTGGTG-3'). All PCR reactions were in 50-µl volumes using 20 pM of individual primers. The PCR conditions used were as follows: denaturation at 95°C for 1 minute, annealing at 58°C for 1 minute and extension at 72°C for 2 minutes. The amplifications were carried out for 32 cycles and a 1/500 dilution of PCR products were used for succesive nested PCR reactions. Nested PCR was carried out using different primers in the adapter (AP2: 5'-GATTCAGTTGGCACA GGTACCTG-3') and forward or reverse primers in the *Ac* element (ACGF2: 5'-GAAGA TGTAGCAAGTGGA TCCTCTCC-3'; ACGF3: 5'-CCGAACAAAAATACCG GATC CCGTCCGATTTCG-3'; ACGR2: 5'-CAATCAGCAGGTGTTGCGGA TCCCC T-3'; ACGR3: 5'-CGGTAACGAAACGGGATCATC-3'). Half the volume of the final PCR products was separated on a 0.8% agarose gel, blotted onto an Amersham Hybond N^+ membrane and analyzed by Southern blot hybridization as described above. Total mRNA (200 ng) isolated from young leaves growing under normal conditions was reverse transcribed with AMV reverse transcriptase (Gibco) as per the manufacturer's instructions and used as template to prepare the complex probe by the

random primer method (Feinberg and Vogelstein 1994). The specific activity of the radiolabeled probe was increased by double-labeling using [^{32}P]dCTP and [^{32}P]dATP in the reaction mixture. Autoradiography was carried out either in a phosphorimager for one day or on X-OMAT Kodak films for three to five days. Positive bands were further characterized using the remaining half volume of the final PCR, as described below.

Cloning and characterization of transcribed Ac tagged sites

PCR products hybridizing to the cDNA probe were isolated from the gel and cloned using the TOPO TA cloning kit (Invitrogen) as per the manufacturer's instructions. The cloned fragments were then sequenced using the ABI automated sequencer and the sequence information obtained was compared to known sequences in public databases using the BLAST algorithm at the National Center for Biotechnology Information (NCBI).

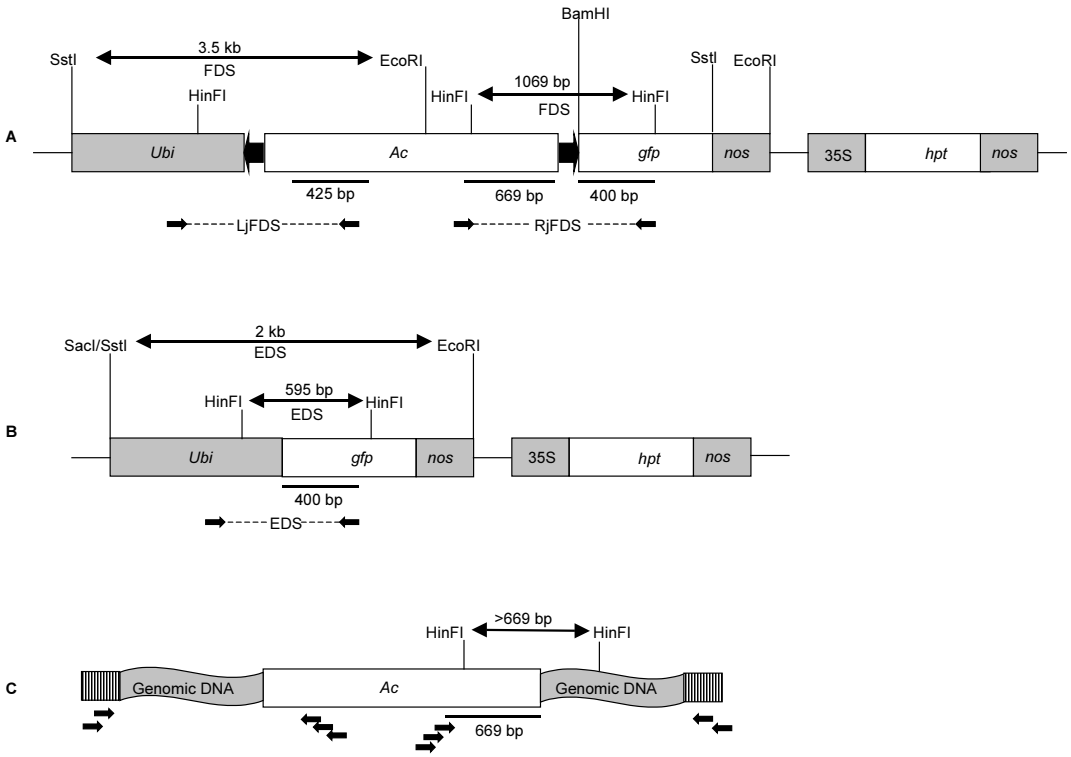


Figure 1. Schematic representation of the *Ac* transposon construct used for rice transformation. The *gfp* gene was used as an excision marker and the *hpt* gene as a selectable marker. The construct is shown in A) its non-excised and B) its excised state, giving rise to the full donor site (FDS) or empty donor site (EDS) products, respectively, as detected by Southern blot hybridization or PCR. Either a combination of *Sst*I and *Eco*RI or simply *Hin*fI was used to detect FDS/EDS products by Southern blot hybridization. In the former case a 425-bp PCR product for *Ac* and a 550-bp PCR product for *gfp* were used as probes. In the latter case a 669-bp *Ac* fragment and a 400-bp *gfp* fragment were used as probes. These were generated by simultaneous digestion of the right junction-FDS PCR product from the original construct in the binary vector with *Hin*fI and *Bam*HI, followed by gel extraction of the two fragments. Detection of left junction (LJ) or

right junction (RJ) FDS and EDS products by PCR was carried out using the primers shown as thick arrows under each construct. Re-insertion events schematically represented in C) were detected by Southern blot hybridization as *Hin*fI bands larger than 669 bp when the 669-bp *Ac* fragment was used as probe. Flanking genomic sequences were amplified by PCR using nested primers (thick arrows) in *Ac* and the *Sal*I adapter (shaded rectangle) at either end of the reinserted *Ac* element.

RESULTS

Transformation of indica rice with Ac and the use of GFP as an excision marker

The *indica* rice varieties Bengal and Pusa Basmati were transformed with the Ubi:*Ac*:smGFP construct (Fig. 1) by particle bombardment. The construct contains the *Ac* autonomous element and the smGFP gene, whose expression is restored upon *Ac* excision. GFP activity was thus used to monitor *Ac* excision. We report results obtained with the variety Bengal. A total of about 1500 plants representing 289 independent transgenic lines were regenerated. Transgenic calli surviving secondary selection (about 3 weeks after bombardment) on media containing hygromycin revealed three categories of GFP activity (Fig. 2A). Approximately 40% of the calli showed no GFP activity (type A callus), while another 40% showed mosaic GFP patterns (type B callus). The remaining 20% expressed GFP uniformly in all callus cells (type C callus). In this last category, all the sub-cultured pieces of the mother callus expressed GFP strongly, indicating very early excision of *Ac*. We allowed all three categories of calli to mature into plants, ensuring the recovery of both early and late transposition events.

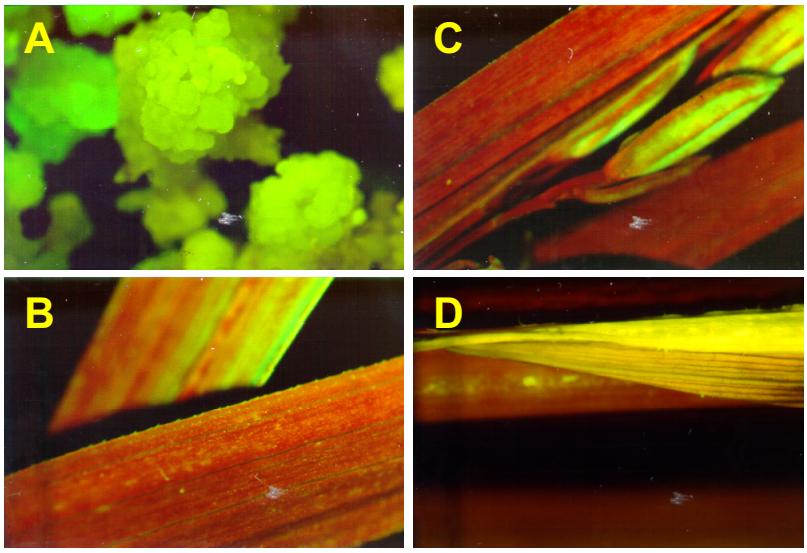


Figure 2. GFP activity in transgenic calli and plant lines. A) Uniform GFP activity in type 3 calli, indicating early excision of *Ac*. B) Mosaic GFP activity in a plant, which was shown to be chimeric for *Ac* excision. C, D) Mosaic GFP activity in two different plant lines, which were shown to be non-chimeric for *Ac* excision.

Molecular analysis for *Ac* excision involved PCR amplification of *Ac*, full donor sites (FDS) and empty donor sites (EDS; Fig. 1) (Table 1). Nearly 90% of the plants regenerated from type A calli produced the expected bands for *Ac* and FDS but no EDS product, indicating that *Ac* did not excise from any copy of the integrated construct. The remaining 10% of these plants produced no products of the expected size, indicating disruptive rearrangements of the construct. Most of the plants regenerated from type B (mosaic) calli (Fig. 2B) produced all three bands as expected: *Ac*, FDS and EDS (Table 1). This suggested that each callus comprised a mosaic somatic cell population reflecting *Ac* excisions in some but not all cells or integrated copies. Other plants exhibiting mosaic GFP activity (Fig. 2C and D) were regenerated from type C (uniform GFP) calli. These plants produced bands for *Ac* and EDS but not for FDS, indicating complete early excision of *Ac* from all the integrated copies of the construct. Nevertheless, none of these plants expressed GFP uniformly over the whole surface of different plant parts.

Table 1. Independent transposition frequency (ITF) in progeny (T_1) plants of different lines.

Features of transgenic line	Excision status[a]					
	EDS		EDS/FDS		FDS	
	Line 4	Line 6	Line 7	Line 8	Line 11	Line 17
Ac copies[b]	1 - 3	1 - 4	1 - 2	1	4 - 5	2
Total No. of T_1 plants	10	12	7	4	9	6
Plants with unique inserts	4	5	7	2	4	0
Total No. of inserts	21	22	10	4	10	12
Unique inserts	6	4	5	2	4	2
ITF[c]	28%	18%	50%	50%	40%	16%

[a] The three categories of excision status are represented as described in the text.
[b] The minimal and maximal *Ac* copy number is indicated.
[c] The ITF is aclculated as the percentage of unique inserts among the total inserts.

Excision and re-insertion of *Ac* in T_0 plants was also analyzed by Southern blot hybridization, using genomic DNA restricted with enzymes *Sst*I and *Eco*RI and a 425-bp fragment corresponding to the 5' end of *Ac* (Fig. 1) a probe. A representative sample showing results obtained for two clones each from eight different lines is shown in Fig. 3, displaying mosaic GFP activity in adult plants. In agreement with the PCR results, lines A19-6 and A19-16 regenerated from type B (mosaic) calli show the expected 3.5-kb FDS fragment whereas the other six plants regenerated from type C (uniform GFP) calli do not display the FDS fragment. All major intense bands and most of the minor bands in these six lines are larger than 2.5-kb, indicating *Ac* re-

insertion after excision. Common intense *Ac* fragments in two clones of the same line are indicative of early transposition events, while unique minor bands represent unique transposition events occurring later in the development of the individual plants.

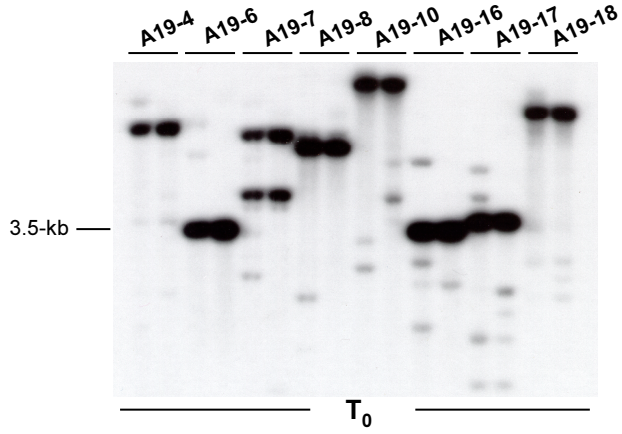


Figure 3. GFP activity in transgenic calli and plant lines. A) Uniform GFP activity in type 3 calli, indicating early excision of *Ac*. B) Mosaic GFP activity in a plant, which was shown to be chimeric for *Ac* excision. C, D) Mosaic GFP activity in two different plant lines, which were shown to be non-chimeric for *Ac* excision.

Ac transposition behavior in progeny suitable for insertional mutagenesis

To monitor the status of transposition events in the next generation, we analyzed lines derived from the three different types of GFP-expressing calli reflecting different status of *Ac* excision by Southern blot hybridization. We selected three to six progeny plants from two to three clones originating from 20 independent transgenic lines. Figure 4 shows the results of such analysis on one line for each of the three categories. The same blots were probed with *Ac* and *gfp* to determine the presence of EDS, FDS and new re-insertion events. The restriction enzyme *Hin*fI was used to digest genomic DNA because it is methylation insensitive, so the expected EDS and FDS fragments are therefore shorter (Fig. 1) and better resolved. Also, since *Hin*fI is a 5-base cutter, it cleaves more frequently in the genome than standard 6-base cutters resulting in better resolution of the DNA fragments and an increased likelihood of identifying all *Ac* re-insertion events.

Figure 4A shows the results of Southern blot analysis of line 6, representing plants with early *Ac* excision (type C calli). Six progeny plants from each of the two regenerated clones (6-1 and 6-2) were analyzed. As expected, both parents showed only the presence of the 0.6-kb EDS fragment after hybridization with *gfp*, while two bands representing new *Ac* insertions were detected after hybridization with the *Ac* probe (data shown for 6-1 only). The two *Ac* specific bands were transmitted to the progeny and segregated as independent unlinked insertions, as revealed by the presence of plants with *Ac* fragments but no EDS fragments and vice versa. New *Ac* bands were also present in the progeny indicating the occurrence of late transposition events. To estimate the copy number in T_0 plants, Southern blot hybridization was carried out after genomic DNA was digested with restriction enzyme *Aat*II, which cuts once in the construct,

followed by probing with the *hpt* coding sequence. Two hybridizing bands were observed (data not shown) indicating integration of at least two copies of the construct.

Figure 4B shows the results of Southern blot analysis of line 7 with two copies, representing plants with mosaic excision pattern (type B calli). This was confirmed by the presence of FDS in both the T_0 parents, while only clone 7.4 showed the presence of EDS. The absence of FDS in the progeny of 7.4 indicated complete excision and the presence of EDS in the progeny from clone 7.5 indicated late excision. Unique *Ac* hybridizing bands in progeny plants from both clones revealed the occurrence of new independent transposition events (data not shown).

Figure 4C shows the results of Southern blot analysis of line 11 representing inactive plants (type A calli). While both the parents showed no transposition activity (presence of FDS only), late germinal excision occurred in some T_1 progeny plants from clone 11-4, as revealed by the presence of EDS and of new *Ac* hybridizing bands (data not shown). On the other hand, all five T_1 plants analyzed from clone 11-3 remained inactive. The occurrence of aberrant gfp-hybridizing bands in some of the progeny plants from line 7 and 11 is discussed later.

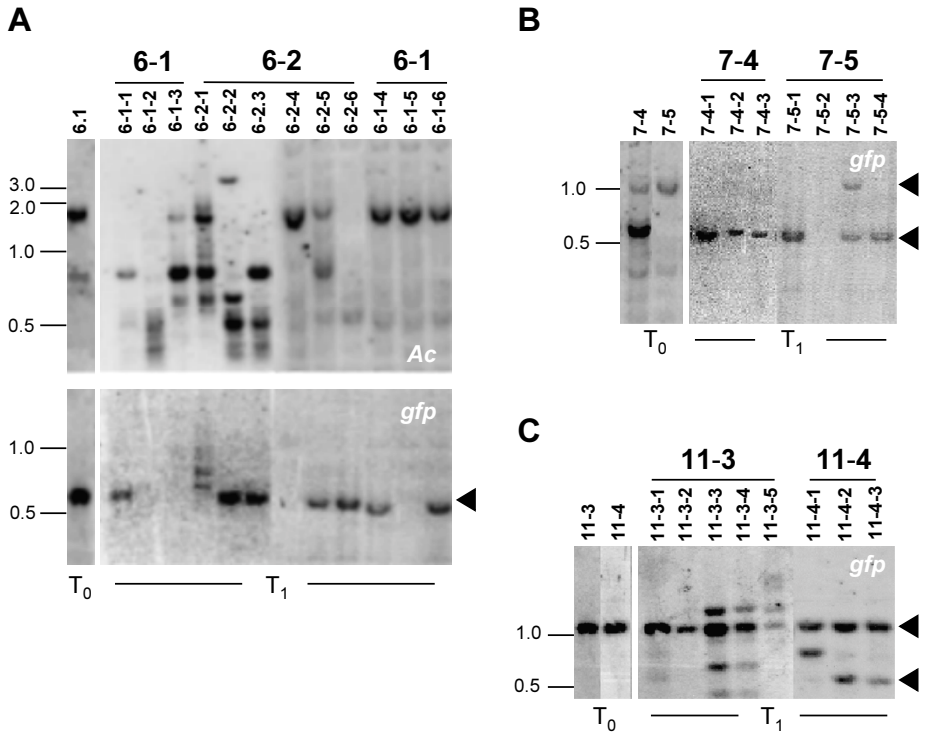


Figure 4. Representative Southern blot hybridization results for T_0 and T_1 plants derived from three different categories of calli. A) Plants arising from type C (uniform GFP) calli. B) Plants arising from type B (mosaic GFP) calli; and C) Plants arising from type A (no GFP) calli. Numbers above each double panel indicate T_0 and T_1 plant numbers. Numbers on the side of each panel indicate the size marker. Results from hybridization experiments with *Ac* (upper) and *gfp* (lower) are shown in A), while only hybridization with *gfp* is shown in B) and C). The 0.6-kb EDS and the 1.1-kb FDS bands are indicated by arrowheads.

Ac transposition events in T_1 progeny plants were used to calculate the independent transposition frequency (ITF) that indicates the frequency of new insertions. The results are summarized in Table 1. Four to fourteen T_1 plants from two lines each derived from the three callus types described above were used to assess unique transposition events, revealing ITF values from 11% to 50%.

Identification and characterization of Ac insertion tagged transcribed genomic sequences

Isolated genomic DNA was used as the template for PCR after *Sal*I digestion and ligation of the *Sal*I adapter as described in Materials and Methods. The primary PCR and two nested PCR amplifications were carried out, sequentially, using nested primers annealing to the adapter and the *Ac* sequence (Fig. 1C). A set of primers annealing at both ends of the *Ac* transposon (forward and reverse *Ac* primers) was used in combination with the two adapter primers on genomic DNA from each plant. T_1 plants were used for this analysis. After the tertiary PCR performed with primers AP2 and ACGF3/ACGR3, discrete bands were obtained from the transgenic lines, whereas no amplification product was obtained from the non-transformed control.

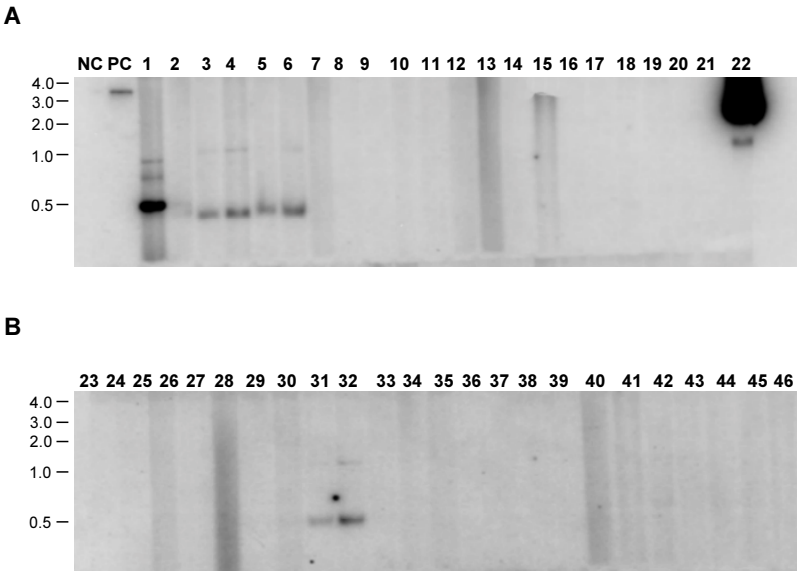


Figure 5. Representative results (T_1 plants) of Southern blot hybridization of PCR amplified genomic sequences flanking *Ac* insertion sites, using wild type leaf cDNA as the probe (reverse transcribed mRNA) of non-transgenic plant. Primers for internal transcribed sequences (ITS) were used to amplify a PCR product which was used as a positive control (PC). The PCR product of the *gfp* gene was used as a negative control (NC). Panel A) shows results obtained when nested PCRs were carried out with *Ac* forward and the adapter primers. Panel B) shows results obtained when the *Ac* reverse primers were used along with the adapter primers. The two panels show tagged transcribed sequences identified as hybridization bands. Numbers on the side of the panels indicate the size marker. Apart from the positive control, clear single or multiple hybridization signals were observed in eight of the 46 lanes.

One or more fragments were amplified when nested PCR was performed with either the *Ac* forward or reverse primers. Tertiary PCR products were blotted and probed with cDNA prepared from young leaves, as described in Materials and Methods. A representative result of such hybridization experiments is shown in Figure 5. One or multiple hybridization signals were observed in eight of the 46 gel lanes. Differences in the intensity of the signals are due to differences in the amount of cDNA probe which hybridize to the DNA fragment. The upper panel of Figure 5 shows hybridization results obtained with amplification products of the tertiary PCR carried out with nested forward *Ac* primers in conjunction with the primers for the *Sal*I adapter. The lower panel shows similar results when nested reverse *Ac* primers were used. Note that lanes in the upper and lower panels do not correspond to the same plant. Lanes 1-6 show results for T_1 plants from line 6. The two minor bands in lane 1 (plant number 6-2-2) were not visible on the gel. Lanes 3 and 4, corresponding to plant numbers 6-2-3 and 6-2-5 that originate from the same T_0 parent, show a band of similar size . Sequencing results confirmed that these two bands were identical. Similarly in lanes 5 and 6 (plant numbers 7-4-2 and 7-4-3) and lanes 31 and 32 (plant numbers 9-2-1 and 9-2-2) the two bands of similar sizes turned out to be identical in the sibling plants. Including the two bands in lane 22 (plant number 12-1-1), eight independent tagged insertion sites were thus identified. This result immediately distinguished those plants in which potential genes had been tagged. Less than half the volume of the tertiary PCR was used for the agarose gel. For plants in which the insertion was located within transcribed sequences, the remaining half volume of the tertiary PCR was separated on a second agarose gel and the relevant fragments isolated and cloned. Uniquely in the case of the results obtained with forward *Ac* primers in plant number 6-2-2, an additional quaternary PCR was carried out to amplify and visualize the minor bands detectable only after Southern blot hybridization. Of a total of 232 bands, 41 hybridized to the leaf cDNA probe. All of these bands were cloned and sequenced, and 33 proved to be unique.

The cloned DNA products were sequenced and the data used to search for similarity to known sequences in the NCBI database. Table 2 shows the homology search results obtained with the eight bands described in Figure 5. We found significant homology between our tagged sequences and rice or *Arabidopsis* genes and ESTs in the database. In particular, we found specific insertions in the rice catalase gene and a homologue of the *Arabidopsis* poly(ADP-ribose) polymerase gene in plants 12-1-1 and 6-2-2 respectively. The former insertion can be visualized as a very intense hybridization signal in Figure 5. These sequencing results exemplify the suitability of identifying tagged genes by sequencing transposon-flanking DNA.

Table 2. Similarity of transcribed genomic sequences (TGS) tagged by Ac to sequences in database.

Plant	No. of TGS	Blast Homology	Organism	Accession	Identity
6.2.2	3	poly(ADP-ribose) polymerase	*Arabidopsis*	AJ131705	95%
		EST	*Oryza sativa*	AU032953	95%
		Genomic BAC clone	*Oryza sativa*	AQ860441	92%
6.2.3	1	EST	*Oryza sativa*	AU090661	100%
6.2.5	1	EST	*Oryza sativa*	AU090661	100%
7.4.2	1	EST	*Oryza sativa*	AU031421	82%
7.4.3	1	EST	*Oryza sativa*	AU031421	82%
9.2.1	1	EST	*Oryza sativa*	AU089823	95%
9.2.2	1	EST	*Oryza sativa*	AU089823	95%
12.1.1	2	CatB gene for catalase	*Oryza sativa*	D64013	95%
		EST	*Oryza sativa*	AU031421	90%

DISCUSSION

An indica perspective

The use of *Ac* for gene tagging in rice is now well established (Izawa et al. 1991; Shimamoto et al. 1993; Enoki et al. 1999; Chin et al. 1999; Nakagawa et al. 2000; Greco et al. 2001). However, all transposon-related studies in rice have been carried out using *japonica* rice varieties. Synteny between cereal genomes (Devos et al. 1995; 1997) has validated the use of rice as a model monocot system (Tyagi et al. 1999; 2000), further justifying the use of model *japonica* varieties. However, the presence of *indica* group-specific genes and group-specific differential expression of numerous genes, suggests that such differences should be identified and characterized. The rate of success in cross-fertilizing *indica* and *japonica* varieties is very low and it is a cumbersome time-consuming operation, thus warranting a separate study in *indica* rice integrated with parallel studies in *japonica*. The efficiency of *Agrobacterium*-mediated transformation of *indica* rice varieties is still poor so particle bombardment remains the method of choice for transformation. We have used an available efficient genotype independent particle bombardment based transformation system to generate a library of *Ac* tagged insertion sites in *indica* rice to complement the ongoing gene tagging efforts for *japonica* rice varieties.

Development of a knockout mutation machine

We obtained 289 independent transformed lines with the *Ac* construct using GFP as an excision marker. The pattern of GFP activity correlated well with the molecular status of *Ac* activity. We observed GFP activity in 60% of the independent lines very early during the primary selection of the transgenic calli. GFP was active in sectors in 40% of the calli (type B calli), and uniformly in the remaining 20% (type C calli), indicating the occurrence of very early excision events. Surprisingly, all the mature plants regenerated from type C calli showed mosaic GFP activity, although molecular analysis revealed complete excision of *Ac*. In such plants, the mosaic activity is likely to reflect sectorial gene silencing. Indeed, variegated GFP activity was also observed when control rice plants were transformed with a Ubi:GFP construct without the *Ac* element (data not shown). Taken in combination with molecular analysis data, GFP expression was always a very good indicator of *Ac* excision as we obtained at least mosaic GFP expression whenever Ac excision occurred.

To increase the chance of recovering late excisions that could generate independent transposition events, we did not select for early excision events at the callus stage. In general, lines with early excision also displayed late secondary transpositions, which resulted in the generation of independent insertions in the next generation, e.g. line 6 with an ITF in progeny of 11-50%. In the case of late excision, the event could involve one or all the integrated copies independently at different times, as observed in progeny from clones 7-4 and 7-5 (Figure 4, Ac hybridization data not shown). Lines that were inactive in the T_0 generation either displayed transpositional activity in T_1 (e.g. 11-4) or remained inactive (e.g. 11-3).

Although particle bombardment can often cause complex patterns of integration, detailed analysis in this study often revealed a single transgenic locus. In line 6, the presence of two Ac and two hpt bands (data not shown) revealed after Southern blot hybridization indicated integration of at least two copies of the construct. PCR analysis of 20 T_1 progeny from each of the clones 6-1 and 6-2 (data not shown) revealed a clear 3:1 segregation ratio for gfp, indicating a single transgenic locus. Particle bombardment-mediated integration of multiple transgene copies as a single genetic locus in rice has been previously described (Kohli et al. 1998).

Excision of Ac from both copies in the two T_0 clones (6-1 and 6-2) took place very early, as indicated by Southern blot hybridization and by uniform GFP activity in the calli. The various segregation combinations involving the two *Ac* bands and the EDS band in the T_1 plants (Fig. 4A) indicated that the re-inserted *Ac* elements were located far enough apart to segregate easily from each other (shown by the presence of only one of the two parental *Ac* bands) and from their original sites of integration (shown by one or both parental *Ac* bands being present but *gfp* absent and *vice versa*). This demonstrates transposition of *Ac* to unlinked sites, as recently described for

rice (Nakagawa et al. 2000), and can be applied to disperse the transposon insertions all over the genome.

The presence of novel *Ac* bands in addition to the two inherited parental insertions in some of the two-copy T_1 plants from line 6 is most probably caused by the transposition of one of the *Ac* elements from unreplicated to replicated DNA, with subsequent increase in copy number. The process of *Ac* transposition and amplification during replication is well known in maize (Greenblatt 1984). Since more than one progeny plant is involved but not all plants, this event may have occurred late during the development of the T_0 plant in one of the tillers.

The absence of EDS bands and the simultaneous occurrence of new aberrant *gfp*-hybridizing bands in some of the T_1 plants (lines 6 and 11) might reflect either intragenic recombination between the copies or deletions caused by *Ac* excision or insertion. Common aberrant bands must have originated from a rearrangement that occurred in the early stages of gametocyte formation. In general, the complex *Ac* insertion patterns observed in progeny arise due to the unique characteristics of rice development. Every tiller can give rise to an independent set of seeds with different transposition patterns. Harvesting seeds from a single plant comprising five or six tillers can therefore generate a mixed stock, with different transposition events. This feature is very useful for the recovery of independent late transposition events from the same plant.

To develop a mutation machine in *indica* rice, we have produced a large number of transformants that should provide an active *Ac* starter position every 10 cM of the genome, thus providing linked transposition to adjacent 10 cM intervals that can saturate the genome. Late and active transpositions from multiple *Ac* transposon-containing lines will generate the number of events required for genome saturation in a fewer number of plants. *Ac* transposons have been reported to remain active through successive generation in *japonica* rice (Enoki et al., 1999). On the basis of the findings reported in this study, we may reasonably believe that a similar behavior should be expected in *indica* rice as well. Through the propagation of about 25,000 individual lines originating from 100 transpositionally active multiple *Ac* lines for 3-4 generations, a population of 100,000 *Ac* inserts for knockout mutagenesis can be generated.

Preferential tagging of transcribed sequences

Cereal genomes contain large stretches of repetitive DNA (including transposons) in addition to gene-rich regions. To be able to select only for heterologous transposon inserts in genes, thus increasing the chances of identifying genes by reverse genetics procedures, we developed the Tagged Transcriptome Display (TTD) technique. The first principle employed in developing TTD was the use of a CpG methylation sensitive enzyme, such as *Sal*I, which is likely to cut only in active chromosomal regions that tend to contain actively transcribed genes. There

are a variety of other methylation sensitive enzymes that can be used to obtain further insertion sequences. The choice of such an enzyme helps to enrich the population of amplified flanking sequences for insertions in transcribed regions. On average, restriction recognition sites for an enzyme like *Sal*I, which has a six-base recognition sequence, occur every 4-kb in the genome. If a 4.5-kb element like *Ac* inserts between two *Sal*I sites in the genome, such that it is closer to one of the sites, it is conceivable that the other end of *Ac* would be too far from the second *Sal*I site to be amplified by the normal PCR conditions we used. As expected, in most of the cases we obtained PCR amplification products using *either* forward or reverse *Ac* primer sets. Only in two plants did we obtain PCR products with *both* the sets of *Ac* primers (data not shown). This shows that the recovery of unmethylated tagged sites was effective in most plants. While probably just one enzyme might not recover all potential gene tagged sites, the presence of multiple inserts in a plant or the use of different enzymes would help to select a higher number of tagged genes.

The second principle tested was the screening of amplified transposon flanking DNA for sequences that are transcribed. We used a labeled cDNA prepared from leaf mRNA to hybridize a Southern blot containing *Ac* tagged sites. In this way we were able to identify 41 (18%) inserts potentially expressed in leaf tissue, out of a total of 232 amplified *Ac* tagged sites. Complementary DNA probes can be also made from mRNA isolated after the exposure of plants to various conditions, thus allowing the recovery of insertions in genes that are transcribed under, for example, particular biotic and abiotic stresses. Apart from revealing tags in transcribed regions, this technique also allows the identification of tissues in which the tagged gene is expressed, which in some cases may hint at the process in which the gene is involved. This is valuable information to help in the search for mutant phenotypes in a systematic manner.

The *Ac* flanking DNA fragments that hybridized to the cDNA probe were isolated and sequenced. The analysis of 30 plants for *Ac* insertion tagged transcribed sites is shown in Figure 5, revealed 11 hybridizing fragments out of 67 amplified flanking sequences. These cloned fragments were sequenced and revealed eight different tagged sites all showing homology to rice ESTs or genes in the public database. This gives a frequency of 12% (8/64) *Ac* tagged sites (non-redundant estimate) with homology to leaf transcripts.

The number of unique rice ESTs in databases is approximately 26,000 with about 400 bp average sequence length, extrapolating to 10.4 Mb total sequence or 2.5% (10.4/430) of the genome. Therefore, a random fragment of rice DNA would have on average a 2.5% chance of being homologous to ESTs in the database. However, the frequency of *Ac* tagged sites homologous to ESTs that we recovered using the TTD technique (12%) appears to be five times higher, suggesting preferential recovery of *Ac* insertions in transcribed sequences. Enoki et al. (1999) have previously demonstrated that *Ac* preferentially transposes into genes in rice by recovering 4% of inserts with flanking DNA showing homology to ESTs. Our results show that the

preferential amplification of unmethylated-tagged sites helps the recovery of at least three times (12/4) more insertions in transcribed sequences than random *Ac* amplification. This shows that the TTD procedure provides an efficient system for the amplification and selection of tagged genes.

This strategy can be elaborated to construct macro-arrays with pools of amplified DNA from different plants. The macro-arrays can then be hybridized to different cDNA probes, providing a transcriptome map of tagged rice genes. This transcriptome based tagging of rice genes is thus an efficient and selective technique of reverse genetics which can be applied to non-model plant genotypes that cannot be analyzed by saturated functional genomics procedures, and to larger cereal genomes to make reverse genetics based insertional mutagenesis more selective towards genes

Ac in rice functional genomics

Rice is a model plant for gene discovery in cereals, where much genome and EST sequence information is becoming available. The rice genome sequence which is rapidly becoming available will reveal about 30,000 genes, half of which have no known function. Reverse genetics using transposon insertional mutagenesis can then be an essential tool to discover gene functions. The active transposon populations in *indica* rice developed here will be available worldwide and can be used to generate lines with multiple independent transposon inserts. These will be directly useful also for forward mutant screenings, with potentially much less background mutation rates than T-DNA insertion populations. Screening of these lines using the TTD technique will help in the rapid identification of genotypes carrying mutants of specific genes in the valuable *indica* rice genotypes.

Acknowledgements

The research was funded in the participating labs by a European Union project BI04 CT972132 on ‘Transposon mutagenesis in rice’. This work has been carried out in compliance with current laws governing genetic experimentation in the United Kingdom and The Netherlands.

LITERATURE CITED

Baker B, Coupland G Fedoroff N, Starlinger P, Schell J (1987) Phenotypic assay for excision of the maize controlling element *Ac* in tobacco. EMBO J 6: 1547-1554

Balcells L, Swinburne J, Coupland G (1991) Transposons as tools for the isolation of plant genes. Trends Biotechnol 9: 31-37

Barlaan EA, Ichii M (1996) Genotypic variation in nitrate and nitrite reductase activities in rice (*Oryza sativa* L). Breeding Sci 46: 125-131

Bingham PM, Levis R, Rubin GM (1981) Cloning of DNA sequences from white locus of *D.melanogaster* by a novel and general method. Cell 25: 693-704

Chin HG, Choe MS, Lee SH, Park SH, Park SH, Koo JC, Kim NY, Lee JJ, Oh BG, Yi GH, Kim SC, Choi HC, Cho MJ, Han CD (1999) Molecular analysis of rice plants harboring an *Ac*/*Ds* transposable element-mediated gene trapping system. Plant J 19: 615-623

Cho YS, Hong SK, Song MT, Moon HP, Lee JH, Kim NS (1998) Comparison of genetic variation among rice varieties detected by RAPD, AFLP and SSRP. Korean J Genetic 20: 117-128

Christensen AH, Quail PH (1996) Ubiquitin based promoters for high level of expression of selectable and or screenable marker genes in the monocotyledonous plants. Transgenic Res 5: 213-218

Davis SJ, Vierstra RD (1998) Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol Biol 36: 521-528

Devos KM, Moore G, Gale MD (1995) Conservation of marker synteny during evolution. Euphytica 85: 367-372

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol 35: 3-15

Edwards K, Johnstone C, Thompson CA (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res 19: 1349-1349

Enoki H, Izawa T, Kawahara M, Komatsu M, Koh S, Kyozuka J, Shimamoto K (1999) *Ac* as a tool for the functional genomics of rice. Plant J 19: 605-613

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6-13

Greco R, Ouwerkerk PBF, Taal AJC, Favalli C, Beguiristain T, Puigdomenech P, Colombo L, Hoge JHC, Pereira A (2001) Early and multiple *Ac* transpositions in rice suitable for efficient insertional mutagenesis. Plant Mol Biol 46:215-227

Greenblatt IM (1984) A chromosome replication pattern deduced from pericarp phenotypes resulting from movements of the transposable element *Modulator* in maize. Genetics 108: 471-485

Hamilton BA, Palazzolo MJ, Chang JH, Vijayraghavan K, Mayeda CA, Whitney MA, Meyerowitz EM (1991) Large-scale screen for transposon insertions into cloned genes. Proc Natl Acad Sci 88: 2731-2735

Haring MA, Rommens CMT, Nijkamp HJJ, Hille J (1991) The use of transgenic plants to understand transposition mechanisms and to develop transposon tagging strategies. Plant Mol Biol 16: 449-461

Izawa T, Miyazaki C, Yamamoto M, Terada R, Iida S, Shimamoto K (1991) Introduction and transposition of the maize transposable element ac in rice (*Oryza-sativa*). Mol Gen Genet 227: 391-396

Izawa T, Ohnishi T, Nakano T, Ishida N, Enoki H, Hashimoto H, Itoh K, Terada R, Wu CY, Miyazaki C, Endo T, Iida S, Shimamoto K (1997) Transposon tagging in rice. Plant Mol Biol 35: 219-229

Koes R, Souer E, Vanhouwelingen A, Mur L, Spelt C, Quattrocchio F, Wing J, Oppedijk B, Ahmed S, Maes T, Gerats T, Hoogeveen P, Meesters M, Kloos D, Mol J (1995) Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutant. Proc Natl Acad Sci 92: 8149-8153

Kohli A, Leech M, Vain P, Laurie DA, 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. Proc Natl Acad Sci 95: 7203-7208

Koprek T, McElroy D, Louwerse J, Williams-Carrier R, Lemaux PG (2000) An efficient method for dispersing *Ds* elements in the barley genome as a tool for determining gene function. Plant J 24: 253-264

Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an insertional mutagen in Arabidopsis. Plant Cell 11: 2283-2290

Li RH, Jiang TB, Xu CG, Li XH and Wang XK (2000) Relationship between morphological and genetic differentiation in rice (*Oryza sativa* L.). Euphytica 114: 1-8

Martin GB (1998) Gene discovery for crop improvement. Curr Opin Biotechnol 9: 220-226

McClintock B (1947) Cytogenetic studies of maize and Neurospora. Carnegie Inst Wash Yearbook 46: 146-152

Nakagawa Y, Machida C, Machida Y, Toriyama K (2000) Frequency and pattern of transposition of the maize transposable element *Ds* in transgenic rice plants. Plant Cell Physiol 41: 733-742

Parinov S, Sundaresan V (2000) Functional genomics in Arabidopsis: large-scale insertional mutagenesis complements the genome sequencing project. Curr Opin Biotechnol 11: 157-161

Pereira A (1998) Heterologous transposon tagging systems. In Transgenic Plant Research (Lindsey K ed.) UK: Harwood Academic Publishers, pp 91-108

Shimamoto K, Miyazaki C, Hashimoto H, Izawa T, Itoh K, Terada R, Inagaki Y, Iida S (1993) Transactivation and stable integration of the maize transposable element *Ds*-cotransfected with the *Ac*-transposase gene in transgenic rice plants. Mol Gen Genet 239: 354-360

Scholz S, Lörz H, Lütticke S (20001) Transposition of the maize transposable element *Ac* in barley (*Hordeum vulgare* L.). Mol Gen Genet 264: 653-661

Sudhakar D, Duc LT, Bong BB, Tinjuangjun P, Maqbool SB, Valdez M, Jefferson R, Christou P (1998) An efficient rice transformation system utilizing mature seed-derived explants and a portable, inexpensive particle bombardment device. Transgenic Res 7: 289-294

Sundaresan V (1996) Horizontal spread of transposon mutagenesis: New uses for old elements. Trends Plant Sci 1: 184-190

Takaiwa F, Oono K, Wing D and Kato A (1991) Sequence of 3 members and expression of a new major subfamily of glutelin genes from rice. Plant Mol Biol 17: 875-885

Takumi S, Murai K, Mori N, Nakamura C (1999) Trans-activation of a maize *Ds* transposable element in transgenic wheat plants expressing the *Ac* transposase gene. Theor Appl Genet 98: 947-953

Tyagi AK, Mohanty A (2000) Rice transformation for crop improvement and functional genomics. Plant Sci 158: 1-18

Tyagi AK, Mohanty A, Bajaj S, Chaudhury A, Maheshwari SC (1999) Transgenic rice: A valuable monocot system for crop improvement and gene research. Crit Rev Biotechnol 19: 41-79

Vain P, Worland B, Kohli A, Snape JW, Christou P (1998) The green fluorescent protein (GFP) as a vital screenable marker in rice transformation. Theor Appl Genet 96: 164-169

Yephremov A, Saedler H (2000) Display and isolation of transposon-flanking sequences starting from genomic DNA or RNA. Plant J 21: 495-505

Chapter 5

Transposon insertional mutagenesis in rice.

Raffaella Greco[1], Pieter B.F. Ouwerkerk[2], Christophe Sallaud[3], Ajay Kohli[4], Lucia Colombo[5], Pere Puigdomènech[6], Emmanuel Guiderdoni[3], Paul Christou[4], J.Harry C. Hoge[2†] and Andy Pereira[1]

[1]Plant Research International, Wageningen, The Netherlands
[2]Institute of Biology, Leiden University, Leiden, The Netherlands
[3]CIRAD-BIOTROP, Montpellier, France
[4]John Innes Centre, Norwich, UK
[5]Dept. of Biology, University of Milan, Milano, Italy
[6]CID-CSIC, Barcelona, Spain

Plant Physiology 125: 1175-1177 (2001)

INTRODUCTION

Transposon mutagenesis systems are being developed in rice to address gene functions using forward and reverse genetics strategies. Although the well-characterized maize transposons seem an obvious choice to develop efficient transposon tagging systems for gene knockouts and gene detection, heterologous transposons have not yet become a genome-wide system for saturation mutagenesis in rice. Our contributions as a multinational European initiative in developing heterologous transposon mutagenesis strategies for functional genomics in rice are described here.

The endogenous retrotransposon *Tos17* (Hirochika, 1997) has been effectively used to make knockouts of rice genes in a reverse genetics approach, but has a high frequency of untagged mutations probably due to somaclonal variation associated to the tissue culture based retrotransposition generation. Recent reports on transposition of *Ac-Ds* heterologous systems in rice have shown their activity and potential use as effective insertional mutagens (Izawa et al., 1997; Chin et al., 1999; Enoki et al., 1999; Nakagawa et al., 2000). The autonomous *Ac* has been shown to maintain a high rate of transposition, though *Ds* elements might undergo silencing in later generations. The use of both systems in reverse genetic approaches has been reported to be potentially efficient.

One principle we sought to employ was to generate rice genotypes with multiple transposons that could effectively saturate the genome with a fewer number of plants. Transformation of T-DNA bearing transposons normally generates a few copies; thus transposon amplification was used to increase copy number.

KNOCKOUT MUTATION MACHINE

We developed a GFP excision assay that allowed the identification of transposon excision events. In a construct with *Ac* containing a double CaMV 35S enhancer adjacent to the autonomous *Ac* promoter (Fig 1), all transformants generated showed very early transposition. About half the lines also showed amplification of *Ac* copy number in the first cell after transformation. This phenomenon could be attributable to the influence of the 35S enhancer on the adjacent endogenous *Ac* promoter, inducing high level of transposase expression, in combination with the transposition of *Ac* during replication (Greco et al., 2001). This allowed the generation of lines containing multiple transposons (ca. four from a single copy T-DNA line) that could generate an average of one to two new inserts per progeny giving a frequency of 15 to 50% of independent transpositions in the next generation (T_1).

Transposition of *Ac* to linked position in the genome, applicable for targeted tagging, was demonstrated by the recovery of a set of six insertions in a 70 Kb interval of sequenced DNA on chromosome 6. Unlinked transposition was also observed by segregation of the T-DNA locus from the *Ac* insertions. The isolation of *Ac* flanking genomic sequences revealed a preferential insertion in protein-coding sequences, as shown by the presence of four times more insertions in genes than randomly expected. This confirms the earlier results of *Ac* insertional preference in rice, in which 4% of the inserts were observed to be in sequences homologous to ESTs (Enoki et al., 1999), and suggests that the preferential transposon insertions in genes could be a valuable asset for generating mutants in rice.

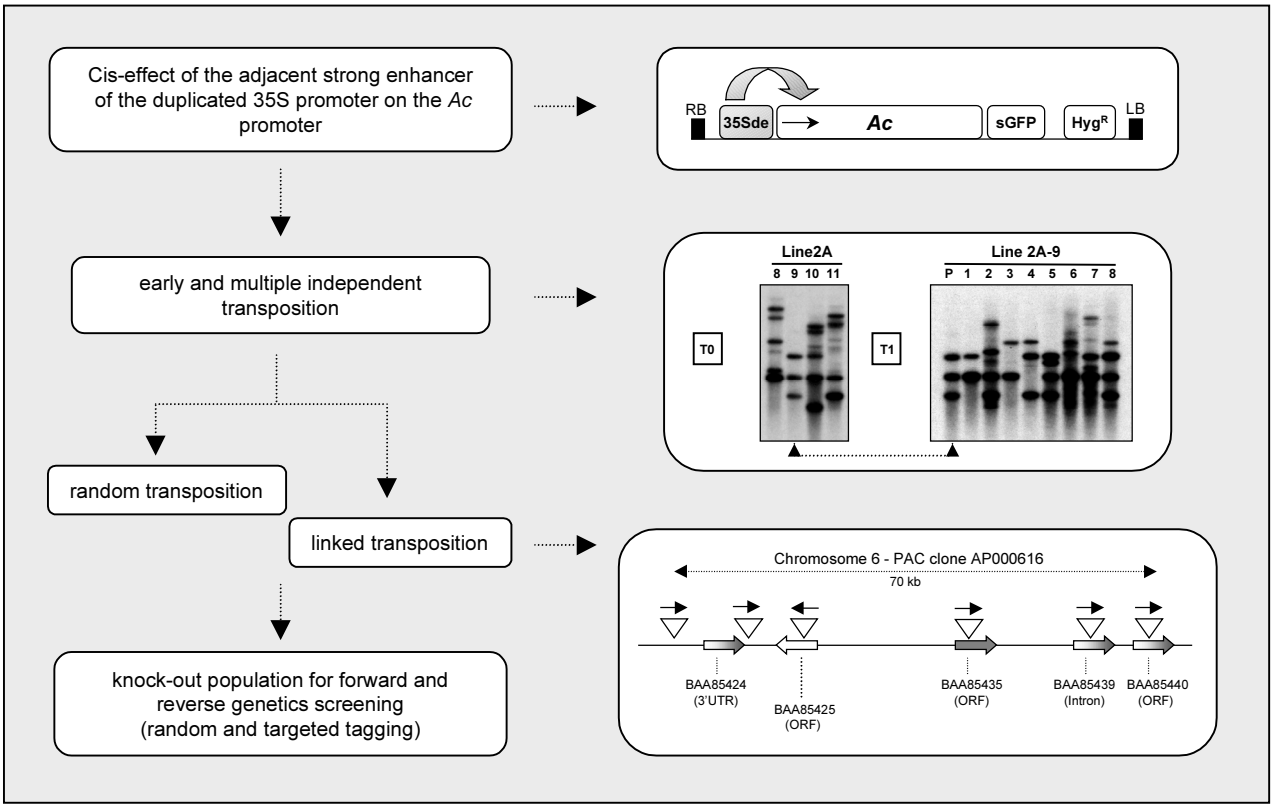


Figure 1. Generation of a knockout population using multiple copy *Ac* lines. The construct used to introduce *Ac* in rice conferring early transposition. The Southern blot represents different regenerants (T_0) from a single copy line (2A) and the T_1 progeny of plant 2A-9, showing transposon amplification and active transposition of the multiple copies. Shown below is local transposition in genes (thick arrows) on chromosome 6, as revealed by sequencing *Ac* insertion sites (triangles) from Line 2A.

Multiple transposon lines and gene insertional specificity enable generation of a tagging population requiring a reduced number of lines to reach saturation. The propagation of these genotypes for three to four generations can generate a population of plants containing four or more *Ac* inserts at different positions in the genome. Using 25,000 lines for three to four generations would generate about 100,000 insertions that are suitable for identifying knockouts for forward as well as reverse genetic strategies. Due to the ability of *Ac* to autonomously transpose, identified knockout insertions will be unstable. For insertions in the coding region of genes, *Ac* excision can produce footprints that can lead to generation of mutant and revertant alleles. These mutant and revertant alleles derived from an identical genetic background can be used to assess the phenotypic effect of the gene mutation (stable), helping reduce the effect of background mutations.

The scope of knockout mutations is limited as the majority of genes display no obvious phenotype, probably due to functional redundancy, in which one or more other homologous loci can substitute for the same function. A way to circumvent this, is the sequential disruption of redundant genes in an individual genotype that might ultimately reveal a mutant phenotype and uncover the gene function.

GENE DETECTION STRATEGY

Gene detection strategies have been developed to address the function of genes that do not directly reveal a knockout phenotype. One way is expression detection that can make use of inserts containing reporter gene constructs such as enhancer traps or gene traps, whose expression depends on transcriptional regulatory sequences of the adjacent host gene. Another way is by creating misexpression mutants, such as activation tags that might reveal a gain-of-function phenotype.

Constructs were made with the aim of generating populations of transposon inserts for knockout mutagenesis as well as gene detection using enhancer traps (ET) and activation tags (AT). The general structure of the ET and AT constructs is outlined in Figure 2A. The *Ds* mobile transposon contains a BAR gene conferring resistance to the herbicide Basta with a GUS reporter gene (with minimal promoter) in the ET constructs or a multiple CaMV 35S enhancer in the AT constructs. The T-DNA contains an immobile *Ac* transposase under control of a strong promoter and a negative selection marker to allow selection of transposase free segregants in a later phase. Using a combination of these greenhouse/field selectable markers, progeny of single locus transformants can be used to identify stable transposants, where the *Ds*-BAR transposes from the T-DNA to unlinked positions. From the ET and AT lines generated in Nipponbare about 50% show transpositional activity. The most active single locus lines containing multiple

transpositions (Fig 2B) are being propagated to advanced generations to make a transposon insertion library.

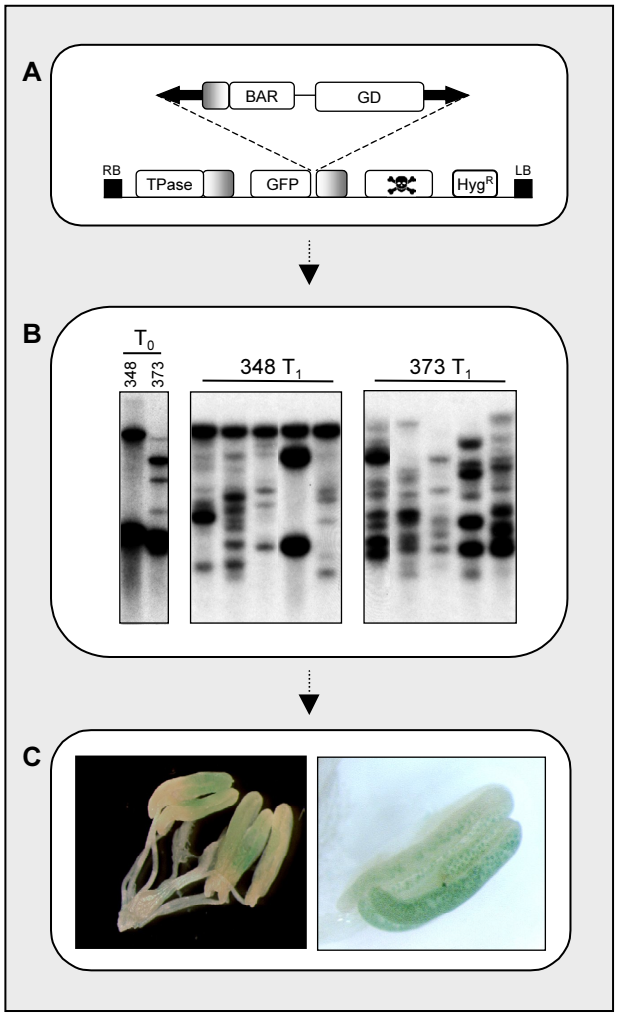


Figure 2. Gene detection strategy. A) Example of a two component *Ac*/*Ds* construct in which the mobile *Ds* transposon contains the sequences for gene detection (GD) and a transposition marker (BAR), and the immobile T-DNA contains the *Ac* transposase source (TPase), an excision marker (GFP), a negative selection marker and a selectable marker for transformation (Hyg^R). B) Example of two single T-DNA copy enhancer trap lines, which display amplification of *Ds* copy number in T_0 and T_1 generation. C) Example of pollen specific GUS staining in anthers of a mature rice flower, identified in a line containing an *Ac*/*Ds* enhancer trap construct.

The *Ds*-ET inserts contain a minimal promoter that can initiate transcription upstream of the GUS marker gene upon insertion near enhancers of host genes in the genome, thus displaying the expression pattern of the adjacent gene (Fig 2C). Such patterns will help identify the adjacent plant gene on the basis of its expression. Gene detection systems in rice have been developed by Chin et al. (1999) using a gene trap *Ds* construct containing a promoter-less GUS reporter gene with upstream splice acceptor sites, which revealed a frequency of detection of gene expression of trapped genes comparable to Arabidopsis. In a gene trap, expression of the reporter gene occurs only upon *Ds* insertion in the correct orientation within a gene. On the other hand enhancer traps are subjected to less constraint and have therefore a higher probability of

detecting expression patterns, making their use especially effective for identification of gene functions in reverse genetics strategies.

Although the frequency of *Ds* transposition in early generations appears to be high, inhibition of transposition is reported in later generations (Izawa et al., 1997). Though this might happen in some lines, active lines have been identified which still show good transposition activity in the T_2/F_3 generation (unpublished data; Nakagawa et al., 2000), suggesting that *Ds* inactivation may not be a general phenomenon. Transposon populations of stable *Ds* insertions sufficient for genome saturation can indeed be generated in a few generations starting from a minimal number of active multiple transposons lines. Assuming preferential transposition of *Ds* in genes as described for *Ac*, a collection of around 100,000 inserts would be adequate to ensure the tagging of almost every gene. Such a population could be produced in three generations after transformation and seed multiplication from 10 starter lines, considering a 20% transposition rate per insertion, four inserts per plant and a 20% frequency of transposition to unlinked positions. By a concerted international effort, the transposon libraries can be produced and made available to rice researchers worldwide.

Acknowledgements

This research was funded in the participating labs by a European Union project BIO4 CT 972132 on ‘Transposon mutagenesis in rice’.

This paper is dedicated to the loving memory of our colleague Dr J. Harry C. Hoge who passed away during the course of this research.

LITERATURE CITED

Chin HG, Choe MS, Lee S-H, Park SH, Koo JC, Kim NY, Lee JJ, Oh BG, Yi GH, Kim SC, Choi HC, Cho MJ, Han CD (1999) Molecular analysis of rice plants harboring an *Ac/Ds* transposable element-mediated gene trapping system. Plant J 19: 615-623

Enoki H, Izawa T, Kawahara M, Komatsu M, Koh S, Kyozuka J, Shimamoto K (1999) *Ac* as a tool for the functional genomics of rice. Plant J 19: 605-613

Greco R, Ouwerkerk PBF, Taal AJC, Favalli C, Beguiristain T, Puigdomènech P, Colombo L, Hoge JHC and Pereira A (2001) Early and multiple *Ac* transpositions in rice generated by an adjacent strong enhancer. Plant Mol Biol 46: 215-227

Hirochika H (1997) Retrotransposons of rice: their regulation and use for genome analysis. Plant Mol Biol 35: 231-240

Izawa T, Ohnishi T, Nakano T, Ishida N, Enoki H, Hashimoto H, Itoh K, Terada R, Wu C, Miyazaki C, Endo T, Iida S, Shimamoto K (1997) Transposon tagging in rice. Plant Mol Biol 35: 219-229

Nakagawa Y, Machida C, Machida Y, Toriyama K. (2000) Frequency and pattern of transposition of the maize transposable element *Ds* in transgenic rice plants. Plant Cell Physiol 41: 733-42

Chapter 6

Transpositional behaviour of an *Ac/Ds* system for reverse genetics in rice.

Raffaella Greco[1], Pieter BF Ouwerkerk[2], Rolf J De Kam[2], Christophe Sallaud[3], Cristina Favalli[4], Lucia Colombo[4], Emmanuel Guiderdoni[3], Annemarie H Meijer[2], J Harry C Hoge[2†] and Andy Pereira[1]

[1] *Plant Research International, Wageningen, The Netherlands*
[2] *Institute of Biology, Leiden University, Leiden, The Netherlands*
[3] *CIRAD-BIOTROP, Montpellier, France*
[4] *Dept. of Biology, University of Milan, Milano, Italy*

Theoretical and Applied Genetics
DOI: 10.1007/s00122-003-1416-8

ABSTRACT

A collection of transposon *Ac*/*Ds* enhancer trap lines is being developed in rice that will contribute to the development of a rice mutation machine for the functional analysis of rice genes. Molecular analyses revealed high transpositional activity in early generations, with 62% of the T_0 primary transformants and more than 90% of their T_1 progeny lines showing ongoing active transposition. About 10% of the lines displayed amplification of *Ds* copy number. However, inactivation of *Ds* seemed to occur in about 70% of the T_2 families and in the T_3 generation. Southern blot analyses revealed a high frequency of germinal insertions inherited in the T_1 progeny plants and transmitted preferentially over the many other somatic inserts to later generations. The sequencing of *Ds* flanking sites in subsets of T_1 plants indicated the independence of insertions in different T_1 families originating from the same T_0 line. Almost 80% of the insertion sites isolated showing homology to the sequenced genome resided in genes or within a range at which neighbouring genes could be revealed by enhancer trapping. A strategy involving the propagation of a large number of T_0 and T_1 independent lines is being pursued to ensure the recovery of a maximum number of independent insertions in later generations. The inactive T_2 and T_3 lines produced will then provide a collection of stable insertions to be used in reverse genetics experiments. The preferential insertion of *Ds* in gene rich regions and the use of lines containing multiple *Ds* transposons will enable the production of a large population of inserts in a smaller number of plants. Additional features provided by the presence of *lox* sites for site-specific recombination or the use of different transposase sources and selectable markers are discussed.

***Key words**: Ac/Ds transposon system, reverse genetics, functional genomics, insertion specificity, rice*

This report is dedicated to the loving memory of our colleague Dr J. Harry C. Hoge.

INTRODUCTION

Cereals are the staple food for more than 90% of the world population and rice alone feeds more than three billion people. Due to its smaller genome size (Arumuganathan and Earle 1991), the availability of dense genetic and physical maps (Chen et al. 2002; Harushima et al. 1998), the ease of transformation (Hiei et al. 1994) and the extensive synteny shared with the other cereals (Gale and Devos 1998), rice has become a model system for monocot species (Izawa and Shimamoto 1996). Deciphering the function of all rice genes is an important step towards the identification of genes of key agronomic importance for the improvement of rice itself and of the other major staple cereals like maize, wheat and barley, in terms of yield and nutritional values.

The complete sequence of the rice genome is now available, thanks to public and private efforts. In the last years, draft genome sequences have been produced for Nipponbare (Barry 2001; Goff et al. 2002) and 93-11 (Yu et al. 2002), cultivars belonging to the two major rice subspecies *japonica* and *indica*, respectively. Concurrently, the International Rice Genome Sequencing Project (IRGSP) has published the complete sequence of Nipponbare chromosomes 1 and 4 (Sasaki et al. 2002; Feng et al. 2002) and recently announced the release of a high-quality draft sequence of the whole rice genome (Buell 2002; http://rgp.dna.affrc.go.jp/rgp/Dec18_NEWS.html). Overall, some 32,000-55,000 rice genes have been predicted, many of them requiring experimental confirmation. Valuable information on the actual number of expressed genes can be obtained from the large EST sequencing projects that are being undertaken by several groups using different rice varieties (Bohnert et al. 2001; Reddy et al. 2002; Wu et al. 2002; Yamamoto and Sasaki 1997; Beijing Genomics Institute at http://btn.genomics.org.cn/rice/index.php). However, most of the genes identified will still lack a putative function. Unravelling their biological role will then be the next compelling and challenging step.

Among the several functional genomics approaches available (reviewed in Pereira 2000), insertional mutagenesis using T-DNA or transposons (retrotransposons or heterologous transposons) is considered to be a very promising tool for high-throughput gene function discovery in rice (reviewed in Jeon et al. 2001). Several collections of rice lines containing the heterologous maize *Ac*/*Ds* transposon system are being established. Both the autonomous *Ac* element (Enoki et al. 1999; Greco et al. 2001b; Kohli et al. 2001) and a two component system based on *Ds* transposition induced by an immobilised *Ac* transposase source (Izawa et al. 1997; Chin et al. 1999; Nakagawa et al. 2000; Upadhyaya et al. 2002), are being used. The latter collections have the advantage that the *Ds* insertions could be stabilised by segregation, allowing easier identification of tagged genes and the construction of insertional databases. In addition, gene and enhancer trapping features (reviewed in Springer 2000) were incorporated in the

transposon vectors, to address the function of genes that would fail to display a visible knockout phenotype, therefore increasing the effectiveness of gene tagging (Chin et al. 1999; Upadhyaya et al. 2002). A *GUS* reporter gene with upstream splice acceptor sites (gene trap) or driven by a minimal promoter (enhancer trap) was placed in this case near the end of the *Ds* transposon. Its expression relies upon the presence of adjacent regulatory sequences, thus revealing the pattern of expression of neighbouring host genes.

Although the transposition rate of *Ac*/*Ds* in rice and the high frequency of germinal insertions support its use as effective mutagenic agent, indications of inactivation of *Ds* in later generations have been reported (Chin et al. 1999; Izawa et al. 1997) that might hamper its efficiency for use in high-throughput functional genomics screenings. However, epigenetic inactivation of endogenous transposable elements is a naturally occurring phenomenon, not restricted to plants, associated with DNA methylation and most probably related to transposon control (reviewed in Martienssen and Colot 2001). Cyclic inactivation of *Ac* or *Ds* transposons was initially described in maize by Barbara McClintock (1959) and since then numerous studies have been carried out, but the molecular mechanisms underlying this process remain still largely unknown. In rice, Izawa et al. (1997) analysed the behaviour of *Ds* in progeny plants originating from a cross between lines carrying a CaMV 35S-immobilized *Ac* transposase (genomic clone) or a non-autonomous *Ds*, obtained by direct DNA transfer. A decrease in the frequencies of *Ds* transpositions was often observed already in F_2 and later, up to the F_5 generation. Among the *Ac*/*Ds* rice populations originated by Agrobacterium-mediated transformation experiments, Chin et al. (1999) employed a *Ds*-gene trap system with a 35S-*Ac* cDNA as a transposase source in the same T-DNA, and monitored transpositions only in T_0 primary transformants and after repetitive ratooning. Nevertheless they observed a decrease in the frequency of secondary transposition, with only 18% of *Ds* elements that showed primary transposition in the T_0 regenerants (80%) being remobilised in the third ratoon generation. Other reports rely on the analysis of a restricted number of successive generations (Nakagawa et al. 2000; Upadhyaya et al. 2002), in which *Ds* still retains transpositional activity. Although more confirmatory data need to be gathered about the behaviour of *Ds* in advanced generations, loss of mobility seems to be an unquestionable fact that has to be taken into account in order to devise efficient tagging strategies using this transposon system.

In this paper we describe further advances in the development of an *Ac*-*Ds* enhancer trap collection by a multinational European Consortium (Greco et al. 2001a) and its evaluation as an efficient tagging tool, in terms of transposition behaviour over successive generations and in relation to gene tagging and enhancer trapping. Using these public resources, a strategy is proposed for the use of these lines for the development of a rice mutation machine for the functional analysis of all rice genes.

MATERIAL AND METHODS

Ac/Ds Enhancer Trap constructs

A diagram of the two enhancer trap constructs employed (ET1 and ET2) is shown in Figure 1. The two constructs differ only for the presence in the latter of the sGFP excision marker. T-DNA: The immobile transposase extending from the *Bst*NI site at position 939 of *Ac* (Genbank Accession X05424) till the end of the element was obtained from a derivative of pKU2 (Baker et al. 1987). This promoter-less *Ac* fragment was fused to a 0.55 kb CaMV 35S promoter fragment derived from pDH51 (Pietrzak et al. 1986) and subsequently excised as a unique *Sal*I-*Not*I fragment. The SU1 gene was obtained from pSSU-SU12 (O'Keefe 1994) as *Eco*RI-*Bam*HI (construct ET1) or *Eco*RI-*Sal*I (construct ET2) fragment. In case of construct ET2, the mobile *Ds* was inserted in between a *Not*I-*Apa*I doubly enhanced CaMV 35S promoter and AMV leader (Sijmons *et al*., 1990) and a *Bam*HI-*Hin*dIII sGFP(S65T)-*Nos* terminator (Chiu *et al*., 1996) fragments. In case of construct ET1, a fragment containing a CaMV 35S terminator and a *lox* site (Hoess et al. 1982) was used as a *Not*I-*Apa*I linker. Mobile Ds: The *5'* terminus of *Ac* till the *Ba*lI site at position 252 (*Ds* left junction) was fused to a intron-less maize Ubiquitin promoter-*BAR*-*Nos* terminator cassette, originally derived from pAHC25 (Christensen and Quail 1996), and subsequently excised as a unique *Apa*I-*Spe*I fragment. A 0.94 kb fragment extending up to the *Bgl*II site in the first exon of the Ubiquitin promoter was used (Christensen et al. 1992). The 3' terminus of *Ac* from the *Pac*I site at position 4302 till the end (*Ds* right junction) was fused to a 61 bp minimal CaMV 35S promoter fragment, a 48 bp oligonucleotide comprising a *lox* site, and a promoter-less *GUS* reporter gene-*Nos* terminator (Jefferson et al. 1987). The whole cassette was then excised as a single *Bam*HI-*Xba*I fragment.

The constructs were assembled by multi-point ligation in the binary vector pMOG22 (Zeneca-MOGEN, Netherlands), containing the hygromycin phosphotranferase (HPT) gene for selection of plant transformants.

Plant transformation and growth conditions

Agrobacterium-mediated transformation of *Oryza sativa* ssp. *japonica* cv. Nipponbare, plant regeneration and growth were performed as previously described (Greco et al. 2001b). Three transformation experiments were carried out, using construct ET1 (first) and construct ET2 (ET2-a: second; ET2-b: third) in *Agrobacterium tumefaciens* LBA4404. In the third experiment, DNA isolation from an aliquot of the cultured Agrobacterium strain containing the transposon construct was performed before co-cultivation, in order to assess the presence of an intact T-DNA in the binary vector.

Marker expression analysis

sGFP: sGFP expression in calli and transgenic plants containing construct ET2 was monitored as previously described (Greco et al. 2001b). BAR: Dehusked seeds were surface-sterilized (2 min EtOH 100%, 1 h NaOCl 5% + Triton X-100 0.1%, 6X rinse with H_20) and sown on 1X MS, 1% sucrose, 0.8% purified agar with 15 mg/L phosphinothricin (Duchefa Biochemie, The Netherlands). SU1: Greenhouse seedlings were sprayed with the pro-herbicide R7042 (DuPont, Wilmington, DE) as described by Koprek et al. (1999). GUS: Histochemical GUS staining was performed as described by Scarpella et al. (2000). Plant tissues were incubated in the staining solution for up to 2 days. Chlorophyll was then removed with 70% (w/v) ethanol and samples stored at 4°C.

Molecular analysis of Ds transposition

Genomic DNA isolations were performed from leaf samples collected either from *in vitro* young plantlets or from adult greenhouse plants, according to Pereira and Aarts (1998), in Eppendorf tubes or in 96 tube-racks after dry grinding using a Mixer Mill MM300 (RETSCH, Germany) with tungsten carbide beads. PCR analysis and Southern blot hybridisation were then employed to analyse transposition of *Ds* at molecular level. For simplicity, only the analysis of plants containing construct ET2 is described in details. PCR analysis: A preliminary PCR analysis was performed in order to select for transgenic lines with intact T-DNA, using primers for the *HPT* gene (HPT-for: 5'-AAAAGTTCGACAGCGTCTCCGACC-3' and HPT-rev: 5'-TCTACACAGCCATCGGTCCAGACG-3'), the immobile *Ac* transposase (Ac971-F: 5'-ACGACTCCATTCCTCAGATGACG-3' and Ac1395-R: 5'-CTTGACTCGGATCTGTAGCTGTACC-3') and the *Ds* transposon (Ubi874-F: 5'-TAAATAGACACCCCCTCCACACC-3' and BAR-R1: 5'-CAGGCTGAAGTCCAGCTGCCAG -3'). For a preliminary screening of excision events, an Empty Donor Site (EDS) fragment of 0.6 kb is expected to be amplified in case of *Ds* excision, using primers in the 35S promoter (35S-for: 5'-ATCCCACTATCCTTCGCAAGACCC-3') and in the *sGFP* gene (sGFP-R2: 5'-GCTTGTCGGCCATGATATAGACG-3'). In case of untransposed *Ds*, Full Donor Site fragments of 0.3 kb and 0.7 kb (left and right junction respectively) will be amplified, using the 35S-for primer in combination with the transposon specific primer *Ds150*-R (5'-GTTTCCGTTTCCGTTTACCGTTTT-3') or the transposon primer *Ds4374*-F (5'-GAACAAAAATACCGGTTCCCGTCC-3') with sGFP-R2. About 20 ng of DNA and standard PCR conditions were used. Southern blot analysis: About 3 μg of genomic DNA were digested with *Eco*RI and subjected to Southern blot analysis as previously described (Greco et al. 2001b). In order to monitor excision events, blots were hybridised with a *sGFP* probe, expecting to detect a 1.85 kb fragment in case of EDS and/or a 3.3 kb fragment in case of FDS (see Fig. 1). Transposed *Ds* elements were revealed after re-hybridisation of the same blots with a *GUS*

probe, as fragments with size larger than 2.3 kb. Untransposed *Ds* will also be revealed as a 3.3 kb FDS band. The number of inserted copies of the original T-DNA was calculated as the number of *HPT*-hybridising fragments. In addition, the *BAR* gene and an *Ac* internal fragment were used to confirm the presence of the *Ds* transposon and the *Ac* transposase respectively (data not shown).

Isolation of Ds flanking sites

Genomics sequences flanking the *Ds* insertions were isolated either by TAIL or Adapter PCR, and then compared to known sequences in public databases using the BLAST algorithm (Altschul et al. 1997). TAIL PCR: TAIL PCR was performed according to Tsugeki et al. (1996) on 10 ng of genomic DNA. The *Ds* transposon primers and the arbitrary degenerate primers used were as reported in Tsugeki et al. (1996) and Liu and Whittier (1995). A re-PCR was generally performed before sequencing the amplified fragments, on agarose-picked tertiary products. Adapter PCR: The adapter PCR was performed according to Tissier et al. (1999) on 400 ng of genomic DNA digested with *Bfa*I and ligated to the adapter. The *Bfa*I adapter primer was used in combination with *Ds* specific primers annealing to the right junction of the transposon (*Ds*3-1 to 4; Tsugeki et al. 1996). A biotinylated *Ds3*-1 primer was used in the first PCR, *Ds3*-2 in the nested PCR, *Ds3*-3 in the reamplification of the eluted fragments and *Ds3*-4 for sequencing.

RESULTS

Enhancer trap constructs

In the two-component *Ac*/*Ds* system employed, an immobilised *Ac* transposase driven by the CaMV 35S promoter was used to mobilise a *Ds* non-autonomous element. These two transposon components together with appropriate markers were placed in the same T-DNA vector, so that transposition could already occur directly after transformation in the transgenic calli and the regenerating T_0 plants without need of further crossing.

Two enhancer trap constructs were made, namely ET1 and ET2 (Figure 1). Both constructs bear the phosphinothricin acetyltransferase gene (*BAR*; DeBlock et al. 1987) for resistance to the herbicide Basta as a marker for positive selection on the mobile *Ds* transposon. To select against the *Ac* transposase source, the cytochrome P450 gene (*SU1*; O'Keefe 1994) from *Streptomyces griseolus*, which converts the pro-herbicide R7042 (DuPont, Wilmington, DE) into a cytotoxic form, is present on the T-DNA. In construct ET2, the *Ds* transposon is inserted between the CaMV 35S promoter and the *sGFP* gene to monitor for excision events. Both constructs carry the

GUS reporter gene driven by a minimal CaMV 35S promoter on the mobile *Ds*, for enhancer trapping. Additionally, construct ET1 bears *lox* sequences for site-specific recombination (reviewed in Ow and Medberry 1995; Ow 1996) in the *Ds* transposon and the T-DNA. Upon crossing with lines expressing a Cre-recombinase, recombination can occur between two *lox* sites. Depending on their mutual orientation, chromosomal rearrangements like deletions, inversions or translocations can be induced.

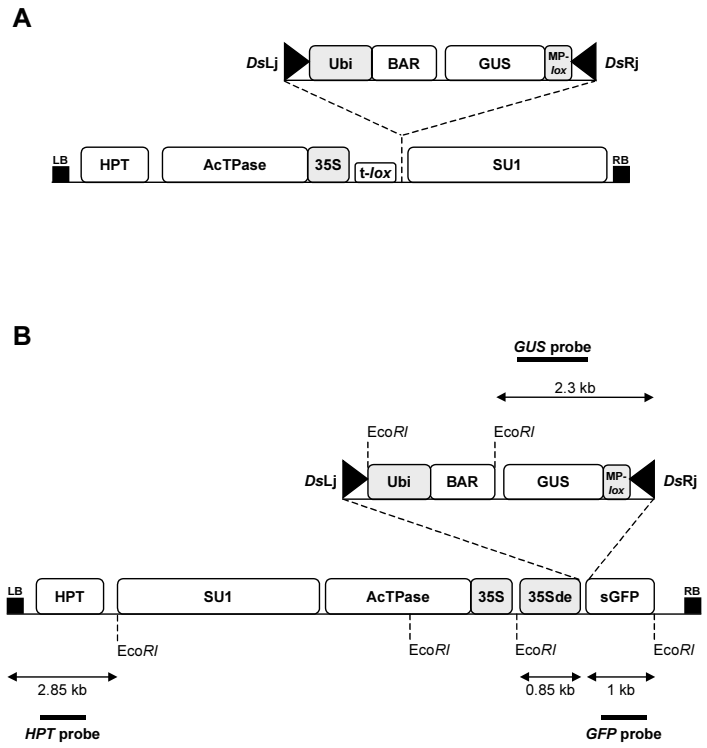


Figure 1. Schematic representation of the enhancer trap constructs used. A) Diagram of ET1. B) Diagram of ET2, with the position of the *Eco*RI restriction sites and the probes used for Southern blot analysis. The *GFP* probe is expected to reveal a 1.85 kb hybridizing fragment in case of excision (EDS), or a 3.3 kb FDS fragment otherwise. Similarly, the *GUS* probe will detect the 3.3 kb FDS fragment if present, or any transposed *Ds* as hybridizing bands bigger than 2.3 kb. The number of *HPT*-hybridising fragments (bigger than 2.85 kb) estimates the number of integrated T-DNA copies. LB, RB, left and right T-DNA borders; HPT, hygromycin phosphotransferase gene; *Ac*TPase, immobilised *Ac* transposase; 35S, CaMV 35S promoter; 35Sde, doubly enhanced CaMV 35S promoter; *Ds*LJ, *Ds*RJ, left and right *Ds* ends; Ubi, maize Ubiquitin promoter; MP-*lox*, CaMV 35S minimal promoter and *lox* site; t-*lox*, CaMV 35S terminator and *lox* site.

Ac/Ds transposition activity in the T_0 primary transformants

Transgenic Nipponbare rice plants containing either constructs ET1 or ET2 were obtained from three independent Agrobacterium-mediated transformation experiments. In total about 1000 T_0 plants were regenerated from 144 independent T_0 hygromycin-resistant calli (henceforth referred to as T_0 original lines), of which 219 were selected as representative sample for molecular analysis. Plants were subjected to PCR and Southern hybridisation to test initially for the presence of each of the components of the original T-DNA (see Material and Methods; data not shown) and then specifically for *Ds* excision, reinsertion and copy number. The results of the

analyses are summarised in Table 1, with the constructs used and the transformation experiments described separately.

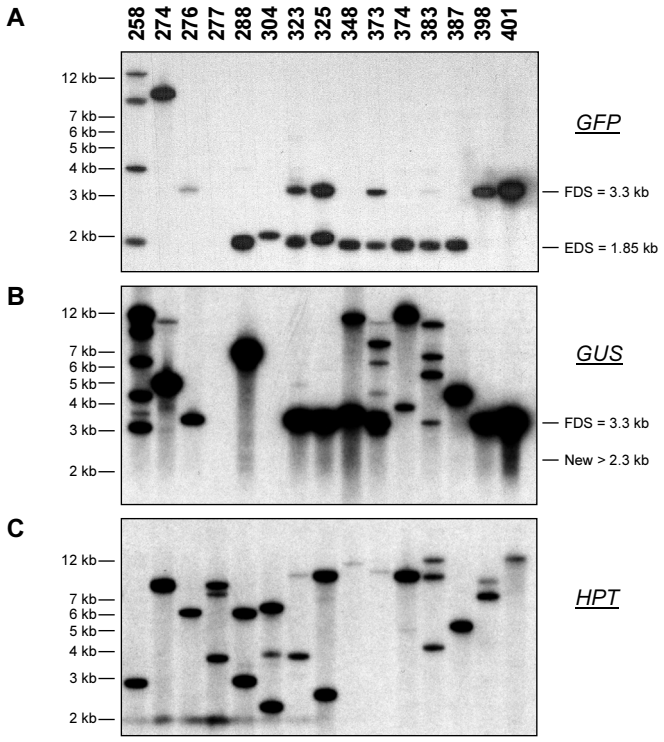


Figure 2. Southern blot analysis of ET2-b enhancer trap lines in T_0 generation. One T_0 regenerant per line is represented. A) Hybridization with a *sGFP* probe detects excision events (EDS) and/or FDS. B) Transposition was monitored by hybridisation with a *GUS* probe. Newly transposed *Ds* (>2.3 kb) and/or *Ds* still residing in the donor T-DNA (FDS) are revealed. C) The number of T-DNA copies integrated in the rice genome is estimated by the number of fragment hybridising to an *HPT*-homologous probe.

Figure 2 shows an example of Southern analysis done on a subset of 15 T_0 lines from the second transformation experiment with construct ET2 (only 1 regenerant per line was analysed in this case). One set of lines showed very early excision (lines 288, 348, 374, 387), probably directly after transformation, as revealed by the presence of only EDS (Fig. 2A). Reinsertion of the excised *Ds* elements at different genomic positions is revealed after hybridisation with a *GUS* probe (Fig. 2B). Among these lines, amplification of *Ds* copy number occurred in lines 348 and 374, since only one *HPT*-hybridising band corresponding to a single T-DNA copy is visible (Fig. 2C), while two *Ds* homologous fragments are detected (Fig. 2B). Another set of lines carried both EDS and FDS (323, 325, 373, 383), indicating later or partial excision in some of the T-DNA copies. Among them, only 373 and 383 revealed *GUS*-hybridising bands corresponding to transposed *Ds* elements (Fig. 2B), which implies a loss of the excised *Ds* in the other two lines. Finally, lines 276, 398 and 401 seemed transpositionally inactive, as revealed by the presence of only FDS after hybridisation with *sGFP* and *GUS*, while another four lines underwent rearrangements (258, 274, 277, and 304).

Overall, the percentage of lines with a correct integration of the original T-DNA appeared to be dependent on the construct used, with a decrease of up to 35% in the first transformation experiment using construct ET2 (ET2-a; Table 1). The lack of complete T-DNA inserts in transformants with this construct turned out to be related to its instability in Agrobacterium. Therefore a pre-selection step was introduced in the next rice transformation experiment at the time of co-cultivation, to ensure that only Agrobacterium cultures with intact T-DNA were used during transformation. In this case, the proportion of lines with undamaged constructs could be raised to 73% (ET2-b; Table 1).

Table 1. Summary of T_0 enhancer trap (ET) lines produced and analysed [a].

Generation	Construct[b]	Analysed	Entire T-DNA	%	Active	%	Single-copy
T_0	ET1	10 {26}	8 {23}	80 {88}	4 {9}	50 {39}	2
	ET2-a	26 {85}	9 {24}	35 {28}	3 {12}	33 {50}	1
	ET2-b	108 {108}	79 {79}	73	53 {53}	67	35
	ET total	144 {219}	96 {126}	67 {58}	60 {74}	62 {59}	38

[a] The numbers reported for each construct correspond to T_0 original lines and T_0 regenerants (brackets), respectively.
[b] Construct ET2 was used in two separate transformation experiments, a and b.

Of the T_0 enhancer trap lines that were not rearranged, 62% showed transpositional activity judged by the presence of *Ds* excision and reinsertion (Table 1). The timing of transposition varied among lines, but also among plants regenerated from the same line, and appeared to occur already in the transgenic calli or during the regeneration of the plantlets. Regenerants with the same pattern of transposition indicate excision events that took place early during callus development and before the formation of the regenerated shoots, while independent insertions result from later excisions in the individual regenerated plants. The percentage of lines showing complete excision (very early, probably at the time of transformation) at the time of the analysis amounted to approximately 55-60%.

On average, the number of integrated T-DNA copies in all the T_0 lines analysed varied between 1 to 3, with about 60% lines being single copy and less than 5% having more than 3 copies (data not shown). In about 10-15% of the lines amplification of *Ds* transposon copy number was observed, as revealed by multiple *Ds* insertions of same intensity as a single copy T-DNA/*Ds* insert.

As a preliminary indication for the occurrence of excision events, calli and regenerant plants were screened for GFP expression. GFP fluorescence is visualised as a result of excision of the

Ds transposon from the T-DNA, which allows the CaMV 35S promoter to drive the expression of the *sGFP* gene. Of 74 lines displaying somatic or germinal excision, only 33 (45%) displayed GFP-positive sectors in calli growing on embryo-induction medium and hardly any GFP activity could be detected in the regenerating plantlets (data not shown). The unexpectedly low percentage of detectable green fluorescent sectors seemed to suggest a low activity of the *Ds* excision marker.

Ac/Ds transposition activity in T_1, T_2 and T_3 generations

To assess the utility of the *Ac-Ds* enhancer trap lines for tagging purposes, transposition activity was monitored in advanced generations. A total of 34 T_0 lines (90 T_0 regenerants) were selected and propagated in T_1, on basis of their transposition behaviour. Among them, 24 showed active excision and transposition in T_0, while 10 displayed only late somatic excision or were completely inactive (visualised as presence of a strong FDS with faint or no *Ds* reinsertion bands on a Southern blot hybridised with a *GUS* probe). In order to propagate only plants containing the *Ds* transposon, T_1 seeds were germinated *in vitro* on medium containing phosphinothricin (PPT) before transfer to the greenhouse. The selection allows the growth of progeny plants bearing *Ds* elements either transposed or not excised. On the other hand, it will prevent the recovery of plants in which the *Ds* transposon is lost due to excision events not followed by reinsertion. A total of 696 T_1 BAR^+ plants were recovered and subjected to molecular analysis by PCR and Southern hybridisation, essentially as described for the T_0 generation.

Table 2. Summary of activity of ET lines through successive generations [a].

Generation	Construct	Analysed	Exp. active	%	Active	% [b]
T_1	ET1	5 {11} [94]	3 {9} [79]	60 {82} [84]	2 {8} [52]	67 {89} [66]
	ET2-a	4 {12} [106]	1 {9} [85]	25 {75} [80]	1 {9} [59]	100 {100} [69]
	ET2-b	25 {67} [496]	20 {57} [449]	80 {85} 90]	19 {45} [214]	95 {79} [48]
	ET total	34 {90} [696]	24 {75} [613]	71 {83} [88]	22 {62} [325]	92 {83} [53]
T_2	ET1	2 {5} [9] (26)	all	100	1 {1} [1] (3)	50 {20} [11] (11)
	ET2-a	1 {8} [20] (104)	all	100	1 {5} [5] (7)	100 {62} [25] (7)
	ET2-b	12 {20} [70] (436)	all	100	9 {13} [26] (63)	75 {65} [37] (14)
	ET total	15 {33} [99] (566)	all	100	11 {19} [32] (73)	73 {58} [32] (13)

[a] The numbers reported for each construct and generation correspond to T_0 original lines, T_0 regenerants or T_1 families (curly brackets), T_1 progeny plants or T_2 families (square brackets), T_2 progeny plants (round brackets) respectively.
[b] The percentage is calculated on the lines, families or plants expected active.

On basis of the Southern analysis, a subset of 99 transpositionally active T_1 plants (originating from 15 T_0 lines and 33 T_1 families) were selected for further propagation in the T_2 generation. Unlike the previous generation, T_2 progeny seeds were directly sown in the greenhouse without selection for BAR resistance, so that plants lacking *Ds* were also included, together with plants lacking the *Ac*-transposase following segregation.

The results of the molecular analysis performed on the T_1 and T_2 lines are reported in Table 2. The occurrence of newly transposed *Ds* in a plant was chosen as a criterion to indicate transpositional activity. Therefore, the number of plants with newly transposed *Ds* is reported as the number of active plants. In general, lines that were inactive or showed only moderate activity in the T_0 generation (FDS with faint GUS-hybridising bands) remained inactive or partially active also in T_1, independently from the construct used (data not shown).

On the other hand, about 92% of the lines considered active in T_0 (EDS or FDS+EDS) retained activity in T_1. Figures 3 and 4 illustrate the transposition behaviour of representative active ET2 lines over successive generations, analysed by Southern hybridisation with a *Ds* specific probe (*GUS*). In general, the insertions present in the T_0 parents were transmitted to the T_1 progenies, indicating their germinal nature (Fig. 3 and 4). Active transposition was evident in all the different T_1 progeny families, visualised as presence of several additional independent *Ds* insertions within and among families. Plants showing a discrete number of bands of higher intensity, usually corresponded to plants in which the *Ac* transposase was lost by genetic segregation yielding stable *Ds* insertions and were not propagated further (i.e. plants 348-C-4, 374-D-1 and 7 in Fig. 4). In total, 83% of the T_1 families (53% of all the T_1 plants) analysed showed new transposed *Ds* inserts.

However, the transposition rate in T_2 generation seemed to decrease considerably. Only about 32% of the T_2 families (58% of the original T_1 families or 13% of the T_2 progeny plants) that were propagated from the most active T_1 plants still displayed new transpositions (i.e. Fig. 3, line 348 and 374). Surprisingly, in the lines exhibiting inhibition of transposition, only the original T_0 insertions seemed to be stably inherited among the many inserts from the T_1 parents, as exemplified by the T_2 progenies of line 373 (Fig. 3) or 13.7 G, H, I and L (Fig. 4 - T_2 panel). This would imply that many of the *Ds* bands present in the T_1 parents represented somatic transposition events that were not inherited in the T_2 progeny. As a consequence, the T_2 progenies originating from the same T_1 family yielded T_2 plants harbouring mainly the same insertions, while propagation of several T_0 regenerants with independent transpositions resulted in the recovery of a higher number of T_2 families with different fixed insertions (Fig. 4).

The same behaviour was observed for a small set of T_3 progeny plants propagated from line 13.11 B (Fig. 4 - T_3 panel). Overall, stabilisation of *Ds* inserts occurred in the T_2 for about

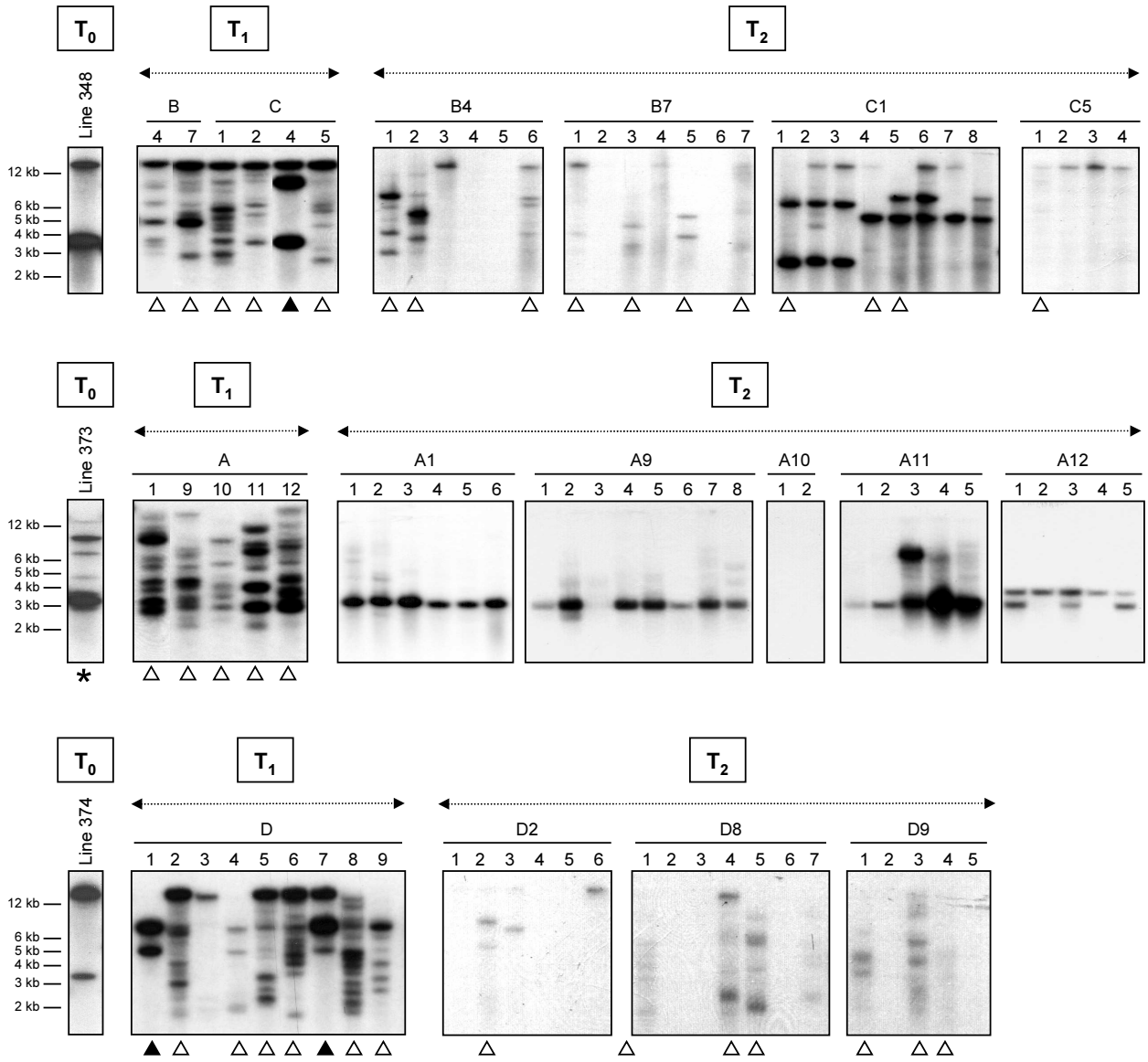


Figure 3. *Ds* transposition behaviour in ET2 lines over three generations. Example of Southern analysis with a *GUS* probe of representative T_1 and T_2 progeny plants from three single copy ET2-b lines described in Fig. 2 (the T_0 pattern is also reported as a comparison). Only line 373 still displayed FDS in T_0, while full excision was detected in the next generation, as confirmed by hybridisation with a *GFP* probe (data not shown). Lines 348 and 374 revealed active independent transposition still occurring in T_2. On the other hand, line 373 represented a clear demonstration of loss of *Ds* mobility in T_2 and stabilisation of T_0 parental insertions. All the three lines displayed amplification of *Ds* copy number. Plants with new transposition events included in the calculations of frequencies are marked by an open arrowhead. Stable T_1 plants lacking the transposase are marked with a filled arrowhead. Plants with FDS are marked with an asterisk.

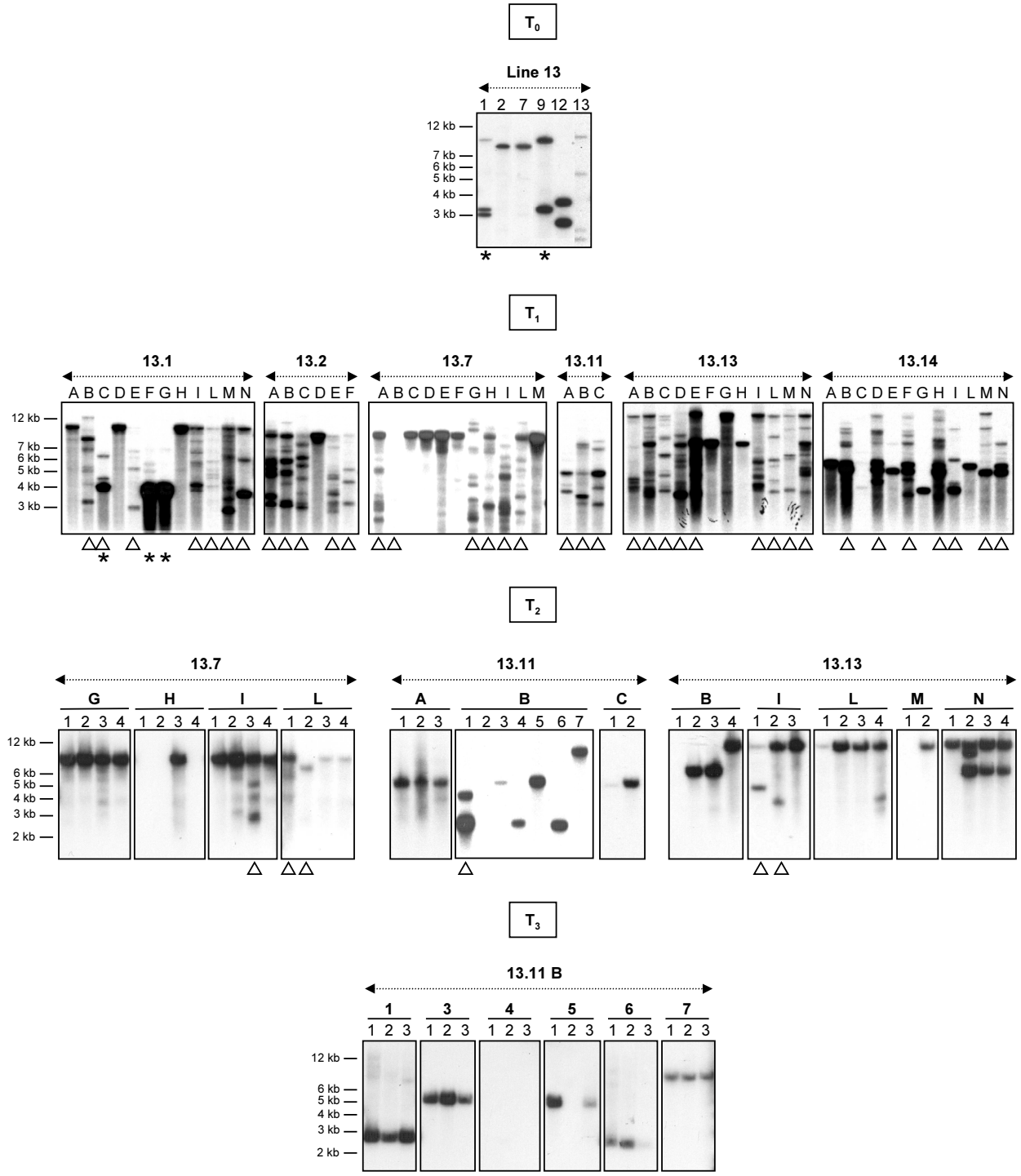


Figure 4. *Ds* transposition behaviour in one ET2 line over four generations. Example of Southern analysis with a GUS probe performed on T_0 regenerants and T_1 to T_3 progeny plants from one ET2-a line, containing 3 copies of the original T-DNA. Several hybridising fragment corresponding to transposed *Ds* are revealed, with fainter bands indicating somatic transpositions. Stabilisation of independent insertions in different T_2 families is shown (i.e. 13.7, 13.11, 13.13). In case of later inactivation, independent stabilised insertions could be revealed within the same T_3 family (13.11 B). Plants still displaying FDS are present in T_0 and T_1 and are marked by an asterisk. An open arrowhead marks the plants with new transposition events included in the calculations of frequencies. Subsets of plants were selected in each generation for further propagation, according to transpositional activity and availability of greenhouse space, and representative samples are shown.

40% of the T_1 families analysed, which could be used for inheritance and segregation in later generations.

In several lines amplification of the number of copies of the *Ds* transposon was observed, as exemplified by the three single-copy lines described in Fig. 3. On the other hand, inhibition of activity of the *Ds* transposon in later generations sometimes led to fixation of a lower number of *Ds* insertions, even in lines with initial higher copy number (Fig. 4).

Enhancer trapping – GUS staining

In order to test the ability of the ET lines to generate GUS expression patterns, a set of active T_1 lines were stained with X-gluc. A total of 536 T_1 plants were tested for GUS staining patterns in roots and leaves. Only one positive plant was identified displaying somatic GUS expression in the leaves. On the other hand, GUS assays on young inflorescences and mature flowers revealed rachilla, stamen and carpel specific staining in three plants out of 146 (2%) analysed so far (Fig. 5).

Genomic sequences flanking the transposon were isolated by TAIL PCR in order to identify the putative genes generating the GUS expression patterns identified. From one to several flanking sequences were isolated in each of the three plants, but no significant homologies were found after comparison to known sequences in public databases.

Inheritance of the GUS expression patterns identified was also checked, but unfortunately none of the *Ds* insertions responsible for the staining was transmitted to the progeny.

Figure 5. GUS staining pattern in three T_1 enhancer trap lines with construct ET2. GUS assays were performed on flowers of T_1 progeny plants originating from line 13 (Fig. 4). GUS expression in line 13.4L was restricted to the rachilla of developing flowers (A). Anthers and pollen specific expression was detected in lines 13.9D (B, C) and 13.14A (D), respectively. Some flowers in line 13.9 D showed also lodicules and carpel specific staining (E).

Reverse genetics

The availability of the rice genome sequence promotes the development of reverse genetics strategies to find inserts in genes discovered by sequencing. To assess the suitability of the *Ac*/*Ds* transposon system to tag genes, genomic sequences flanking the *Ds* transposons (Insertion Tagged Sites, ITS) were isolated for three subsets of T_1 lines with active somatic transposition or with stable insertions, due to loss of the *Ac* transposase by genetic segregation. The pattern of *Ds* transposition analysed by Southern hybridisation of some of the selected plants can be visualised in Figure 3 and 4. A summary of the significant homologies obtained after comparison of the flanking sequences obtained to known genes or proteins in public databases is shown in Table 3.

Table 3. Summary of Blast analyses of *Ds* flanking sequences (ITS) from T_1 ET2 plants.

	T_1 lines		
	ET2-a	ET2-b	ET2-b
Total number of plants[a]	17 (1)	8 (4)	14 (9)
Transpositional status	active	active	stable
Total number of ITS isolated	67	32	18
Average number of inserts / plant	3.9	4	1.3
ITS with homology to rice sequences[b]	31	14	12
Total number of rice chromosomes with ITS	8	5	4
ITS with homology to transcribed rice sequences[c]	16 (9)	10 (9)	7 (6)
Frequency of ITS in transcribed regions[d]	24% (52%)	31% (71%)	39% (58%)

[a] Within brackets the number of original T_0 lines.
[b] The homologies reported refer to Blast searches against the "nt" or "nr" databases at NCBI. On the other hand, all the ITS gave homology to rice contigs when blasted against the rice database at the Beijing Genomics Institute.
[c] The numbers refer to insertions in ORF and 5' regulatory sequence, up to 0.5 kb upstream the ATG. Within brackets the number of ITS in ORF.
[d] The percentage is calculated on the number of total ITS or on the number of ITS with homology to rice sequences (in brackets).

The ITSs were isolated using TAIL PCR and Adapter PCR (see Material and methods). In general, TAIL PCR proved to be efficient in case of plants with single or low copy number of the transposon, as for the set of stabilised T_1 plants. When multiple *Ds* insertions containing plants were employed, techniques involving adapter mediated PCR resulted in a higher percentage of recovery of good quality sequences (data not shown).

A first subset of 17 active T_1 plants, derived from one original T_0 ET2-a line (line 13; Fig. 4), was analysed by Adapter PCR. In total 67 unique ITSs were obtained, accounting for an average number of 3.9 *Ds* inserts per plant. A similarity search against public databases revealed that 46% of the flanking sites (31/67) showed significant nucleotide homology to rice genomic

sequences spread over 8 chromosomes. Among them, the percentage of insertions in transcribed regions amounted to 52% (16/31) or 24% of the overall flanking sites isolated (16/67).

A second set of 8 active T_1 plants (from 4 T_0 ET2-b lines) was analysed by TAIL PCR. Of 32 total ITSs isolated, 10 represented insertions in codogenic regions out of 14 showing significant similarity to rice genomic sequences over six chromosomes (listed in Table 4), indicating that 31% of the total *Ds* insertions identified were in transcribed genes. The number of *Ds* inserts in these plants was also averaging 4.

The last set of 14 stable T_1 plants (from 9 original ET2-b lines) analysed by TAIL PCR yielded 18 total flanking sites, with a decrease in the average number of *Ds* inserts per plant to 1.3. After comparison to public databases, about 39% (7/18) of the insertions could be assigned to transcribed genes. A list of the ITSs with significant homology to rice sequences (12/18) is presented in Table 4.

DISCUSSION

In an effort to build a transposon population for functional genomics studies in rice, more than 1000 transgenic lines containing *Ac*/*Ds* enhancer trap constructs were produced and analysed for activity. As problems in the continuous mobility of the *Ac*/*Ds* transposons in rice have been reported earlier, we assessed the behaviour of *Ds* over several generations to understand the different parameters involved. The potential use of the lines generated for gene tagging purposes was therefore evaluated by testing the transposition ability through three successive generations, the frequency of GUS staining patterns obtained and the nature of the insertion tagged sites generated.

Ds transposition behaviour

Excision occurred in about 80% of the intact T_0 lines analysed (data not shown), varying in time from very early at callus level or later during plant regeneration. Eighty percent of the excised *Ds* underwent reintegration at different genomic locations, accounting for about 62% of the T_0 plants displaying complete transposition. Early transpositions directly after transformation are likely attributable to the high levels of transposase induced by the CaMV 35S promoter driving the immobilised *Ac*. This phenomenon was initially reported in Arabidopsis (Scofield et al. 1992; Swinburne et al. 1992) and later described in rice (Chin et al. 1999; Izawa et al. 1997), in case of *Ds* excision promoted by either CaMV 35S-*Ac* transposase expressed from the cDNA or from the genomic clone. However, a higher frequency of early transpositions often resulted in a lower amount of independent insertions in Arabidopsis (Scofield et al. 1992; Swinburne et al.

Table 4. Genomic sequences flanking *Ds* insertions (ITS) in T_1 progeny plants of ET2-b lines.

Ds	Plant	ITS	Accession	Position	Chr	Annotation	Upstream/Downstream[a]
Active	102-A2	A01	AAK55456	(blastx)		Putative reverse transcriptase	
		D05	AP002745	132,900	1	Put. endo-beta-1,4-glucanase	
		H06	AP002522	40,628	1	Putative polyprotein	
	102-A4	E02	AL732376	6,904	12	Intergenic	Ups: Trans. factor OSMYB2
		G08	AC082645	119,500	3	Intergenic	
		H07	AP004364	54,690	1	Hypothetical protein	
		H11	AP003247	73,898	1	Hypothetical protein	
	288-A2	H02	AL662987	87,077	4	Put. DNA helicase (Sen1-like)	
	373-A11	A02	AC103891	40,698	3	Intergenic	Ups: Putative RING zinc finger protein / Dwn: Putative protein kinase
	373-A12	A08	AP005178	133,618	7	Putative protein kinase Xa21	
		B12	BAB61155	(blastx)		Putative DnaJ protein	
		C08	AAK54295	(blastx)		Putative reverse transcriptase	
	374-D6	A04	AL928749	2,138	12	Intergenic	Ups: Hypothetical protein
		H03	AAK27805	(blastx)		Putative reverse transcriptase	
Stable	102-A8	C04	AP003793	99,943	1	Putative H+-pyrophosphatase	
	209-C3	10_21	AP003754	40,874	7	Put. glucuronosyl transferase	
	274-A6	C01	AL606648	132,513	4	UDP-galactase-4-epimerase	
	274-B2	A02	AL606694	97,261	4	Intergenic	Dwn: Hypothetical protein
	274-C7	D01	AP003052	108,460	1	Putative phosphoenolpyruvate carboxylase	
		E08	AP003413	19,641	1	Intergenic	
	348-B1	B09	Q9FWA4	(blastx)	3	Probable glycosyltransferase	
	348-C4	7_18	AP002521	8,562	1	Arginyl-tRNA synthetase	
	374-C2	8_19	AL606455	49,301	4	Intergenic	
	374-D7	8_17			Repeat		
	387-C2	10_19			Repeat		
		13_22	BAB87820	(blastx)		P450 [Triticum aestivum]	

[a] In case of *Ds* insertions in intergenic regions, the nearest genes predicted within a distance of 5 kb upstream (ups) or downstream (dwn) are reported.

1992). In our population, a high frequency of transmission of the T_0 parental *Ds* insertions (germinal) in the T_1 generation was observed, in agreement with previous reports (Chin et al. 1999). In some lines, the presence of the same parental band in plants belonging to the same T_1 family confirmed the occurrence of early transpositions in the To generation. Nevertheless, the percentage of plants within those T_1 families having also independent transpositions was high, suggesting the occurrence of additional secondary transpositions in the germ cells of the T_0 regenerants. Moreover, active transposition was still detectable in T_1, as indicated by the presence of numerous somatic insertions.

In spite of the high mobility of *Ds* observed in the T_1 plants, only about 30% of the T_2 families analysed included siblings showing new transpositions, which are being multiplied further. Considering the original T_1 families, about 40% underwent complete inactivation in T_2. Surprisingly, most of the inactive T_2 plants stably inherited the original insertions present in the T_0 parent. Inhibition of transposition in later generations seemed to be associated with inactivation of the BAR selectable marker. When applying PPT selection on T_2 and T_3 seeds, the recovery of resistant plants was much lower than expected from the normal genetic segregation ratios (Enckevort and Pereira, unpublished results). In agreement with previous reports (Schwartz and Dennis 1986; Wang et al. 1996), this would suggest the occurrence of methylation of *Ds* as a cause of inhibition of transposition, though in this case methylation seems to spread over the whole transposon instead of being restricted to the terminal regions.

On the other hand, inactive T_2 and T_3 plants could be considered as a source of stabilised insertions to be used directly in reverse genetics experiments. The pattern of *Ds* transposition in the T_2 generation clearly indicated stabilisation of parental insertions in about 70% of the families analysed, with plants belonging to the same T_2 family having often the same *Ds* insert and different T_2 families having different *Ds* inserts. Given the prevalent inheritance of T_0 insertions, the probability to get T_2 progeny plants with different *Ds* inserts is then correlated to the amount of different T_0 primary regenerants, and therefore T_1 families, propagated. In order to obtain a collection with the maximum number of independent stabilised insertions, an ample number of lines in earlier generations (T_0 and T_1) are being propagated. T_2 and T_3 progenies will then be utilised as a source of stabilised lines.

Previous studies in maize demonstrated that silent *Ac* transposons could be reactivated following tissue culture (Brettell and Dennis 1991; Peschke et al. 1987). Accordingly, remobilisation of inactive *Ds* in rice was achieved in plants regenerated from cultured rice seeds containing inactive *Ds* and active *Ac* transposase (Izawa et al. 1997; Ki et al. 2002). Remobilisation of the *Ds* transposon appeared to be stage-specific, with a maximum of activity at the time of plantlets regeneration, and linked with de-methylation of its terminal regions (Ki et al. 2002). At the same time, an increased level of *Ac* transposase transcript was shown to occur in the regeneration stage with higher transpositional activity, suggesting a link between transposase activity and methylation of *Ds*. The addition of a regeneration step via tissue culture as a way to reactivate transposon activity in an heterologous system proved to be successful also with *Ac*/*Ds* in *Brassica oleracea* (Mckenzie et al. 2002) and could then represent a useful strategy to obtain high level of transposition for efficient insertional mutagenesis in rice.

Interestingly, the autonomous *Ac* element was shown to transpose efficiently in rice even in later generations (Enoki et al. 1999; Izawa et al. 1997; Greco et al. 2001b). However, once inactivated, the process appeared to be irreversible, in contrast with previous reports from maize

(Brettell and Dennis 1991). This could suggest a different sensitivity (response) of the *Ac*/*Ds* transposon system to silencing (methylation) and subsequent inhibition of activity, when the *Ac* transposase is expressed under the control of its own promoter. If this is true, perhaps a transposase source consisting of an immobilised *Ac* retaining its own promoter, next to a strong enhancer, would provide an adequate level of transposition and be less prone to inactivation. Induction of early excision and transposition by a strong enhancer adjacent to the endogenous *Ac* promoter was previously demonstrated for a two-component *Ac*/*Ds* system in Arabidopsis (Balcells and Coupland 1994). Similarly, the CaMV 35S enhancer next to an autonomous *Ac* element also influences its transposition behaviour in rice (Greco et al. 2001b). Constructs based on this principle are currently under preparation.

Among the active lines produced, about 10% displayed amplification of *Ds* copy number. This phenomenon was initially described for *Ac* transposition in maize (Greenblatt 1984) and is due to transposition during replication, possibly associated with high *Ac* transposition activity (Peterson and Yoder 1995). Later, proliferation of *Ac* was shown to occur also in heterologous hosts, like tomato (Yoder 1990) and rice (Greco et al. 2001b). Moreover, amplification of *Ds* copy number was reported after ratoon culture in rice (Chin et al. 1999). Here, restriction of *Ds* proliferation to T_0 and T_1 actively transposing plants could be also related to high levels of transposition promoted by the CaMV 35S promoter driving the *Ac* transposase. The propagation of such lines with multiple independent *Ds* insertions will help in reducing the number of lines needed to saturate the genome.

Some of the enhancer trap lines developed (construct ET1) bear *lox* sequences for site-specific recombination (reviewed in Ow and Medberry 1995; Ow 1996) in the *Ds* transposon and the T-DNA. Combining insertional mutagenesis with site-specific recombination provides additional advantages for mutant identification and gene cloning (van Haaren and Ow 1993). The general principle is that, once transposition has occurred, site-specific recombination can induce chromosomal rearrangements between transposed *Ds* and the T-DNA. Due to prevalent *Ac*/*Ds* transposition to linked loci, this will mainly results in the occurrence of intra-chromosomal deletions or inversions. Mapping the donor T-DNA in the lines used will then provide a way to induce rearrangements at specific locations. Such an approach has been applied in Arabidopsis, tobacco and tomato, using *Ac*/*Ds* and the Cre-*lox* (Medberry et al. 1995; Osborne et al. 1995; Stuurman et al. 1996) or the R-*RS* system (Machida et al. 1995). Using the latter system, Nakagawa et al. (2001) described the recovery of genomic deletions in rice. In case of early transposition with amplification of *Ds* copy number in a single cell, as it was observed for *Ac* (Greco et al. 2001b), the multiple transposed *Ds* at cis-linked loci will provide a substrate for recombination with the possibility to create sequential deletions. At the same time, the presence of sequences for site-specific recombination could allow Cre recombinase-mediated delivery of

foreign DNA in the rice genome, encompassing the limitations due to transformation process (reviewed in Ow 2002).

Enhancer trapping efficiency

Gene detection systems in rice have been previously described, which made use of T-DNA (Jeon et al. 2000) or transposons. In the latter case, *Ds* elements modified to serve as gene or enhancer traps have been employed (Chin et al. 1999; Upadhyaya et al. 2002) and the reported frequencies of GUS staining patterns identified were comparable to Arabidopsis (Sundaresan et al. 1995). In our system, the frequency of lines showing GUS expression was surprisingly low. Perhaps too few independent insertions were represented in the subset of T_1 plants analysed, while the high rate of somatic transpositions could have hampered the detection, especially in case of staining restricted to small sectors of the plant. All the three GUS-positive patterns identified failed to be inherited, strongly suggesting the occurrence of *Ds* insertions in sectors of the inflorescences used for screening, which were not transmitted to the next generation. Indeed, several flowers from the same plant were stained, but GUS expression was identified only in some of them. Alternatively, the presence of a *lox* site between the minimal promoter and the GUS gene might also be reducing the expression of the GUS marker. Improved protocols to increase the sensitivity of staining detection are currently being tested on T_2 stable lines.

Gene insertional specificity

A pilot sequencing experiment of Insertion Tagged Sites (ITSs) was performed on three subsets of T_1 lines, active or stabilised by genetic segregation of the transposase.

Plants with active transposition showed a higher number of Ds insertions in comparison to stabilised plants, in accordance with the high somatic activity revealed in the former by Southern analysis. As previously discussed, most of those inserts will likely not be transmitted to the next generation. Nevertheless, the different ITSs obtained clearly indicate that many independent insertions can be potentially recovered after stabilisation, even from plants originating from a single T_0 line, as in case of line 13.

The percentage of insertions in transcribed regions of the rice genome is about 30% of all the ITSs, but rises to more than 50% on considering only the ITSs that give homology to rice sequences, indicating that more than half of the insertions are actually in genes. A compilation of the tagged genes shown in Table 4 points out that more inserts are in genes than upstream/downstream, demonstrating again the insertional specificity. This preference for Ds insertion in transcribed genes confirm previous reports about the behaviour of Ac (Enoki et al. 1999; Greco et al. 2001b) and implies that a lower number of lines will be needed to saturate the genome for insertions.

Furthermore, as the lines contain enhancer trap constructs, they are supposed to reveal also neighbouring genes that reside within 5 kb upstream or downstream the insertion, the range at which strong enhancers in the genome can work. Examination of the genes flanking the insertion tagged sites (Table 4) reveal a number of interesting genes like transcription factors, whose expression patterns can be monitored using a reverse genetics approach.

Based on those results, extensive sequences of flanking sites is now being carried out on stabilized T_2 lines that underwent inactivation of transposition.

Selectable markers

The effectiveness of a high-throughput transposon tagging system for gene function identification relies on the choice of efficient selectable markers. A critical step is the selection of lines carrying independent stable insertions. Double assays based on the use of a counter-selectable marker linked to the transposase source and a positive selection marker inside the mobile transposon were devised and successfully employed in Arabidopsis, to select for independent unlinked stable insertions (Fedoroff and Smith 1993; Sundaresan et al. 1995). Particularly efficient turned out to be the use of the *BAR* and *SU1* genes as positive and negative selection markers in greenhouse-based screenings in Arabidopsis (Tissier et al. 1999; Marsch et al. 2002). To date, only the *BAR* gene was proven to work efficiently in rice as transposon marker (Chin et al. 1999; Greco et al. 2001b), while the effectiveness of *SU1* is still controversial. The functionality of *SU1* in monocot species was previously reported by Koprek et al. (1999), which demonstrated its potential applicability in barley for large-scale greenhouse screenings. In rice, Chin et al. (1999) described the use of this marker in their original T-DNA, but did not report about its efficiency. Preliminary tests performed on seedlings of some of our rice enhancer trap lines grown in the greenhouse and sprayed with the R7402 pro-herbicide did not result in an effective selection. However, further testing using different experimental conditions and larger samples are required for a more proper evaluation.

On the other hand, sGFP seems to offer a promising alternative (Upadhyaya et al. 2000; our unpublished results). Although its suitability as excision marker could be debatable (Greco et al. 2001b; this work), the possibility of easy detection at seed level as a segregation marker could allow effective visual selection for independent insertions. Transposon constructs using the *sGFP* gene as negative selection marker are currently being tested. Alternatively, hygromycin painting on greenhouse leaves could represent another effective way to screen for transposase presence and segregation (Cotsaftis et al. 2002). Effective selection for plants with stable *Ds* insertions could be achieved using a leaf painting assay for Basta and hygromycin separately on greenhouse seedlings. The Basta resistant/hygromycin sensitive plants can then be selected as stable lines.

Conclusions

International public efforts to build a rice gene machine require a concerted effort from a number of different labs providing genotypes and information available to the public. The use of a number of different systems like T-DNA, retrotransposons and cut-and-paste type of transposons such as the *Ac*/*Ds* system will ensure that biases due to insertional specificity will be avoided.

The important consistent observation derived from this study is that the *Ds* transposons, like the autonomous *Ac* (Enoki et al. 1999; Geco et al. 2001b), insert at a higher frequency in gene rich regions and tag genes. This suggests that a much smaller population of inserts would be required to be able to tag most of the genes in the genome. In addition, our choice of using multiple transposon lines allows us to be able to generate a large population of transposon inserts in a smaller number of plants. Following the identification of an insert in a gene of interest, segregation analysis has normally to be carried out and in this effort other inserts can be segregated out.

The results reported here help in the development of more efficient *Ac*/*Ds* transposon populations using different markers and insertion types. Parallel efforts are thus being made to produce a rice activation tagging population. In this case, the *Ds* mobile transposon is modified to carry strong enhancer sequences that could cause dominant mutant phenotypes by misexpressing adjacent genes. This system has been successfully employed for gene function discovery in Arabidopsis using T-DNA (i.e. Weigel et al. 2000) and recently described in rice (Jeong et al. 2002). However, its effectiveness in terms of mutant phenotype recovery in Arabidopsis was demonstrated to be even higher when transposons were used as vectors (Marsch et al.2002; Wilson et al. 1996), offering a powerful strategy in rice to unravel novel gene functions that would be otherwise not possible due to gene redundancy.

Acknowledgements

We would like to thank Gert van Arkel and Wim Dirkse for assistance in the isolation of flanking sequences; Hans Sandbrink and Joost de Groot for assistance in Blast analyses and Dick Geurtjes for taking care of the plants in the greenhouse. This research was funded in the participating labs by a European Union project BIO4 CT 972132 on “Transposon mutagenesis in rice”. The work has been carried out in compliance with the current laws governing genetic experimentation in The Netherlands.

LITERATURE CITED

Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402

Arumuganathan K, Earle E (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208-218

Baker B, Coupland G, Fedoroff N, Starlinger P, Schell J (1987) Phenotypic assay for excision of the maize controlling element *Ac* in tobacco. EMBO J 6: 1547-1554.

Balcells L, Coupland G (1994) The presence of enhancers adjacent to the *Ac* promoter increases the abundance of transposase mRNA and alters the timing of *Ds* excision in Arabidopsis. Plant Mol Biol 24: 789-798

Barry G (2001) The use of the Monsanto draft rice genome sequence in research. Plant Physiol 125: 1164-1165

Bohnert H, Ayoubi P, Borchert C Bressan R, Burnap R, Cushman J, Cushman M, Deyholos M, Fischer R, Galbraith D et al. (2001) A genomics approach towards salt stress tolerance. Plant Physiol Biochem 39: 295-311

Brettell RI, Dennis ES (1991) Reactivation of a silent *Ac* following tissue culture is associated with heritable alterations in its methylation pattern. Mol Gen Genet 229: 365-372

Buell CR (2002) Current status of the sequence of the rice genome and prospects for finishing the first monocot genome. Plant Physiol 130: 1585-1586

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B et al. (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14: 537-545

Chin HG, Choe MS, Lee SH, Park SH, Koo JC, Kim NY, Lee JJ, Oh BG, Yi GH, Kim SC, Choi HC, Cho MJ, Han CD (1999) Molecular analysis of rice plants harboring an *Ac*/*Ds* transposable element-mediated gene trapping system. Plant J 19: 615-623

Chiu W, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J (1996) Engineered GFP as a vital reporter in plants. Curr Biol 6: 325-330

Christensen AH, Quail PH (1996) Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res 5: 213-218

Christensen AH, Sharrock RA, Quail PH (1992) Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Mol Biol 18: 675-689

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

De Block M, Botterman J, Vanderwiele M, Dockx J, Thoen C, Gossele V, Movva R, Thompson C, Van Montagu M, Leemans J (1987) Engineering herbicide resistance in plants by expression of a detoxifying enzyme. EMBO J 6: 2513-2518

Enoki H, Izawa T, Kawahara M, Komatsu M, Koh S, Kyozuka J, Shimamoto K (1999) *Ac* as a tool for the functional genomics of rice. Plant J 19: 605-613

Fedoroff NV, Smith DL (1993) A versatile system for detecting transposition in Arabidopsis. Plant J 3: 273-289

Feng Q, Zhang Y, Hao P, Wang S, Fu G et al. (2002) Sequence and analysis of rice chromosome 4. Nature 420: 316-320

Gale M, Devos K (1998) Comparative genetics in the grasses. Proc Natl Acad Sci 95: 1971-1974

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

Greco R, Ouwerkerk Pieter BF, Sallaud C, Kohli A, Colombo L, Puigdomenech P, Guiderdoni E, Christou P, Hoge JHC, Pereira A (2001a) Transposon insertional mutagenesis in rice. Plant Physiol 125: 1175-1177

Greco R, Ouwerkerk Pieter BF, Taal Anke JC, Favalli C, Beguiristain T, Puigdomenech P, Colombo L, Hoge JHC, Pereira A (2001b) Early and multiple *Ac* transpositions in rice suitable for efficient insertional mutagenesis. Plant Mol Biol 46: 215-227

Greenblatt IM (1984) A chromosome replication pattern deduced from pericarp phenotypes resulting from movements of the transposable element Modulator in maize. Genetics 108: 471-485

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148: 479-494

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

Hoess R, Ziese M, Sternberg N (1982) *P1* site-specific recombination: nucleotide sequence of the recombining sites. Proc Natl Acad Sci 79: 3398-3402

Izawa T, Ohnishi T, Nakano T, Ishida N, Enoki H, Hashimoto H, Itoh K, Terada R, Wu C, Miiyazaki C, Endo T, Iida S, Shimamoto K (1997) Transposon tagging in rice. Plant Mol Biol 35: 219-229

Izawa T, Shimamoto K (1996) Becoming a model plant: The importance of rice to plant science. Trends Plant Sci 1: 95-99

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

Jeon JS, An G (2001) Gene tagging in rice: A high throughput system for functional genomics. Plant Sci 161: 211-219

Jeon JS, Lee S, Jung KH, Jun SH, Jeong DH, Lee J, Kim C, Jang S, Lee S, Yang K, Nam J, An K, Han MJ, Sung RJ, Choi HS, Yu JH, Choi JH, Cho SY, Cha SS, Kim SI, An G (2000) T-DNA insertional mutagenesis for functional genomics in rice. Plant J 22: 561-570.

Jeong D, An S, Kang H, Moon S, Han J, Park S, Lee H, An K, G A (2002) T-DNA Insertional Mutagenesis for Activation Tagging in Rice. Plant Physiol 130: 1636-1644

Ki C, Je B, Piao H, Par S, Kim M, Park S, Park J, Park S, Lee E, Chon N, Won Y, Lee G, Nam M, Yun D, Lee M, Cha Y, Le K, Eun M, Han C (2002) Reprogramming of the activity of the activator/dissociation transposon family during plant regeneration in rice. Mol Cells 14: 231-237

Kohli A, Xiong J, Greco R, Christou P, Pereira A (2001) Tagged Transcriptome Display (TTD) in indica rice using *Ac* transposition. Mol Gen Genom 266: 1-11

Koprek T, McElroy D, Louwerse J, Williams Carrier R, Lemaux PG (1999) Negative selection systems for transgenic barley (*Hordeum vulgare* L.): comparison of bacterial *codA* and cytochrome *P450* gene-mediated selection. Plant J 19: 719-726

Liu YG, Whittier RF (1995c) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25: 674-681

Machida Y, Onouchi H, Torikai S, Kudo M, Koizumi J, Machida C (1995) Use of site-specific recombination systems for isolation of mutants of *Arabidopsis thaliana*: the tagging/deletion strategy. In: Modification of gene expression and non-mendelian inheritance. Proceedings of the US-Japanese Joint Meeting. Edited by Oono K, Takaiwa F. Mito-Sato Press: 199-223

Marsch Martinez N, Greco R, Van Arkel G, Herrera Estrella L, Pereira A (2002) Activation tagging using the *En-I* maize transposon system in Arabidopsis. Plant Physiol 129: 1544-1556

Martienssen R, Colot V (2001) DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science 293: 1070-1074

McClintock B (1959) Genetic and cytological studies of maize. Carnegie Inst Wash Year Book 58: 452-456

Mckenzie N, Wen L, Dale P (2002) Tissue-culture enhanced transposition of the maize transposable element *Dissociation* in *Brassica oleracea* var. '*Italica*'. Theor Appl Genet 105: 23-33

Medberry S, Dale E, Qin M, Ow D (1995) Intra-chromosomal rearrangements generated by Cre-*lox* site-specific recombination. Nucleic Acids Res 23: 485-490

Nakagawa Y, Machida C, Machida Y, Toriyama K (2000) Frequency and pattern of transposition of the maize transposable element *Ds* in transgenic rice plants. Plant Cell Physiol 41: 733-742.

Nakagawa Y, Machida C, Machida Y, Toriyama K (2001) A system to induce the deletion of genomic sequences using R/*RS* site-specific recombination and the *Ac* transposon in transgenic rice plants. Theor Appl Genet 102: 1136-1141

O'Keefe D, Tepperman J, Dean C, Leto K, Erbes D, Odell J (1994) Plant expression of a bacterial cytochrome p450 that catalyzes activation of a sulfonylurea pro-herbicide. Plant Physiol 105: 473-482

Osborne B, Wirtz U, Baker B (1995) A system for insertional mutagenesis and chromosomal rearrangement using the *Ds* transposon and Cre-*lox*. Plant J 7: 687-701

Ow D (1996) Recombinase-directed chromosome engineering in plants. Curr Opin Biotechnol 7: 181-186

Ow D (2002) Recombinase-directed plant transformation for the post-genomic era. Plant Mol Biol 48: 183-200

Ow D, Medberry S (1995) Genome manipulation through site-specific recombination. Crit Rev Plant Sci 5: 521-527

Pereira A (2000) A transgenic perspective on plant functional genomics. Transgenic Res 9: 245-260

Pereira A, Aarts MGM (1998) Transposon tagging with the *En-I* system. In: Martinez-Zapater J, Salinas J (eds) Arabidopsis protocols. Humana Press, Totowa, NY, pp 329-338

Peschke V, Phillips R, Gengenbach B (1987) Discovery of transposable element activity among progeny of tissue culture-derived maize plants. Science 238: 804-807

Peterson PW, Yoder JI (1995) Amplification of *Ac* in tomato is correlated with high *Ac* transposition activity. Genome 38: 265-276

Pietrzak M, Shillito RD, Hohn T, Potrykus I (1986) Expression in plants of two bacterial antibiotic resistance genes after protoplast transformation with a new plant expression vector. Nucleic Acids Res 14: 5857-5868

Reddy A, Ramakrishna W, Sekhar A, Ithal N, Babu P, Bonaldo M, Soares M, Bennetzen J (2002) Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. ssp. *indica* cv. Nagina 22). Genome 45: 204-211

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T et al. (2002) The genome sequence and structure of rice chromosome 1. Nature 420: 312-316

Scarpella E, Rueb S, Boot K, Hoge J, Meijer A (2000) A role for the rice homeobox gene *Oshox1* in provascular cell fate commitment. Development 127: 3655-3669

Schwartz D, Dennis E (1986) Transposase activity of the *Ac* controlling element in maize is regulated by its degree of methylation. Mol Gen Genet 205: 476-482

Scofield SR, Harrison K, Nurrish SJ, Jones JD (1992) Promoter fusions to the *Activator* transposase gene cause distinct patterns of *Dissociation* excision in tobacco cotyledons. Plant Cell 4: 573-582

Sijmons PC, Dekker BM, Schrammeijer B, Verwoerd TC, van den Elzen PJ, Hoekema A (1990) Production of correctly processed human serum albumin in transgenic plants. Biotechnology 8: 217-221

Springer P (2000) Gene traps: Tools for plant development and genomics. Plant Cell 12: 1007-1020.

Stuurman J, de Vroomen M, Nijkamp H, van Haaren M (1996) Single-site manipulation of tomato chromosomes in vitro and in vivo using Cre-*lox* site-specific recombination. Plant Mol Biol 32: 901-913

Sundaresan V, Springer P, Volpe T, Haward S, Jones JD, Dean C, Ma H, Martienssen R (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev 9: 1797-1810

Swinburne J, Balcells L, Scofield SR, Jones JD, Coupland G (1992) Elevated levels of *Activator* transposase mRNA are associated with high frequencies of *Dissociation* excision in Arabidopsis. Plant Cell 4: 583-595

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JD (1999) Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. Plant Cell 11: 1841-1852

Tsugeki R, Kochieva EZ, Fedoroff NV (1996) A transposon insertion in the Arabidopsis *SSR16* gene causes an embryo-defective lethal mutation. Plant J 10: 479-489

Upadhyaya NM, Zhou XR, Zhu QH, Ramm K, Wu L, Eamens A, Sivakumar R, Kato T, Yun DW, Santhoshkumar C, Narayanan KK, Peacock JW, Dennis ES (2002) An i*Ac*/*Ds* gene and enhancer trapping system for insertional mutagenesis in rice. Funct Plant Biol 29: 547-559

van Haaren M, Ow D (1993) Prospects of applying a combination of DNA transposition and site-specific recombination in plants: a strategy for gene identification and cloning. Plant Mol Biol 23: 525-533

Wang L, Heinlein M, Kunze R (1996) Methylation pattern of *Activator* transposase binding sites in maize endosperm. Plant Cell 8: 747-758

Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, Nguyen JT, Sato S, Wang ZY, Xia Y, Dixon RA, Harrison MJ, Lamb CJ, Yanofsky MF, Chory J (2000) Activation tagging in Arabidopsis. Plant Physiol 122: 1003-1013

Wilson K, Long D, Swinburne J, Coupland G (1996) A *Dissociation* insertion causes a semidominant mutation that increases expression of *TINY*, an Arabidopsis gene related to *APETALA2*. Plant Cell 8: 659-671

Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, Fujii F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano M, Matsumoto T, Sasaki T (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell. 14: 525-535

Yamamoto K, Sasaki T (1997) Large-scale EST sequencing in rice. Plant Mol Biol 35: 135-144

Yoder JI (1990) Rapid proliferation of the maize transposable element *Activator* in transgenic tomato. Plant Cell 2: 723-730

Yu J, Hu S, Wang J, Wong GK, Li S et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296: 79-92

Chapter 7

Transcription and transposition of the maize *En/Spm* transposon system in rice.

Raffaella Greco[1], Pieter B.F. Ouwerkerk[2], Anke J.C. Taal[2], Christophe Sallaud[3], Emmanuel Guiderdoni[3], Annemarie H. Meijer[2], J. Harry C. Hoge[2†] and Andy Pereira[1]

[1] *Plant Research International, PO Box 16, 6700 AA, Wageningen, The Netherlands*
[2] *Institute of Biology, Leiden University, Leiden, The Netherlands*
[3] *CIRAD-BIOTROP, Montpellier, France*

(Submitted for publication)

ABSTRACT

Transposition of the maize *En/Spm* system in rice was investigated using a two-component construct containing a CaMV 35S-immobilised transposase and a modified *I/dSpm* transposon. Mobilization of *I/dSpm* in somatic sectors was demonstrated by sequencing of excision products and isolation of flanking genomic sequences in T_0 and T_1 progeny plants. Since the transposition efficiency appeared considerably lower compared to maize or other heterologous systems like Arabidopsis, we examined *En/Spm* transcription and splicing in the transgenic rice plants. Northern blot analysis revealed the presence of transcripts encoding the active TnpA and TnpD transposases with a higher amount of the latter, in reverse proportion to that observed in maize and Arabidopsis. RT-PCR analysis confirmed the occurrence of correct splicing and the formation of the two other alternatively spliced transcripts (TnpB and TnpC), as previously described for maize. Two alternative splice donor sites at the end of exon 1 were identified in maize at position 578 and 704. We observe that rice is similar to maize, as TnpA is preferentially spliced at position 578. We also show that Arabidopsis prefers splicing at position 704, similar to other dicots like tobacco. These observations indicate differences in splicing of the maize *En/Spm* element between dicot and monocot hosts. Nevertheless, the ratio at which the transcripts for the active transposases are produced seems to determine the efficiency of transposition, independently from the host. As a result, a limiting amount of TnpA might be responsible for the lower transposition activity of *En/Spm* in rice. Alternatively, reduced mobility of the modified *I/dSpm* element used may have resulted from the absence of critical sequences necessary for transposition. The influence of endogenous autonomous *En/Spm* related elements present in the rice genome on the transposition behaviour of the exogenous maize element was also considered.

Keywords: *En/Spm, transcription, transposition, rice, Arabidopsis*

This report is dedicated to the loving memory of our colleague Dr J. Harry C. Hoge.

INTRODUCTION

Rice has become a model system for the monocot species (Izawa and Shimamoto 1996), due to its compact genome, detailed molecular genetic maps and relative ease to transform. The whole genome sequence of rice is now available as draft shotgun sequences (Barry 2001; Goff et al. 2002; Yu et al. 2002), as well as systematic sequencing of chromosomes (Feng et al. 2002; Sasaki et al. 2002). The genome sequencing efforts reveal between 30-55,000 genes, whose functions still have to be elucidated.

Large scale insertional mutagenesis approaches can provide a significant contribution to the functional analysis of the discovered rice genes, as it has been amply demonstrated in Arabidopsis (reviewed in Parinov and Sundaresan 2000). In rice, insertional mutagenesis with T-DNA (Jeon et al. 2000; Jeong et al. 2002), endogenous retrotransposons (Hirochika 2001) and heterologous transposons like *Ac* and *Ac*/*Ds* (Izawa et al. 1997; Chin et al. 1999; Enoki et al. 1999; Nakagawa et al. 2000; Greco et al. 2001b; Kohli et al. 2001; Upadhyaya et al. 2002) have been employed. Both T-DNA and endogenous retrotransposons employ a tissue-culture regeneration phase that yield a high proportion of untagged mutations and complicates forward mutant analysis. On the other hand, the maize *Ac*/*Ds* transposon system has been shown to be active in rice for a few generations after transformation, but tends to be inactivated in later generations (Izawa et al. 1997). Alternative strategies are therefore still needed to broaden the effectiveness of insertional mutagenesis in rice.

The maize *En/Spm* transposon system seems promising in this respect. Although its use has not yet been reported in rice, it has been demonstrated to function in heterologous hosts like tobacco (Masson and Fedoroff, 1989; Pereira and Saedler, 1989), potato (Frey et al. 1989) and Arabidopsis (Cardon et al. 1993b). One of the main advantages in using the *En/Spm* system is the lower frequency of linked transposition if compared to *Ac*/*Ds* (Aarts et al. 1995), which could ease the recovery of stable independent insertions. In Arabidopsis, where the system was shown to transpose at higher efficiency, *En*/*Spm* mutagenised populations have been developed and successfully employed for gene tagging (Aarts et al. 1995; Wisman et al. 1998; Speulman et al. 1999; Tissier et al. 1999; Marsch-Martinez et al. 2002).

In general, the use of heterologous transposons allows one to modify the mobile element to carry markers for positive selection or gene detection (reviewed in Pereira 2000). In this latter case, a cassette containing a reporter gene like GUS or luciferase, preceded by splicing acceptor sites (gene and exon traps) or a minimal promoter (enhancer traps), is placed next to one of the borders of the transposon. Upon splicing and formation of a chimeric transcript or transcriptional activation by neighbouring regulatory sequences, expression of the reporter gene will eventually reflect the pattern of expression of the endogenous gene. In case of activation tagging, the mobile

element is carrying a strong enhancer that will provide mis-regulation of adjacent genes. For maximum efficiency, a minimal distance between the gene detection cassette and the surrounding genomic sequence would be required. This implies the possibility of using minimal transposon borders that can however still promote efficient transposition. Gene detection has been applied effectively using the *Ac*/*Ds* system in Arabidopsis (Fedoroff and Smith 1993; Springer et al. 1995; Sundaresan et al. 1995; Tsugeki et al. 1995; Wilson et al. 1996; Martienssen 1998) and rice (Chin et al. 1999; Greco et al. 2001a; Upadhyaya et al. 2002), but only to a very limited extent with the *En*/*Spm* system (Marsch-Martinez et al. 2002).

In this work, we present the first report on the transposition behaviour of the maize *En*/*Spm* system in rice. From the analysis of transgenic rice lines containing an *En*/*Spm* enhancer trap two-component system, we demonstrate excision and reinsertion of the *I*/*dSpm* transposon. As only somatic events could be revealed, indicating a low efficiency of transposition, we analyse the alternatively spliced transcription products of the *En/Spm* element in rice in comparison to Arabidopsis and discuss the results found earlier in maize and other plants.

MATERIAL AND METHODS

Vector construction

The immobilised *En* transposase driven by the CaMV 35S promoter was obtained from Aarts et al. (1995). The left (*I*-Lj) and right (*I*-Rj) ends of the mobile *I/dSpm* element in construct *En*-ET1 were derived from *En*-1, originally cloned from the *waxy* locus of maize (Pereira et al. 1985; GenBank Accession M25427). The *I*-Lj extends from position 1 up to the *Sal*I site at position 267, while the *I*-Rj extends from the *Nla*III site at position 7647 until the end of the element at position 8287.

The two constructs used were assembled in the binary vector pMOG22 (Zeneca-MOGEN, Netherlands) by multi-point ligation of the fragments subsequently described. *En*-ET1: 1) a *Xho*I-*Not*I fragment containing the immobilised CaMV 35S-*En* transposase; 2) a *Not*I-*Hind*III linker fragment, containing a CaMV 35S terminator and a *lox* site (Hoess et al. 1982); 3) a *Hind*III-*Spe*I fragment containing the *I*-Rj fused to an intron-less Ubiquitin promoter-*BAR*-Nos terminator cassette originally derived from pAHC25 (Christensen and Quail 1996); 4) a *Xba*I-*Bam*HI fragment containing a 61 bp minimal CaMV 35S promoter, a 48 bp oligonucleotide comprising a *lox* site, and a promoter-less *GUS* reporter gene-*Nos* terminator (Jefferson et al. 1987) fused to the *I*-Lj; 5) a *Bam*HI-*Eco*RI containing the *SU1* gene, with petunia Rubisco small subunit *SSU301* promoter and terminator, derived from pSSU-SU12 (O'Keefe 1994). *En*-TPase: 1) an

*Eco*RI/*Bam*HI fragment containing the *SU1* gene; 2, 3) a *Bam*HI/*Avr*I and an *Avr*I/*Hind*III fragments comprising the immobilized CaMV 35S-*En* transposase.

Plant transformation and growth conditions

Agrobacterium-mediated transformation of *Oryza sativa* ssp. *japonica* cv. Nipponbare, plant regeneration and growth were performed as previously described (Greco et al. 2001b). The *Agrobacterium tumefaciens* strains EHA105 or LBA4404 were employed.

For growth in tissue culture, progeny seeds were dehusked, surface-sterilized (2 min ethanol 100%, 1 h NaOCl 5% and Triton X-100 0.1%, six times rinsing with sterile H_20) and sown on Murashige-Skoog medium supplemented with 1% sucrose, 0.8% purified agar and 25 mg/l hygromycin. Plants were grown in a climate chamber under long day conditions (16h light, 8 h dark, 22°C) for about three weeks, before transfer to the greenhouse.

DNA isolation and molecular analysis of transposition

Genomic DNA was isolated from leaves of tissue culture-grown plantlets or greenhouse plants according to Pereira and Aarts (1998). PCR analysis: All PCR reactions were performed on 20 ng of genomic DNA according to standard conditions. Primers in *HPT* (HPT-F: AAAAGTTCGACAGCGTCTCCGACC and HPT-R: TCTACACAGCCATCGGTCCAGACG) and *En* transposase (En-1010F: CTGCAGCCAAACACATTTTCGC; En-1474R: ACCATGAGTGACACTGTCGAATCC) were used to check the presence of these components in the integrated constructs. Excision of *I/dSpm* in *En*-ET1 plants was monitored by amplification of Empty Donor Site (EDS) fragments using primers in the CaMV 35S terminator (t35S-R2: TCAACACATGAGCGAAACCC) and *SSU301* terminator (tSSU301-F1: GTTGGTTGAGAGTCTTGTGGCCT). This primer combination is expected to give a product only in case of excision. Nested PCR reactions were also performed using internal primers (t35S-R1: AGAGAGAGACTGGTGATTTCAGCG; tSSU301-F2: GAGGATGGAGAAGATGAAGCTG). Nested EDS products for each individual plant were cloned in pGEM-T (Promega) and several clones for each ligation were sequenced, in order to screen for the occurrence of independent excision events. The presence of untransposed *I/dSpm* in the donor T-DNA was revealed by amplification of Full Donor Site (FDS) fragments using the same primers described above in combination with primers in the *I/dSpm* Left (En-140R: AAGCGTCGGTTTCATCGGGAC) or Right (En-8023F: CAGGTAGCTTACTGATGTGCG) junctions. Southern blot analysis: Southern blot analyses were performed as previously described (Greco et al. 2001b). About 2 μg of DNA from different regenerants lines restricted with *Eco*RI were used. The same blots were hybridised with different

probes, after stripping in a boiling 0.1% SDS solution. Excision events were monitored by hybridisation with the petunia rubisco *SSU301* terminator, which is expected to detect a 4.6 kb EDS fragment in case of excision and/or a 6.4 kb FDS fragment in case of *I*/*dSpm* still residing in the donor T-DNA. Transposed *I*/*dSpm* elements were expected to be detected as fragments hybridising to a *GUS* probe with a minimum size of 2.1 kb. To determine the number of copies of the original T-DNA, blots were probed with the *HPT* gene. In addition, an *En*/*Spm* *Xba*I internal fragment was used to confirm the presence of the transposase.

RNA isolation and expression analysis

Northern blot analysis: Total RNA from three weeks old *in vitro* plantlets was isolated as described by Verwoerd et al. (1989) and poly(A)$^+$ purified using the "mRNA purification kit" (PHARMACIA), following the manufacturer's instructions. About 6 μg of poly (A)$^+$ RNA for each sample were separated on a 1.5% formaldehyde agarose gel and blotted in 10x SSC onto a Hybond-N$^+$ membrane (Amersham, UK). Pre-hybridization and hybridization were performed at 65°C in 10% Dextran Sulfate, 1% SDS, 1M NaCl, 0.4 mg/ml sonicated salmon sperm DNA. After hybridisation, the membrane was washed twice in 1X SSC, 1% SDS at 65°C. Hybridisation signals were quantified using a Bas2000 PhosphorImager (Fujix) and TINA software (Raytest, Germany). To detect all the maize *En*/*Spm* transcripts, a probe spanning part of ORF2 and exons II-III-IV was obtained as a 1.4-kb *Eco*RI fragment (from pos. 4459 till 5836 of *En*-1).

RT-PCR analysis: Semi-quantitative RT-PCR was performed according to Frohman et al. (1988) on cDNA synthesised starting from 500 ng of poly(A)$^+$ RNA with Superscript reverse transcriptase (Invitrogen), as recommended by the manufacturer. The *En/Spm* specific primers used were designed based on Masson et al. (1989) at the following positions in the *En*-1 sequence: 560-Forward (primer 1); 2647-Reverse (primer 6); 3781-Forward (primer 7); 5299-Reverse (primer 11); 6059-Forward (primer 12); 6402-Reverse (primer 13); 7761-Reverse (primer 14). To ease the interpretation of the results, primers located at similar positions were called as in Masson et al. (1989). Aliquots of the amplified products were separated on agarose gel, blotted and hybridised to an *En*/*Spm* specific probe.

Isolation of I/dSpm flanking sites

Genomic *I*/*dSpm* insertion sites were isolated by inverse-PCR as described by Earp et al. (1990). About 400 ng of genomic DNA were digested with *Hinf*I or *Nla*III. After self-ligation and linearization with *Sal*I, primers in the *GUS* gene (GUS-F1: GTGAACAACGAACTGAACTGG; GUS-F2: GGTGATTACCGACGAAAACGGC) or in the CaMV 35S promoter (p35S-F2:

CGCACAATCCCACTATCCTTCGCAA and p35S-F1) in combination with primers in the left border of the *I/dSpm* element (En-140R and En-21R: CTTTGACGTTTTCTTGTAGTG) were used in primary and nested PCR reactions to amplify the flanking genomic DNA. The inverse-PCR products were cloned in pGEM-T (Promega) and sequenced. Similarity searches in public databases were performed using the Blast algorithm (Altschul et al. 1997) at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) or at Gramene (http://www.gramene.org). Additional information on chromosomal position was derived from the International Rice Genome Sequencing Project (http://rgp.dna.affrc.go.jp/IRGSP/index.html).

RESULTS

En/I transposition activity in rice

The *En*/*I* heterologous maize transposon system was evaluated as an additional tool for transposon tagging in rice. About 14 transgenic rice lines (65 plants) were produced containing a two-component construct (*En*-ET1, Fig. 1) with an immobilized *En* transposase source, driven by the CaMV35S promoter (Aarts et al, 1995) and a mobile *I* element. The latter was modified to bear the *BAR* resistance gene (DeBlock et al. 1987) as a positive selectable marker and the *GUS* reporter gene driven by a CaMV35S minimal promoter for enhancer detection. The *SU1* gene was employed as a counter-selectable marker on the T-DNA against the transposase source. Also lines containing only the immobilized transposase in combination with the negative selection marker were made, for further crossing with *I* stable plants (*En*-TPase; Fig. 1).

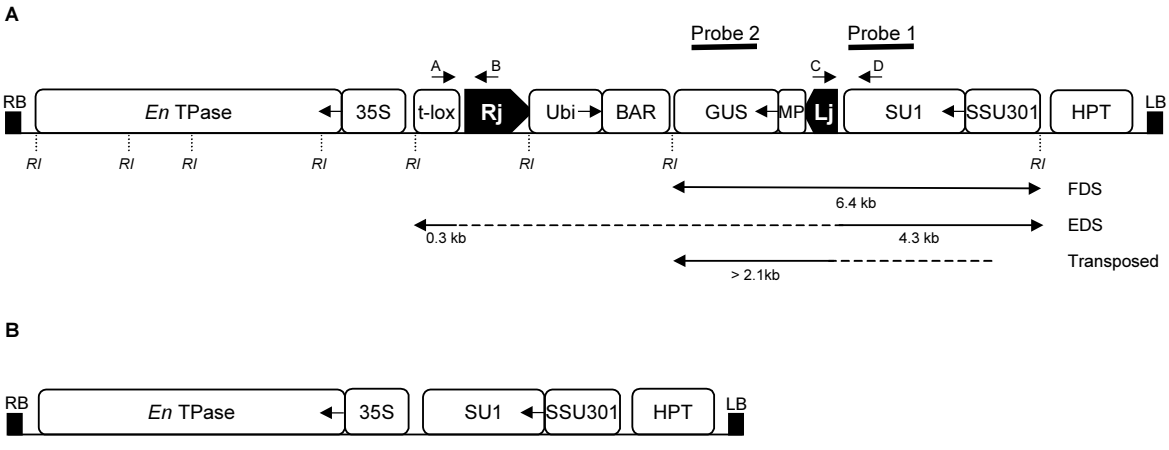


Figure 1. Diagram of the *En*/*Spm* transposon constructs used for rice transformation. A) *En*-ET1 two-component construct. The position of the primers used to detect EDS (A+D) or FDS (A+B, C+D) are indicated, along with the probes used for Southern analysis of transposition (probe 1, excision; probe 2, transposition) and the size of the

expected hybridising bands. B) *En*-Tpase transposase construct. LB, RB, left and right T-DNA borders; *En*TPase, immobilised *En* transposase; Lj, Rj, *I/dSpm* left and right borders; HPT, hygromycin phosphotransferase gene; MP, CaMV 35S minimal promoter and *lox* site; SSU301, petunia Rubisco small subunit *SSU301* promoter; t-*lox*, CaMV 35S terminator and *lox* site; Ubi, maize Ubiquitin promoter; 35S, CaMV 35S promoter.

Six T_0 independent *En*-ET1 lines (for a total of 20 plants) were identified having an intact single copy T-DNA, after PCR analysis and Southern hybridisation with primers or probes corresponding to the different components (see Material and methods; data not shown). All the other transgenic plants produced, excluding the transposase lines, revealed partial or severe rearrangements probably due to the instability of the complex T-DNA construct.

Excision in the *En*-ET1 lines was monitored by PCR using primers to amplify the Empty Donor Site (EDS; Fig. 1). A faint EDS band could only be amplified in four T_0 primary regenerants originating from three independent lines, indicating the occurrence of rare somatic excision events. Subsequently, Southern blot hybridisation analysis with a *SSU301* terminator-homologous probe could reveal only the Full Donor Site (FDS) fragment, but failed to reveal any EDS-hybridising fragment (data not shown). An analogous result was obtained after testing the T_1 progeny of the PCR positive plants, thereby confirming the absence of germinal excisions.

Table 1. Empty donor site footprints in T_0 *En*-ET1 regenerants and T_1 progeny plants.

	Waxy (maize) GATTCGTATATACCTCGTCTG *En* at *wx*-844 GATTCGTATATA - ***En*** - ATACCTCGTCTG		
		T_0	T_1
EDS Plant 4.2	GATTCGTATATA - T - ATACCTCGTCTG	2/3	5/8
	GATTCGTATATA * TACCTCGTCTG	1/3	2/8
	GATTCGTATATA * * * CCTCGTCTG	-	1/8
EDS Plant 7.2	GATTCGTATATA - T - ATACCTCGTCTG	3/5	Not tested
	GATTCGTATATA * TACCTCGTCTG	1/5	
	GATTCGTATATA * * * CCTCGTCTG	1/5	
EDS Plant 13.4	GATTCGTATATA - T - ATACCTCGTCTG	2/4	Not tested
	GATTCGTATATA * TACCTCGTCTG	1/4	
	GATTCGTACATA * TACCTCGTCTG	1/4	
EDS Plant 13.7	GATTCGTATATA - T - ATACCTCGTCTG	4/5	1/4
	GATTCGTATATA * * * CCTCGTCTG	1/5	2/4
	GATTCGTATATA * TACCTCGTCTG	-	1/4

Sequence analysis of the EDS fragments demonstrated the presence of the characteristic molecular footprint left behind after *I/dSpm* excision (Table 1). As expected for late excision in somatic sectors, different footprints were revealed in each of the T_0 plant analysed, and in the T_1 progenies as well. Interestingly, an excision product leaving an additional thymidine between the triplets of the Target Site Duplication (TSD) was preferentially found in all the T_0 plants and in

some of the T_1 progenies. The same footprint was found as the most abundant after sequence analysis of *En/Spm* excision products in potato (Frey et al. 1989).

Reintegration of the *I/dSpm* element at different locations in the rice genome, following somatic excision, was further demonstrated by the isolation of multiple genomic sequences flanking the *I/dSpm* transposon in two of the T_0 plants (Table 2). Similarity searches in public databases revealed homology to rice genomic sequences for all the Insertion Tagged Sites (ITS) isolated. Their chromosomal distribution would suggest that the original T-DNA donor sites were located on chromosome 4 in plant 4.2 and chromosome 3 in plant 13.7. The inserts at a chromosome are distributed over a loosely linked 50 cM interval, 19.6-70.9 on chromosome 4 and 89.2-142.3 on chromosome 3. This is unlike the closely linked transpositions of the *Ac/Ds* system. Previous studies in maize reported a frequency of transposition to unlinked genomic positions for the *En/Spm* transposon system of about 75% (Peterson 1970). Accordingly, in rice transposition of the *I/dSpm* inserts occurred to loosely linked sites as well as unlinked sites.

The finding of transposon reinsertion sites, together with the presence of EDS, demonstrates the occurrence of the complete process of *En/I* transposition in rice. However, the absence of germinal excisions indicates a low frequency of transposition events.

Table 2. *I/dSpm* Insertion tagged sites (ITS) in *En*-ET1 transgenic T_0 rice plants.

Plant	ITS	Size[a]	Clone[b]	E-value[c]	Chromosome[d]	Gene/Protein prediction[e]
4.2	H1	72	AP003266	6.6e-07	1 (163.8)	Ups 2.9 kb: hypothetical BAB64192 Dwn 1 kb: hypothetical BAB64193
	H2	93	AL606611	1.8e-10	4 (19.6)	Exon of CAD40195 (transposase)
	N1	307	AL731591	5.6e-39	4 (49.7-52.6)	Ups 1.6 kb: GRMG00000065950 (put. AP2 domain transcription factor)
	N2	194	AL606441	1.1e-36	4 (70.9)	Intergenic
13.7	H1	128	AC133860	3.5e-22	3 (89.2-90.3)	Ups 3.5 kb: GRMG00000090202 (put. lipase-related protein)
	H2	87	AC099043	4.6e-12	3 (142.3)	Ups 0.7 kb: GRMG00000026400 (cold acclimation protein WCOR413-like) Dwn 0.9 kb: GRMG00000026401 (put. selenium-binding protein)
	H3	77	AC120505	1.2e-10	3 (96.6)	Dwn 2 kb: GRMT00000077656 (hypothetical)
	N1	141	AC128644	4.1e-25	11 (20.3-27.8)	Dwn 5 kb: GRMG00000091698 (put. polyprotein)

Splicing of the maize En/Spm transposon in rice

In order to investigate why the *En*/*Spm* system transposes at a low efficiency in rice, we decided to carry out molecular analyses on the transcription and splicing of the element. Earlier studies in maize have shown that the active transposases encoded by the *En*/*Spm* autonomous element originate from alternative splicing of the same primary transcript (Pereira et al. 1986; Masson et al. 1989). Of the four products identified, TnpA and TnpD seem so far required for transposition (Frey et al. 1990; Gierl et al. 1988; Masson et al. 1989, 1991). It was proposed that TnpA binds to the sub-terminal regions of the element promoting their joining, while TnpD stabilises the TnpA-DNA complexes (Gierl et al. 1988; Raina et al. 1998). Moreover, the 2.5 kb transcript corresponding to TnpA is by far the most abundant transcript in the maize cell, as it was shown to be present in at least 100 folds more than the 6 kb TnpD transcript (Pereira et al. 1986).

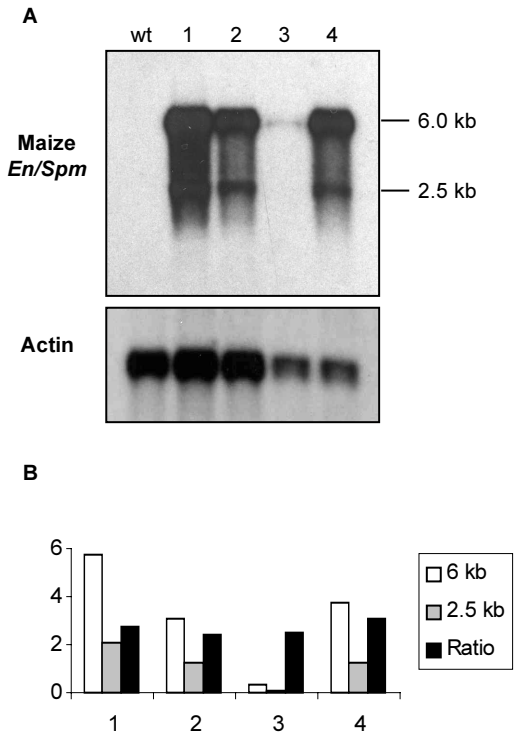


Figure 2. Northern blot analysis of *En*/*Spm* specific transcripts in transgenic rice plants. A) Poly $(A)^+$ RNA was isolated from wild type (wt) and transgenic rice plants containing construct *En*-ET1 (1: plant 4.2; 2: plant 13.7) and *En*TPase (3: plant 1.4; 4: plant 7.9). Hybridisation was performed with a probe specific for the 3' region of *En*/*Spm*. The same blot was then re-hybridised with an actin tobacco cDNA probe for loading control. B) Signals were quantified by measuring the OD/mm^2 and the ratio between the 6 kb TnpD and the 2.5 kb TnpA transcripts calculated for each plant.

The transcription of *En*/*Spm* in rice was monitored by RNA gel blot analysis on poly$(A)^+$ RNA isolated from two T_1 *En*-ET1 enhancer trap regenerants that showed transposition activity (Table 1 and 2) and from two *En* transposase lines, with three and one T-DNA copy respectively (Fig. 2). The probe used was expected to reveal all the alternatively spliced products, since it spanned a region comprising the 3' portion of the primary transcript (Pereira et al. 1986). Indeed, the 2.5 kb fragment corresponding to TnpA could be revealed, though much less abundant than the larger transcript(s), which is in sharp contrast with the situation in maize cells (Pereira et al.

1986; Masson et al. 1989). Smaller transcripts of 0.2-0.7 kb in size, shown to be present in maize plants containing active *En/Spm* elements (Masson et al. 1989), were not visible in rice. No cross-hybridisation with rice endogenous *En*/*Spm* elements was observed, as shown by the absence of hybridisation with mRNA from non-transgenic rice.

To confirm the presence of all the different alternatively spliced transcripts, RT-PCR analysis was performed on the same lines using primers designed along the maize *En*/*Spm* element, similarly to Masson et al. (1989; Fig. 3 and Material and methods). In the analysis we also included an Arabidopsis line containing the same *En/Spm* transposase source used for rice, known to promote active transposition of the *I/dSpm* element in Arabidopsis (TEn5; Aarts et al. 1995). A total of six primer combinations were chosen to check the presence of the large TnpB, TnpC, TnpD transcripts and the TnpA transcript.

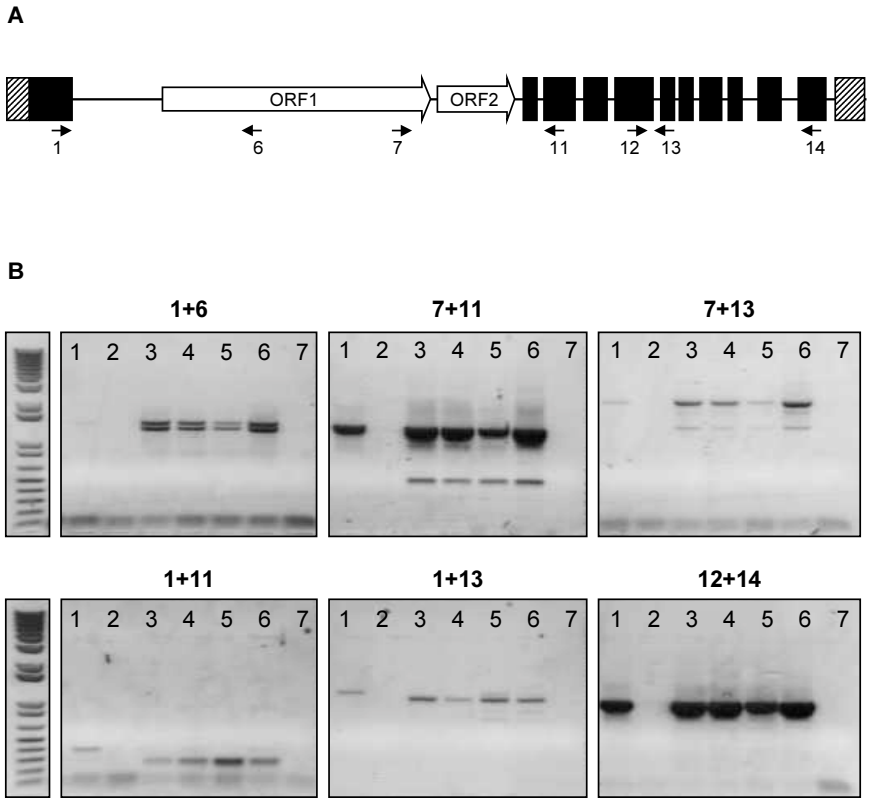


Figure 3. Analysis of *En/Spm* splicing products in transgenic rice and Arabidopsis plants. A) Diagram of the maize *En/Spm* transposon with the location of the primers used for RT-PCR. Exons, ORFs in the first intron and the sub-terminal repetitive regions are represented as black boxes, open arrows and hatched boxes, respectively. B) Ethidium bromide stained agarose gel of RT-PCR products obtained with the primer combinations described in Table 3. Lane 1, Arabidopsis TEn5; lane 2, rice wild-type; lane 3, rice *En*-ET1 plant 4.2; lane 4, rice *En*-ET1 plant 13.7; lane 5, rice *En*-Tpase plant 1.4; lane 6, rice *En*-Tpase plant 7.9; lane 7, water control. The 1 Kb Plus DNA ladder is shown at the left side.

The RT-PCR products obtained were of the expected size (Pereira et al. 1986; Masson et al. 1989), confirming that the element could be in principle correctly processed (Fig. 3; Table 3). However, differences in *En/Spm* splicing between rice and Arabidopsis could be revealed, as previously described for maize and tobacco (Masson et al. 1989). Transcripts containing a longer exon 1 spliced at position 704 were preferentially found in Arabidopsis, similarly to tobacco. On the other hand, rice rather uses a splicing site at position 578 of exon 1 or both, as seen in maize (Masson et al. 1989). Accordingly, the size of the RT-PCR products we obtained in Arabidopsis were similar to the ones obtained for tobacco, while splicing products in rice were similar to maize (Table 3). Moreover, the preference of Arabidopsis for the 704 splicing site seems to be absolute, as no different products could be amplified. In rice, the TnpA transcript seems though to be produced exclusively using the splicing site at position 578.

Table 3. RT-PCR analysis of *En/Spm* splicing products in transgenic rice and Arabidopsis plants.

Primers	Product	Fragment length (kb) [a]			
		Maize [b]	*Rice*	*Tobacco* [b]	*Arabidopsis*
1 + 6	TnpB,C,D	*1.38* [c]	1.38	*1.38* [c]	1.38
	"	*1.26* [d]	1.26	*(1.26)* [d]	-
7 + 11	TnpD	*1.28*	1.28	*1.28*	1.28
	TnpC	*1.18*	(1.18)	*(1.18)*	(1.18)
	TnpB	*0.49*	(0.49)	-	(0.49)
7 + 13	TnpD	*2.14*	2.14	-	2.14
	TnpC	*2.05*	(2.05)	-	-
	TnpB	*1.36*	(1.36)	-	-
1 + 11	TnpA	*(0.36)* [c]	-	*0.36* [c]	0.36
	"	*0.24* [d]	0.24	*(0.24)* [d]	-
1 + 13	TnpA	*1.1* [d]	1.1	*1.23* [c]	1.23
12 + 14	all	*0.97*	0.97	*0.97*	0.97

[a] Less abundant products are reported within brackets.
[b] The sizes were calculated according to Masson et al. (1989).
[c] Splicing using the donor site at position 704 of exon 1.
[d] Splicing using the donor site at position 578 of exon 1.

In addition, all the transposase transcripts were shown to be produced in rice, with a relative higher abundance of TnpD over TnpB and TnpC. In agreement with the results of the RNA gel blot experiments, also the ratio between the TnpD and TnpA transcripts seemed to be altered in favour of the larger product, consistently in all the lines containing the transposase. In contrast in

Arabidopsis, only the TnpA and TnpD transcripts could be detected, even after hybridisation of the RT-PCR products with an *En/Spm* specific probe (data not shown). Additional products probably originating from aberrant splicing were also detected in both species, though at very low amount.

Independence of rice and maize En/Spm elements

We assessed the ability of endogenous *En*/*Spm*-like encoded transposases in the rice genome to trans-activate introduced *I*/*dSpm* defective elements from maize. For this purpose, we employed transgenic rice plants in which the transposase fragment was not transferred in the T-DNA, but contained an intact non-autonomous *I*/*dSpm* element. If trans-activation by endogenous rice transposons could take place, it would result in excision of the non-autonomous maize element. However, PCR experiments done to check the presence of EDS in progeny plants from several *I/dSpm* lines revealed absence of excision (data not shown). This suggests that the *En/Spm* related elements in rice do not excise the maize *I/dSpm* elements, and their excision is dependent on the maize *En/Spm* transposase expression. The interaction of the maize *En/Spm* transposase expression with rice endogenous elements was also tested and will be reported elsewhere.

DISCUSSION

With the production of transgenic rice lines containing the maize *En*/*Spm* transposon system, we describe for the first time transposition of this heterologous system in rice. We demonstrate that the complete process of transposition, that is excision followed by reinsertion at different genomic positions, is occurring in our lines. However, the overall frequency of transposition appeared to be low and restricted to somatic events, indicating that this maize system is working at low efficiency in rice. This is surprising, considering the high activity of *En/I* in Arabidopsis and in comparison to the other maize transposon system *Ac*/*Ds*, whose mobility is high in early generations in rice.

Low transposition efficiency might be considered the result of incorrect splicing of the maize *En*/*Spm* element in rice. As a consequence, the right transposase products TnpA and TnpD may not be produced, or may be produced in altered proportions that influence negatively the efficiency of transposition. Splicing of *En/Spm* was previously investigated in potato (Frey et al. 1989) and tobacco (Pereira and Saedler 1989), where the native maize element generally

transposes at low efficiency, except for some tobacco lines with multiple autonomous *En/Spm* elements. Northern analyses revealed higher amount of the longer TnpD transcript in relation to the TnpA, providing a clue that alternative splicing was not working efficiently in dicots. Similarly in rice, we could demonstrate only somatic transposition and Northern analysis revealed a major abundance of the large 6 kb transcripts relatively to the 2.5 kb TnpA specific transcript. On the other hand, the *En/Spm* system transposes efficiently in Arabidopsis (Aarts et al 1995; Cardon et al 1993b), where splicing produces the TnpA and TnpD products in a proportion similar to that seen in maize.

A more detailed analysis of *En/Spm* splicing by RT-PCR, revealed differences between dicot and monocot host systems, in agreement with previous studies highlighting difficulties in the processing of monocot transcripts in dicot species (Goodall and Filipowicz 1991; Keith and Chua 1986). In tobacco, alternative choice of splicing sites at the end of exon 1 resulted in the production of transcripts thought to have a lower efficiency of translation (Masson et al. 1989). In this work, we demonstrate that *En/Spm* splicing in the dicot Arabidopsis is similar to tobacco, while splicing in rice is similar to maize (Masson et al. 1989). Yet both the transcripts encoding the active TnpA and TnpD transposases are formed in all the hosts. The absence of the TnpB and TnpC splicing products in Arabidopsis, supports the notion that these products are not necessary for transposition. Aberrant transcripts of smaller size were also detected in *En/Spm* transgenic potato, tobacco and Arabidopsis lines (Frey et al. 1989; Masson et al. 1989; Pereira and Saedler 1989; Cardon et al.1993b). In maize relatively abundant shorter transcripts, mainly derived from transcription of defective elements, can produce aberrant transposase products that function as repressors of the wild-type transposase, thus inhibiting transposition (Cuypers et al. 1988). Such transcripts were detected in very low amounts in our transgenic rice lines, and might not be enough to trigger this response. In conclusion, it seems that the presence of the correctly spliced TnpA and TnpD transcripts does not necessarily correlate with the ability of the element to transpose, while their relative proportion does. This observation supports the idea that host factors influencing the efficiency at which the different products are produced (and therefore the final abundance of the active transposases in the cell) play a more crucial role in mediating efficient transposition. Moreover, splicing efficiency seems to be independent from the host being dicot or monocot. In rice, it therefore seems that a limiting amount of TnpA is formed that is probably not sufficient to promote efficient transposition.

Another reason for the lower transposition efficiency could be the lack of critical sequences in the modified *I/dSpm* mobile element used. Studies on the behaviour of natural deletion derivatives of the *En*/*Spm* transposon in maize indicated that the size of the sub-terminal regions has a dramatic influence on the transposition process. In these regions, extending for 180 bp from the left end and 299 bp from the right end, reside several copies of the 12-bp consensus site at

which TnpA is supposed to bind to promote excision (Gierl et al. 1985, 1988; Masson et al. 1987). There are 9 and 15 copies of this consensus sequence at the left and right terminus, respectively (Masson et al. 1987). Deletions involving loss of some of these binding sites at both termini of *En*/*Spm* cause decreased frequency and timing of transposition, while absence at one terminus completely abolish excision (Schiefelbein et al. 1985; Schwarz-Sommer et al. 1985; Tacke et al. 1986; Menssen et al. 1990; Aukerman and Schmidt, 1993). Previous reports on transposition of *En*/*Spm* in the heterologous systems Arabidopsis and tobacco made use of unmodified autonomous and non-autonomous elements (Pereira and Saedler 1989; Cardon et al. 1993b; Aarts et al. 1995; Wisman et al. 1998; Speulman et al. 1999). When the *I/dSpm* transposon was modified to carry selectable markers, left and right arms not smaller than 1015 bp and 786 bp respectively were used (Masson and Fedoroff 1989; Cardon et al. 1993a; Tissier et al. 1999). In our lab, *I/dSpm* elements modified to function as gene traps in Arabidopsis revealed excision ability only in case a minimal left junction (*I*-Lj) of at least 0.19 kb was used, in combination with a 0.64 kb right junction (*I*-Rj; Aarts, 1996). Shorter *I*-Lj abolished excision, presumably due to the reduction in the TnpA binding site sequences or their accessibility. The excision frequencies for these modified *I/dSpm* were nevertheless lower than for wild-type elements (Aarts 1996). Though the insertion of transcribed genes on the *I/dSpm* element was shown to have a negative influence on the transposition rate in tobacco and Arabidopsis (Cardon et al. 1993a; Tissier et al. 1999), an additional negative effect due to a sub-optimal length of the left border can not be ruled out. Based on these results, a longer fragment of 0.27 kb and a 0.64 kb fragment were used as left and right junctions of the *I/dSpm* element in the enhancer trap construct *En*-ET1 developed for rice transformation. Theoretically, as all the TnpA binding sites are supposed to be present in the transposon ends used, transposition is expected to occur if no other factors are negatively influencing the process. Even though a similar construct was employed in our lab for the successful development of an activation tagging population in Arabidopsis (Marsch-Martinez et al. 2002), the low transposition efficiency obtained in rice still hints to the possibility that additional transposon specific sequences are required. A GC-rich region adjacent to the sub-terminal left border and extending up to 860 bp from the left end of the *En*/*Spm* element, was indeed found to be required for efficient transposition in maize (Masson et al. 1987), and might be important in the related monocot rice.

On the other hand, the presence of several *En/Spm* related sequences in the rice genome (Goff et al. 2002) suggests the possibility that endogenous elements might also interact with other members of the same family and alter their transposition behaviour. In addition, cross-talks between endogenous and exogenous related transposons could occur, resulting in transactivation of either mobile members. Interaction between transposable elements was already described in rice, even for elements belonging to distinct transposon families. For the recently discovered

activity of the MITE elements *mPing*, it was suggested that cross-talk between related but different transposase sources (*Ping* or *Pong*) and the defective elements could promote mobility, at least in rice cell cultures (Jiang et al. 2003). Similarly, amplification of *Stowaway* MITEs in the rice genome appeared to be due to cross-mobilisation by distantly related *Mariner*-like autonomous transposable elements (Feschotte et al. 2003). The presence of consensus binding sites for the active transposase in the termini of the cross-mobilised elements is though a pre-requisite for mobilisation. This phenomenon occurs also in maize, where *Ac* autonomous elements can promote excision of a distantly related *Ds* defective element sharing only short homologous sequences in the terminal inverted repeats (Shen et al. 1988). In addition *En/Spm* related sequences present in the maize genome, like the *Doppia* element that shares homology to the TnpA and TnpD proteins (Bercury et al. 2001) might as well interact with each other. Indeed, the *En/Spm* element was shown to excise related members with homologous ends, in the presence of another element function termed Mediator (Muszynski et al. 1993). Transactivation of exogenous maize *I*/*dSpm* defective elements by endogenous *En*/*Spm*-like encoded transposases in our transgenic rice lines was tested, but no evidence of excision could be revealed. Indeed, no active members of the *En*/*Spm*-like transposon family were described in rice so far. Nevertheless, the absence of cross-mobilisation does not exclude the presence of active elements and the possibility of interference cannot be ruled out.

In conclusion we demonstrate that the maize *En/Spm*-*I/dSpm* two component system can transpose in rice. Further improvements, by providing the TnpA and TnpD cDNAs for transposase expression as successfully done in tobacco (Cardon et al. 1993a) and using longer *I/dSpm* transposon ends, could make *En/Spm* a practical alternative to the other maize *Ac/Ds* system for transposon mutagenesis in rice.

Acknowledgments

This research was funded in the participating labs by a European Union project BIO4 CT 972132 on “Transposon mutagenesis in rice”. The work has been carried out in compliance with the current laws governing genetic experimentation in The Netherlands.

LITERATURE CITED

Aarts MGM (1996) An *En/Spm* based transposable element system for gene isolation in *Arabidopsis thaliana*. PhD Thesis Wageningen University, Wageningen, The Netherlands.

Aarts MGM, Corzaan P, Stiekema WJ, Pereira A (1995) A two-element *Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana*. Mol Gen Genet 247: 555-564

Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402

Aukerman M, Schmidt R (1993) A 168 bp derivative of *Suppressor-mutator/Enhancer* is responsible for the maize *o2-23* mutation. Plant Mol Biol 21: 355-362

Barry G (2001) The use of the Monsanto draft rice genome sequence in research. Plant Physiol 125: 1164-1165

Bercury S, Panavas T, Irenze K, Walker E (2001) Molecular analysis of the *Doppia* transposable element of maize. Plant Mol Biol 47: 341-351

Cardon GH, Frey M, Saedler H, Gierl A (1993a) Definition and characterization of an artificial *En/Spm*-based transposon tagging system in transgenic tobacco. Plant Mol Biol 23: 157-178

Cardon GH, Frey M, Saedler H, Gierl A (1993b) Mobility of the maize transposable element *En/Spm* in *Arabidopsis thaliana*. Plant J 3: 773-784

Chin HG, Choe MS, Lee SH, Park SH, Park SH, Koo JC, Kim NY, Lee JJ, Oh BG, Yi GH, Kim SC, Choi HC, Cho MJ, Han CD (1999) Molecular analysis of rice plants harboring an *Ac/Ds* transposable element-mediated gene trapping system. Plant J 19: 615-623.

Christensen AH, Quail PH (1996) Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res 5: 213-218

Cuypers H, Dash S, Peterson P, Saedler H, Gierl A (1988) The defective *En*-I102 element encodes a product reducing the mutability of the *En/Spm* transposable element of *Zea mays*. EMBO J 7: 2953-2960

De Block M, Botterman J, Vanderwiele M, Dockx J, Thoen C, Gossele V, Movva R, Thompson C, Van Montagu M, Leemans J (1987) Engineering herbicide resistance in plants by expression of a detoxifying enzyme. EMBO J 6: 2513-2518

Earp DJ, Lowe B, Baker B (1990) Amplification of genomic sequences flanking transposable elements in host and heterologous plants: a tool for transposon tagging and genome characterization. Nucleic Acids Res 18: 3271-3278

Enoki H, Izawa T, Kawahara M, Komatsu M, Koh S, Kyozuka J, Shimamoto K (1999) *Ac* as a tool for the functional genomics of rice. Plant J. 19: 605-613

Fedoroff NV, Smith DL (1993) A versatile system for detecting transposition in Arabidopsis. Plant J 3: 273-289

Feng Q, Zhang Y, Hao P, Wang S, Fu G et al. (2002) Sequence and analysis of rice chromosome 4. Nature 420: 316-320

Feschotte C, Swamy L, Wessler SR (2003) Genome-wide analysis of *Mariner*-like transposable elements in rice reveals complex relationships with *Stowaway* miniature inverted repeat transposable elements (MITEs). Genetics 163: 747-758

Frey M, Reinecke J, Grant S, Saedler H, Gierl A (1990) Excision of the *En/Spm* transposable element of *Zea mays* requires two element-encoded proteins. EMBO J 9: 4037-4044

Frey M, Tavantzis SM, Saedler H (1989) The maize *En*-1/*Spm* element transposes in potato. Mol Gen Genet 217: 172-177

Frohman M, Dush M, Martin G (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci 85: 8998-9002

Gierl A, Lutticke S, Saedler H (1988) TnpA product encoded by the transposable element *En*-1 of *Zea mays* is a DNA binding protein. EMBO J 7: 4045-4053

Gierl A, Schwarz-Sommer Z, Saedler H (1985) Molecular interactions between the components of the *En-I* transposable element system of *Zea mays*. EMBO J 4: 579-583

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

Goodall GJ, Filipowicz W (1991) Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. EMBO J 10: 2635-2644

Greco R, Ouwerkerk PBF, Sallaud C, Kohli A, Colombo L, Puigdomenech P, Guiderdoni E, Christou P, Hoge JHC, Pereira A (2001a) Transposon insertional mutagenesis in rice. Plant Physiol 125: 1175-1177

Greco R, Ouwerkerk PBF, Taal AJC, Favalli C, Beguiristain T, Puigdomenech P, Colombo L, Hoge JHC, Pereira A (2001b) Early and multiple *Ac* transpositions in rice suitable for efficient insertional mutagenesis. Plant Mol Biol 46: 215-227

Hirochika H (2001) Contribution of the *Tos17* retrotransposon to rice functional genomics. Curr Opin Plant Biol 4: 118-122

Hoess R, Ziese M, Sternberg N (1982) *P1* site-specific recombination: nucleotide sequence of the recombining sites. Proc Natl Acad Sci 79: 3398-3402

Izawa T, Ohnishi T, Nakano T, Ishida N, Enoki H, Hashimoto H, Itoh K, Terada R, Wu C, Miiyazaki C, Endo T, Iida S, Shimamoto K (1997) Transposon tagging in rice. Plant Mol Biol 35: 219-229

Izawa T, Shimamoto K (1996) Becoming a model plant: The importance of rice to plant science. Trends Plant Sci 1: 95-99

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

Jeon JS, Lee S, Jung KH, Jun SH, Jeong DH, Lee J, Kim C, Jang S, Lee S, Yang K, Nam J, An K, Han MJ, Sung RJ, Choi HS, Yu JH, Choi JH, Cho SY, Cha SS, Kim SI, An G (2000) T-DNA insertional mutagenesis for functional genomics in rice. Plant J. 22: 561-570

Jeong D, An S, Kang H, Moon S, Han J, Park S, Lee H, An K, G A (2002) T-DNA Insertional Mutagenesis for Activation Tagging in Rice. Plant Physiol 130: 1636-1644

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA transposon family in rice. Nature 421: 163-167

Keith B, Chua N (1986) Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco. EMBO J 5: 2419-2425

Kohli A, Xiong J, Greco R, Christou P, Pereira A (2001) Tagged Transcriptome Display (TTD) in *indica* rice using *Ac* transposition. Mol Gen Genom 266: 1-11

Marsch-Martinez N, Greco R, Van Arkel G, Herrera Estrella L, Pereira A (2002) Activation tagging using the *En-I* maize transposon system in Arabidopsis. Plant Physiol 129: 1544-1556

Martienssen R (1998) Functional genomics: probing plant gene function and expression with transposons. Proc Natl Acad Sci 95: 2021-2026

Masson P, Fedoroff NV (1989) Mobility of the maize *Suppressor-mutator* element in transgenic tobacco cells. Proc Natl Acad Sci 86: 2219-2223

Masson P, Rutherford G, Banks JA, Fedoroff N (1989) Essential large transcripts of the maize *Spm* transposable element are generated by alternative splicing. Cell 58: 755-765

Masson P, Strem M, Fedoroff N (1991) The tnpA and tnpD gene products of the *Spm* element are required for transposition in tobacco. Plant Cell 3: 73-85

Masson P, Surosky R, Kingsbury JA, Fedoroff NV (1987) Genetic and molecular analysis of the *Spm*-dependent *a-m2* alleles of the maize a locus. Genetics 117: 117-137

Menssen A, Hohmann S, Martin W, Schnable PS, Peterson PA, Saedler H, Gierl A (1990) The *En/Spm* transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a *dSpm* element in the *A2* gene. EMBO J 9: 3051-3057

Muszynski MG, Gierl A, Peterson PA (1993) Genetic and molecular analysis of a three-component transposable-element system in maize. Mol Gen Genet 237: 105-112

Nakagawa Y, Machida C, Machida Y, Toriyama K (2000) Frequency and pattern of transposition of the maize transposable element *Ds* in transgenic rice plants. Plant Cell Physiol 41: 733-742

O'Keefe D, Tepperman J, Dean C, Leto K, Erbes D, Odell J (1994) Plant expression of a bacterial cytochrome p450 that catalyzes activation of a sulfonylurea pro-herbicide. Plant Physiol 105: 473-482

Parinov S, Sundaresan V (2000) Functional genomics in Arabidopsis: large-scale insertional mutagenesis complements the genome sequencing project. Curr Opin Biotechnol 11: 157-161

Pereira A (2000) A transgenic perspective on plant functional genomics. Transgenic Res 9: 245-260

Pereira A, Aarts MGM (1998) Transposon tagging with the *En-I* system. In: Martinez-Zapater J, Salinas J (eds) Arabidopsis protocols. Humana Press, Totowa, NY, pp 329-338

Pereira A, Saedler H (1989) Transpositional behavior of the maize *En/Spm* element in transgenic tobacco. EMBO J 8: 1315-1321

Pereira A, Cuypers H, Gierl A, Schwarz-Sommer Z, Saedler H (1986) Molecular analysis of the *En/Spm* transposable element system of *Zea mays*. EMBO J 5: 835-841

Pereira A, Schwarz-Sommer Z, Gierl A, Bertram I, Peterson PA, Saedler H (1985) Genetic and molecular analysis of the *Enhancer* (*En*) transposable element system of *Zea mays*. EMBO J 4: 17-23

Peterson PA (1970) The *En* mutable system in maize. III. Transposition associated with mutaional events. Theor Appl Genet 40: 367-377

Pietrzak M, Shillito RD, Hohn T, Potrykus I (1986) Expression in plants of two bacterial antibiotic resistance genes after protoplast transformation with a new plant expression vector. Nucleic Acids Res 14: 5857-5868

Raina R, Schlappi M, Karunanandaa B, Elhofy A, Fedoroff N (1998) Concerted formation of macromolecular *Suppressor-mutator* transposition complexes. Proc Natl Acad Sci 95: 8526-8531

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T et al. (2002) The genome sequence and structure of rice chromosome 1. Nature 420: 312-316

Shen W, Ramos C, Hohn B (1998) Excision of *Ds1* from the genome of maize streak virus in response to different transposase-encoding genes. Plant Mol Biol 36: 387-392

Schiefelbein JW, Raboy V, Fedoroff NV, Nelson OE, Jr. (1985) Deletions within a defective *Suppressor-mutator* element in maize affect the frequency and developmental timing of its excision from the bronze locus. Proc Natl Acad Sci 82: 4783-4787

Schwarz-Sommer Z, Gierl a, Berndtgen R, Saedler H (1985) Sequence comparison of 'states' of *a1-m1* suggests a model of *Spm* (*En*) action. EMBO J 4: 2439-2443

Speulman E, Metz PL, van Arkel G, te Lintel Hekkert B, Stiekema WJ, Pereira A (1999) A two-component *Enhancer-Inhibitor* transposon mutagenesis system for functional analysis of the Arabidopsis genome. Plant Cell 11: 1853-1866

Springer PS, McCombie WR, Sundaresan V, Martienssen RA (1995) Gene trap tagging of *PROLIFERA*, an essential MCM2-3-5-like gene in Arabidopsis. Science 268: 877-880

Sundaresan V, Springer P, Volpe T, Haward S, Jones JD, Dean C, Ma H, Martienssen R (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev 9: 1797-1810

Tacke E, Schwarz-Sommer Z, Peterson P, Saedler H (1986) Molecular analysis of states of the *A* locus of *Zea mays*. Maydica 31: 83-91

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JDG (1999) Multiple independent defective *Suppressor-mutator* transposon insertions in Arabidopsis: A tool for functional genomics. Plant Cell 11: 1841-1852

Tsugeki R, Kochieva EZ, Fedoroff NV (1996) A transposon insertion in the Arabidopsis *SSR16* gene causes an embryo-defective lethal mutation. Plant J 10: 479-489

Upadhyaya NM, Zhou XR, Zhu QH, Ramm K, Wu L, Eamens A, Sivakumar R, Kato T, Yun DW, Santhoshkumar C, Narayanan KK, Peacock JW, Dennis ES (2002) An i*Ac/Ds* gene and enhancer trapping system for insertional mutagenesis in rice. Funct Plant Biol 29: 547-559

Verwoerd T, Dekker B, Hoekema A (1989) A small-scale procedure for the rapid isolation of plant RNAs. Nucleic Acids Res 17: 2362

Wilson K, Long D, Swinburne J, Coupland G (1996) A *Dissociation* insertion causes a semidominant mutation that increases expression of *TINY*, an Arabidopsis gene related to APETALA2. Plant Cell 8: 659-671

Wisman E, Cardon Guillermo H, Fransz P, Saedler H (1998) The behaviour of the autonomous maize transposable element *En/Spm* in *Arabidopsis thaliana* allows efficient mutagenesis. Plant Mol Biol 37: 989-999

Yu J, Hu S, Wang J, Wong GK, Li S et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296: 79-92

Chapter 8

Influence of maize *En/Spm* transposase expression in rice on endogenous homologous transposons.

Raffaella Greco[1], Pieter B. F. Ouwerkerk[2] and Andy Pereira[1]

[1] *Plant Research International, Wageningen, The Netherlands*
[2] *Institute of Biology, Leiden University, Leiden, The Netherlands*

(Submitted for publication)

ABSTRACT

The maize *Enhancer/Suppressor-Mutator* (*En/Spm)* transposase was transformed into rice to study its expression and interaction with homologous rice transposons. A low copy sub-family of rice transposons termed *En/SpmB,* most similar to the maize *En/Spm* element containing TnpD and TnpA like transposase products, was identified from rice genome sequences. Rice subspecies *indica* cultivars seem to contain complete *En/SpmB* like elements at different chromosomal positions indicating recent transpositional activity. However, the subspecies *japonica* cultivar Nipponbare carries a single complete *En/SpmB* element that is mutated and includes another *En/Spm* related transposon *hipa* in the transposase gene. Alternative transcripts produced from the *En/SpmB* element are formed by either termination before or read-through the *hipa* insertion site. The longer transcript reading over the *hipa* insertion site, is not due to transposon excision, but seems to be due to precise splicing of the *hipa* insert using non-canonical splice sites. This splice product is however 'suppressed' in Nipponbare genotypes expressing the maize *En/Spm* transposase, probably due to binding of the maize transposase products to the *hipa* terminal repeats and inhibiting transcription read-through. These experiments reveal an interaction between heterologous and endogenous *En/Spm* elements in rice that support their conserved role as 'controlling elements' in plant genomes.

Keywords*: Transposon, Rice, Suppressor, En/Spm, hipa*

INTRODUCTION

Plant genomes like those of cereals comprise many transposon families contributing much to their expansion. This is especially true for the maize genome with more than 70% retrotransposons, considered as the major candidates for the increase in genome size (San Miguel et al. 1996, 1998; Bennetzen 2000), or cut-and-paste type transposons. This latter class is particularly very mobile in maize, contributing to their discovery like in the form of the *Ac*/*Ds* and *En/Spm* transposon families (McClintock 1951, 1954; Peterson 1953).

Transposable elements (TE) constitute a large portion of the rice genome as well, accounting for about 25% of the genome (Goff et al. 2002; Yu et al. 2002). The majority are Class I Long Terminal Repeat (LTR) retrotransposons that comprise 70% of the transposable element sequences. Miniature Inverted-repeat Transposable Elements (MITEs) are about 16%, while about 14% of all the TE consist of class II cut-and-paste transposons. Among the latter, the *En/Spm* family seems to be the most represented (Goff et al. 2002). Members of the *Tos* family of LTR retrotransposons were originally the only endogenous elements demonstrated to be active in rice, which has now led to the generation of extensive collections of insertional mutants produced through tissue culture (Hirochika et al. 1996, 2001). Recently, also the activity of a class of MITE elements (*mPing* and *Pong*) was reported in cell cultures of *indica* (Jiang et al. 2003) and in anther cultures and intact plants of *japonica* rice varieties (Kikuchi et al. 2003; Nakazaki et al. 2003). However, no evidence of activity for any members of the *En/Spm* family or for any other cut-and paste transposable element in the rice genome has been reported so far.

The *En/Spm* autonomous element was first isolated in maize (Pereira et al. 1985; Masson et al. 1987) and belongs to the superfamily of CACTA-like transposons. The maize *En/*Spm family comprises autonomous elements, encoding the two transposase proteins TnpA and TnpD necessary for transposition (Frey et al. 1990; Masson et al. 1991), as well as defective elements (*I*/*dSpm*), which have lost the capacity to encode for the transposases and are therefore unable to transpose in the absence of an autonomous partner (reviewed by Kunze and Weil 2002). Related sequences were found in several plant genomes other than maize and activity of some of them was demonstrated, as in the case of *Tam1* from Antirrhinum (Nacken et al. 1991), *Tpn1* from Japanese morning glory (Inagaki et al. 1994), *Psl* from Petunia (Snowden and Napoli, 1998) and *Cs1* from sorghum (Chopra et al. 1999). In general, *En*/*Spm* elements from different species share limited nucleotide homology, mainly restricted to the short Terminal Inverted Repeats (TIRs) and sub-terminal regions required in *cis* for transposition. In addition, the structure of the sub-terminal regions can be conserved rather than the sequence itself, as exemplified by the presence of direct and inverted repeats of different sequence stretches in related elements from different plants. Consequently, also the consensus binding-site for TnpA appears not to be

conserved and accordingly, TnpA-like transposases are generally very divergent as well. In case of autonomous elements indeed, only the gene coding for the TnpD transposase display a significant degree of conservation. The co-conservation of TnpD and the CACTA termini supports the hypothesis that these terminal repeats are substrates for the TnpD transposase protein (Frey et al. 1990). In a proposed model for *En/Spm* transposition, the TnpA transposase binds the sub-terminal repeats juxtaposing the transposon ends at the TIRs, while TnpD is supposed to stabilize the complex by binding to TnpA and to promote endonucleolitic cleavage at the CACTA sequences (Gierl et al. 1988; Trentmann et al. 1989; Frey et al. 1990; Raina et al. 1998). Binding of TnpA to the sub-terminal repeats can also suppress expression of maize genes containing *I/dSpm* insertions (McClintock 1954; Grant et al. 1990, 1993). This effect is thought to result from inhibition of transcription through the element that, in absence of TnpA, could otherwise be spliced out from a primary chimeric transcript, partially restoring gene expression.

En/*Spm*-like transposons in rice were first described by Motohashi et al. (1996) with the identification of a defective 1536 bp element (*Tnr3*) that was inserted in a p-SINE retrotransposon. *Tnr3*-related sequences, some having sizes compatible with encoding active transposases, were shown to be present in multiple copies in the genome of different rice cultivars, while absent in other cereals like maize, barley and wheat (Motohashi et al. 1996). Since then, several other defective and putative autonomous *En*/*Spm*-like elements have been identified in computational surveys of genomic rice sequences (Mao et al. 2000; Tarchini et al, 2000; Panaud et al. 2002; Wicker et al. 2003). Apart from the CACTA sequences in the TIRs, these elements often do not share nucleotide sequence homology with each other and therefore define different sub-families of *En*/*Spm* transposons in the rice genome.

The presence of several *En/Spm* related sequences in the rice genome suggests that endogenous elements may influence transposition of exogenous members of the same family, and vice versa. This is especially relevant when developing transposon tagging strategies using heterologous transposons. In our lab, transgenic rice lines were previously produced containing the maize *En/Spm* transposon system (R. Greco and A. Pereira, unpublished data). To explore the possibility of cross-talk between endogenous and exogenous elements, we compared transcription of *En/Spm* endogenous elements that may potentially retain activity in wild-type and transgenic rice plants containing the maize *En*/*Spm* transposon. This analysis could indeed reveal interaction with the maize transposase resulting in specific suppression of an endogenous transposon transcript. The possible mechanisms involved in this process are examined here.

MATERIALS AND METHODS

Plant material

Wild type *Oryza sativa* ssp. *japonica* cvs. Nipponbare, Eyi105, Taipei 309 and ssp. *indica* cvs. Pusa Basmati and A19 were obtained from Dr Paul Christou (Fraunhoefer Institute, Germany) or Dr Annemarie H. Meijer (Institute of Biology, Leiden University, The Netherlands). The transgenic plants containing the maize *En/Spm* transposon system were produced in Nipponbare (R. Greco and A. Pereira, unpublished data). Lines containing a two component enhancer trap construct with an immobilized *En/Spm* transposase and a mobile *I/dSpm* transposon, or only the immobilized transposase, were used.

RT-PCR and RACE analysis of En/SpmB

RT-PCR reactions were performed on single-strand cDNA synthesized from 500 ng of poly(A)$^+$ RNA with Superscript reverse transcriptase (Invitrogen), as described by Frohman et al. (1989). The *En/SpmB* specific primers used are listed in Table 2, and all the possible forward and reverse combinations were employed. The concentrations of the cDNA samples were equalized using actin primers. The "SMART RACE cDNA Amplification Kit" (Clontech) was used for the isolation of full-length cDNAs. The *En/SpmB* specific primers F and H (nested) or G and G2 (nested) were used for 3' and 5' RACE respectively, in combination with the adapter primers supplied with the kit. RT-PCR and RACE products were cloned in pGEM-T Easy (Promega) and sequenced. Multiple sequence alignments were performed using CLUSTAL X (Thompson et al. 1997) and manually edited using GENEDOC (Nicholas et al. 1997). Prediction of exon/intron boundaries was performed with GENSCAN (Burge and Karlin, 1997; http://genes.mit.edu/GENSCAN.html). Similarity searches in public databases were performed using the Blast algorithm (Altschul et al. 1997) at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) or at Gramene (http://www.gramene.org).

RT-PCR and excision analysis of hipa

Transcription of *hipa* was monitored using primers in the GENSCAN predicted ORF1 (position 1695-forward and 2758-reverse) and ORF2 (pos. 7502-forward and 8738-reverse). These primer combinations were expected to amplify products of 1064 bp and 1237 bp on genomic DNA or 929 bp and 1066 bp on cDNA if splicing occurred as predicted. Excision of the *hipa*-like transposon from *En/SpmB2* was monitored using primers L1 and M2 (position +358 in the flanking genomic sequence, corresponding to position 300,914 in AF172282), followed by a

nested PCR using primers L1 and M1 (position +172 or 301,100 in AF172282). Alternatively, primers L1+L2 followed by L1+L were used, expected to amplify both *En/SpmB2* and *En/SpmB3*.

Northern blot analysis

Total RNA was isolated from three weeks old tissue culture-grown plantlets as described by Verwoerd et al. (1989). About 6 μg of poly $(A)^+$ RNA purified using the "mRNA purification kit" (Pharmacia) was separated on a 1.5% formaldehyde agarose gel and blotted in 10X SSC onto Hybond-N^+ membrane (Amersham, UK). Pre-hybridization and hybridization were performed at 65°C in 10% Dextran Sulfate, 1% SDS, 1M NaCl and 0.4 mg/ml sonicated salmon sperm DNA. After hybridization, the membrane was washed twice in 1X SSC with 1% SDS at 65°C. Hybridization signals were quantified using a Bas2000 PhosphorImager (Fujix) and TINA software (Raytest, Germany).

Southern blot analysis

Genomic DNA isolations and DNA gel blot hybridizations were performed as previously described (Greco et al. 2001). About 3 μg of DNA from each plant were restricted separately with *Bam*HI, *Bgl*II, *Eco*RI, *Eco*RV and *Hin*dIII. The same blots were then hybridized independently with *En/SpmB* homologous probes amplified with primers D + E (probe 1) and H + I (probe 2).

RESULTS

Identification of putative autonomous En/Spm elements in the rice genome

To identify potential *En*/*Spm* autonomous elements in rice we performed translated blast searches against the GenBank database using the maize TnpA and TnpD transposase proteins as queries. Among the many rice genomic sequences encoding predicted peptides homologous to TnpD, the four displaying the highest similarities contained also a predicted gene downstream whose product was homologous to TnpA, though similarities in this case were considerably lower (Table 1). In one of these genomic clones (AF172282), the two tandemly arrayed predicted genes were previously annotated as part of a CACTA-like transposable element, namely *En*/*SpmB* (Tarchini et al. 2000), while no clear annotation existed for the other three clones. Manual inspection of the surrounding genomic context to identify the conserved CACTA terminal inverted repeats and the tri-nucleotide target site duplication could tentatively define putative autonomous elements of 9539 bp (AC713907), 9854 bp (AC116603) and 11315 bp (AC078829). As predicted,

nucleotide sequence alignment did not reveal significant homology among these elements nor with the maize *En*/*Spm*, apart from the termini and the TnpD open reading frame. Nevertheless, computational prediction of splicing sites suggested a similar complex organization as the maize *En*/*Spm*, exemplified by *En*/*SpmB* (Fig. 1), which seemed a good candidate to represent a cognate active autonomous element.

Table 1. Putative autonomous *En/Spm* elements in the rice genome.

GenBank Accession	Chr	Length (bp)	ORF1[a] (905 aa)	E-value	ORF2[a] (146 aa)	E-value	TnpA[a] (621 aa)	E-value
AF172282	11	8216	561/946 (59%)	0.0	28/51 (54%)	0.020	37/75 (49%)	0.060
AC078829	10	11315	377/609 (61%)	0.0	32/51 (62%)	1e-04	130/327 (39%)	3e-07
AL713907	12	9539	514/799 (64%)	0.0	38/68 (55%)	6e-05	27/56 (48%)	1.1
AC116603	10	9854	489/732 (66%)	0.0	32/51 (62%)	7e-06	80/204 (39%)	0.004

[a] Predicted peptide similarities with the transposase products of the maize *En*-1 (Pereira et al. 1986; GenBank accession M25427).

En/SpmB is a 8216 bp long transposon with imperfect Terminal Inverted Repeats (TIR) ending with the typical CACTA sequence and repetitive sub-terminal regions of about 400 bp in size. Only 9 of 13 nucleotides are conserved in the TIRs of *En/SpmB* compared to the maize *En/Spm*, with the first stretch of 6 nucleotides, including the CACTA sequence, being identical. As for other transposons of the same family, insertion of *En/SpmB* produces a 3 bp direct duplication of the target site. The presence of two predicted open reading frames coding for peptides with homology to the TnpA and TnpD transposases, demonstrated to be necessary for transposition of the maize *En/Spm* element (Frey et al. 1990; Masson et al. 1991), would classify *En/SpmB* as a potentially autonomous transposon. Imperfect repeats of a common motif (GTAGGGAAA) are present in several copies in direct and inverse orientation in the sub-terminal regions that, similarly to the maize *En/Spm* element, may represent the sites at which the TnpA-like transposase can bind to mediate the transposition process. However, these motifs do not share homology to the consensus binding site for the maize TnpA (Gierl et al. 1988), indicating that the elements belong to different subfamilies.

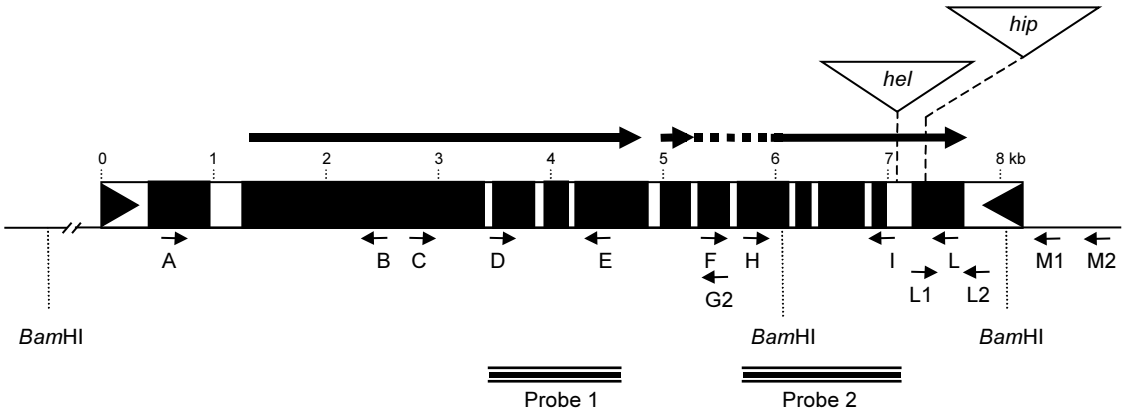


Figure 1. Diagram of *En/SpmB*. Position 1 and 8,216 of *En/SpmB* correspond to position 309,482 and 301,273 in the BAC clone AF172282. Black rectangle boxes and triangles represent exons and terminal regions, respectively. The *helitron* (hel) and *hipa*-like (hip) transposon insertions present in some *japonica* cultivars (*En/SpmB2*) are indicated as open triangles. Thick arrows indicate the predicted open reading frames. The positions of the primers used for transcription analyses are shown, along with the probes and the *Bam*HI restriction sites used for Northern and Southern analysis.

Transcription of the rice En/SpmB is influenced by the maize En/Spm transposon

Transcription of the *En*/*SpmB* transposable element was analyzed in wild-type and transgenic Nipponbare rice plants containing the maize *En/Spm* transposase (see Material and Methods). An indication that *En*/*SpmB* was transcribed emerged from RNA gel blot hybridization experiments using a probe corresponding to the 3' region of the element (probe 2; Fig. 1). As for the maize *En*/*Spm* (Pereira et al. 1986), two main hybridizing bands were revealed in all the plants (Fig. 2). The larger fragment (about 6 kb) could represent the TnpD homologous transcript, while the 2 kb band could correspond to the endogenous TnpA-like transposase. However, unlike in maize, the long form appeared far more abundant than the shorter form. In addition, an alternative transcript of about 2.3 kb was visible in the wild-type.

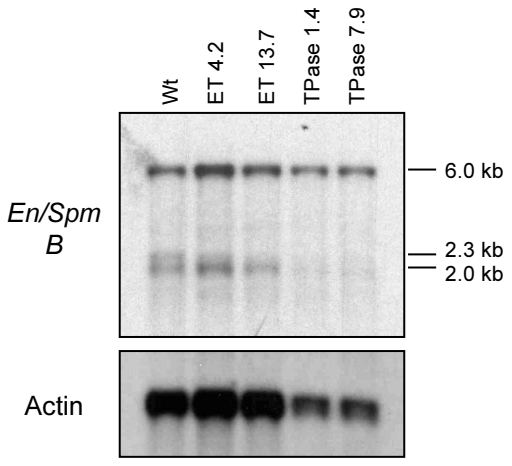


Figure 2. *En/SpmB* endogenous transcripts. Northern blot analysis was performed on poly $(A)^+$ RNA isolated from rice plantlets. Lane 1: wild-type rice. Lane 2, 3: rice *En/Spm* enhancer traps. Lane 4, 5: *En/Spm* transposase lines. To detect endogenous *En/SpmB* transcript, a 0.95 kb PCR-amplified fragment obtained using primers H and I was used as a probe (Fig. 1), corresponding to the 3' region of the putative transposases (TnpA or TnpD). As a loading control, the same blot was re-hybridized with an actin tobacco cDNA probe.

Splicing was further investigated by RT-PCR analysis using primers designed along the whole *En*/*SpmB* element, based on computational prediction of splicing sites on the genomic sequence (Fig. 1; Table 2). Sequencing of the amplified products revealed in total 12 exons and 11 introns, together with alternative splicing sites generating the multiple minor products (Table 2). From these experimentally defined splicing products two main ORFs could be identified. The first long ORF spanned 3441 bp from the beginning of exon 2 (position 1240) until the end of exon 5 (position 4899), encoding a 1147 aa putative peptide with homology to TnpD of maize. A second ORF of 1587 bp in length, starting at position 5032 in the sixth exon and ending in the last exon at position 7497, would encode a 529 aa peptide partially homologous to TnpA. However, a stop codon at position 5069 is present that interrupts the reading frame. A shorter peptide of 359 aa would be produced instead, from a 1077 bp ORF starting from a second ATG at position 5874 (exon 8).

Table 2. Exon/intron boundaries and primer positions in *En/SpmB*.

Exon	Begin	End	Primer	Position	Direction[a]
1st	418[b]	893	A	512	F
2nd	1213[b]	3455	B, C	2110, 2925	R, F
3rd	3534	3842	D	3557	F
4th	3915	4035			
5th	4113	4948[c]	E	4428	R
6th	5031	5228			
7th	5310	5505	F, G2, G	5346, 5363, 5500	F, R, R
8th	5757	6128	H	5811	F
9th	6204	6318			
10th	6392	6730			
11th	6819	6932[d]	I	6856	R
12th	7243	7497	L1, L	7249, 7378	F, R
3' UTR	7498	7739	L2	7725	R

[a] Forward (F) or reverse (R).

[b] Predicted sites not confirmed experimentally.

[c] Splicing of this exon can occur also using alternative donor sites at positions 4207 and 4678.

[d] In most transcripts, this donor site is skipped and transcription proceeds until the alternative polyadenylation site at position 7018.

Remarkably, amplification using a reverse primer in the last exon (L) in combination with the forward primers F or H allowed the identification of the predicted products (1.2 and 1 kb respectively) only in the wild-type and not in the transgenic rice lines containing the maize *En*/*Spm* transposon (Fig. 3A). When primer I was used in place of primer L, products of the expected size could be detected in all the different plants. Moreover, the products amplified with

primer L were less abundant than the products obtained with primer I, although the similar sizes should give equivalent amplification efficiency.

Accordingly, 3' RACE experiments carried out with primer H and an oligo-d(T) adapter primer revealed an ubiquitous 1 kb main fragment and a less abundant 1.4 kb form selectively present in the wild-type (Fig. 3B). In contrast a smaller 0.5 kb fragment seems to be present in a higher amount in the transgenic *En/Spm* lines that turned out to be the result of aberrant splicing. Sequencing of the RACE products suggested that the 1 kb form originates by a premature polyadenylation within the last intron, at position 7018. As a result of transcription through the last intron, the ORF terminates with the stop codon at position 6971 (458 or 288 aa peptide product) and thus the last exon, containing the sequence homologous to primer L, is missing in this transcription product. On the other hand, the 1.4 kb form contained all the expected exons and was polyadenylated, as predicted, at position 7739. This form likely corresponds to the 2.3 kb transcript visible in the Northern analysis (Fig. 2). The presence of this larger form only in the wild-type would explain the selective amplification with primer L and suggests an influence of the exogenous maize transposon suppressing transcripts of the endogenous *En/SpmB* element. The question then arose whether this influence was mediated through alternative splicing (polyadenylation) of the same endogenous element or on an additional copy of a related element in the genome.

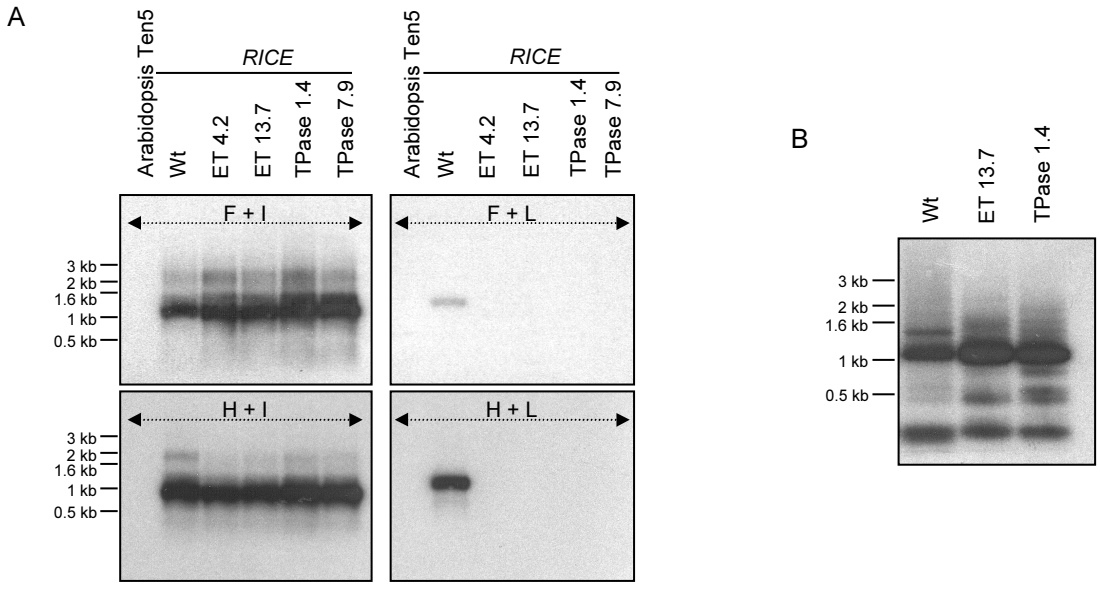


Figure 3. RT-PCR and RACE analysis of *En/SpmB* transcripts. A) Amplified fragments from wild-type and transgenic rice cDNAs using *En/SpmB* specific primers F or H in combination with I or L were fractionated on an agarose gel and hybridized with probe 2 (Fig. 1). As a negative control, cDNA prepared from an Arabidopsis *En/Spm* transposase line was used (Ten5; Aarts et al. 1995). B) *En/SpmB* specific transcripts revealed after hybridization with the same probe, using 3'RACE reaction products obtained with primer H and a 3'-adapter primer on wild-type and transgenic rice cDNAs.

All the primer combinations used in RT-PCR were also tested on genomic DNA. Surprisingly, PCR amplification of *En/SpmB* carried out using primer L in combination with F or H did not give any product, neither in the wild-type nor in the transgenics, although the expected fragments were of a size compatible with the Taq polymerase amplification conditions. The same primers F and H in combination with I did amplify a clear band in all samples (data not shown). No amplification was detected as well when *japonica* varieties other than Nipponbare were used (Eyi 105 and Taipei 309; data not shown). Interestingly, the genomic sequence of *En*/*SpmB* deposited in the GenBank database and used for design of the primers was originally identified in the double haploid line YT14, derived from a cross between the two *japonica* cultivars Yashiro-mochi and Tsuyuake (Tarchini et al. 2000). This would suggest that some kind of modifications occurred independently in the *En/SpmB* elements in the *japonica* genotypes employed in our analysis.

En/SpmB in some japonica cultivars contains two transposon insertions

Since the transcription analysis was performed in Nipponbare, we examined the genomic structure of *En/SpmB* and the presence of additional elements sharing nucleotide sequence similarity in the genome of this rice cultivar. The results of the Blast searches in public databases (updated July 2003), indicated the presence of at least two *En*/*SpmB* related elements in the Nipponbare genome, identified by the genomic clones reported in Figure 4 (for simplicity of description, the two elements are named *En*/*SpmB2* and *En*/*SpmB3*, to distinguish them from the original *En*/*SpmB* element).

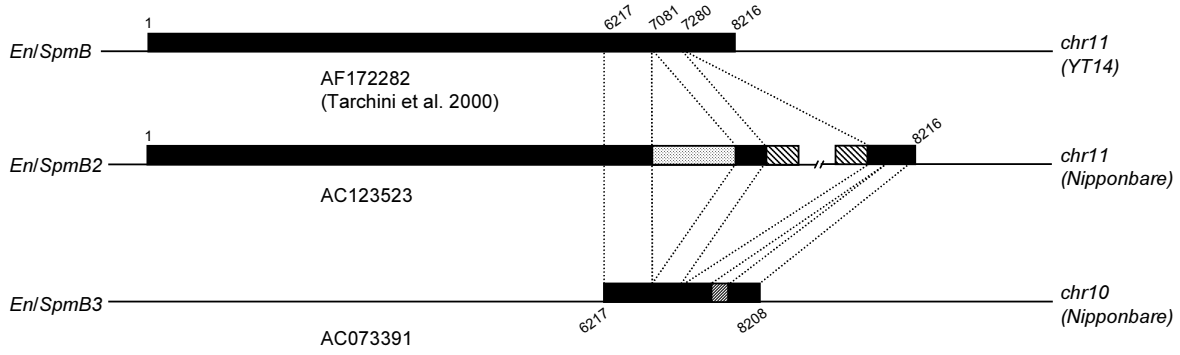


Figure 4. *En/SpmB* homologous sequences in the rice Nipponbare genome. Genomic clones displaying partial/total homology with *En/SpmB* (Tarchini et al. 2000). *En/SpmB2*: GenBank clone AC123523 on chromosome 11, homologous to *En/SpmB* from position 97,146 (pos. 8216 of *En/SpmB*) till 108,538 (pos. 1 of *En/SpmB*). Homology extends also to the flanking *adh1-adh2* region. The dotted-box represents the *helitron* insertion and the striped-boxes the borders of the *hipa*-like insertion. *En/SpmB3*: GenBank clone AC073391 on chromosome 10, homologous to *En*/*SpmB* from position 38,116 (pos. 6217 of *En*/*SpmB*) till 40241 (pos. 8208 of *En*/*SpmB*).

En/*SpmB2* is identical to *En*/*SpmB* except for the presence of two insertions. The first insertion of 1280 bp is located in the seventh intron, at position 7081 of *En*/*SpmB*, after the alternative polyadenylation signal at position 7018 as revealed by our transcription analysis. The inserted DNA shares homology to an intron sequence present in several rice LRR-type disease resistance proteins. However, a more detailed inspection identified this insertion as a potential defective *helitron*-like transposon (Table 3). *Helitrons* constitute a new class of DNA transposable elements recently identified in eukaryotic genomes including plants (Kapitonov and Jurka, 2001; Lal et al, 2003), which transpose via a rolling-circle mechanism similar to bacterial Insertion Sequences, do not produce target site duplication upon insertion and lack terminal inverted repeats. Instead, they insert specifically into an AT target site and have a conserved 5' TC and a 3' CTRR, preceded by a 15-20 bp palindrome separated by 10-12 nucleotide from the 3' end (Kapitonov and Jurka, 2001). Autonomous *helitrons* are supposed to encode a DNA helicase and a nuclease/ligase (Kapitonov and Jurka, 2001), though a presumed recently active autonomous element from maize codes for an RNA helicase (Lal et al, 2003). So far, only three families of putative autonomous and defective *helitrons* were identified in the rice genome by computational prediction (Kapitonov and Jurka, 2001) not sharing homology with the *helitron* in *En/SpmB*, which may therefore identify a different family.

In addition, *En*/*SpmB2* carries a CACTA-like transposable element inserted at position 7280 in the last exon, with a typical trinucleotide target site duplication (Table 3). Sequence comparisons revealed homology to *hipa* (Panaud et al. 2002), a rice *En*/*Spm*-like transposon with characteristic conserved CACTG termini. The size of the *hipa*-like element in *En/SpmB2* appears to be around 1.8 kb from the GenBank clone AC123523. However, the existence of an unresolved gap between the assembled contigs in this region makes this estimate not precise. Attempts to isolate the full-length insertion sequence by PCR on genomic DNA proved unsuccessful, probably due to the formation of secondary structures inhibiting amplification. From similarity searches for homologous sequences in the rice genome and Southern analysis (see next section), it seems to be a defective element of about 1.3 kb in size.

Table 3. Transposon insertions in *En/SpmB* in some *japonica* cultivars[a].

Helitron-like insertion (pos. 7081)
tcatatttatca**a** TCTATCTATCTA…CGATACATGCCCGCGTATACGCGCGGGCTACCTTTCTAG **t**tgatactaatag
Hipa-like insertion (pos. 7280)
aga CACTGGTAGAGAAACCATCTTTGGTCGGTC…CCGGGACTAAAAAGCATTTCTCCACCAGTG **aga**

[a] Transposon sequences are represented in capital letters, with the terminal conserved sequences (TIRs in case of *hipa*) double-underlined. The palindrome in the 3' terminal region of the *helitron* is single-underlined. Boldface lower case letters in the flanking genomic sequences represent the target insertion site (duplicated in case of *hipa*).

As *En*/*SpmB2* is inserted at exactly the same position in the *adh1-adh2* locus of Nipponbare chromosome 11, as the original *En/SpmB* in YT14 (Tarchini et a. 2000), we assume that we are dealing with the same transposon element. We therefore hypothesized that the *helitron* and *hipa* insertions described above occurred independently in *En*/*SpmB2* in the Nipponbare genome. The presence of these two insertions would explain the absence of amplification when using primer L on genomic DNA, due to the increased size of the expected product and the complex repetitive structure of the *hipa*-like transposon. Moreover, as the *japonica* cultivars Taipei 309 and Eyi 105 display the same amplification pattern, the insertion events are present in these *japonica* genotypes as well (see next section).

The second element (*En/SpmB3*) is present on chromosome 10 and contained only the terminal part of *En*/*SpmB*, from position 6217 till 8208 (Fig. 4). Besides the different genomic location, sequence divergence clearly defined this truncated element as independent.

The presence and structure of *En*/*SpmB* homologous sequences in the *japonica* genome was also monitored by DNA gel blot analysis using five different restriction enzymes (see Material and methods). Besides Nipponbare, the two other *japonica* rice cultivars tested in PCR analysis (Eyi 105, Taipei 309) were analyzed. Hybridization was carried out with probes corresponding to part of the ORF of the TnpD-like protein (probe 1; Fig. 5A) or to the 3' region of *En*/*SpmB* (probe 2; Fig. 5A).

The results obtained confirmed the presence of only two *En*/*SpmB* related copies, same in all the three *japonica* varieties, one complete copy at the *adh1-adh2* locus on chromosome 11 bearing the *helitron* and *hipa* insertions (*En*/*SpmB2*) and a truncated copy (*En*/*SpmB3*) on chromosome 10. As a representative example, Southern analysis of genomic DNA restricted with *Bam*HI is shown in Figure 5B. An intact *En*/*SpmB* element at the *adh1-adh2* locus should display a single band of 8.6 kb or two bands of 8.6 and 1.9 kb upon hybridization with probe 1 and 2, respectively. If the *helitron* and *hipa* insertions are present (*En*/*SpmB2*), a larger band estimated to be about 4,5-5 kb in size (depending on the *hipa*-like size) should be revealed by probe 2, in addition to the 8.6 kb, while the hybridization pattern with probe 1 would remain unaltered. If the truncated *En*/*SpmB3* copy is present, no hybridizing fragment would be expected using probe 1, while probe 2 would detect a 7.3 kb band. As expected for the absence of intact *En/SpmB*, a 8.6 kb hybridizing fragment was revealed with probe 1, while three fragments of 8.6, 7.3 and 4.5 kb hybridized with probe 2. The hybridization patterns obtained with the other restriction enzymes used were in agreement with these results (data not shown). In general, the lower intensity of some of the hybridizing fragments (i.e. the 7.3 kb) can be explained by the poorer homology to the probe used. From the size of the 4.5 kb hybridizing fragment, we estimated the *hipa* insertion to be around 1.3 kb.

In conclusion, at least two *En/SpmB*-containing *adh1-adh2* alleles exist in *japonica* cultivars. One allele contains the intact *En/SpmB* element, as seen in the YT14 cultivar, while the other allele bears the *En/SpmB2* element in cultivars such as Nipponbare, Eyi 105 and Taipei 309.

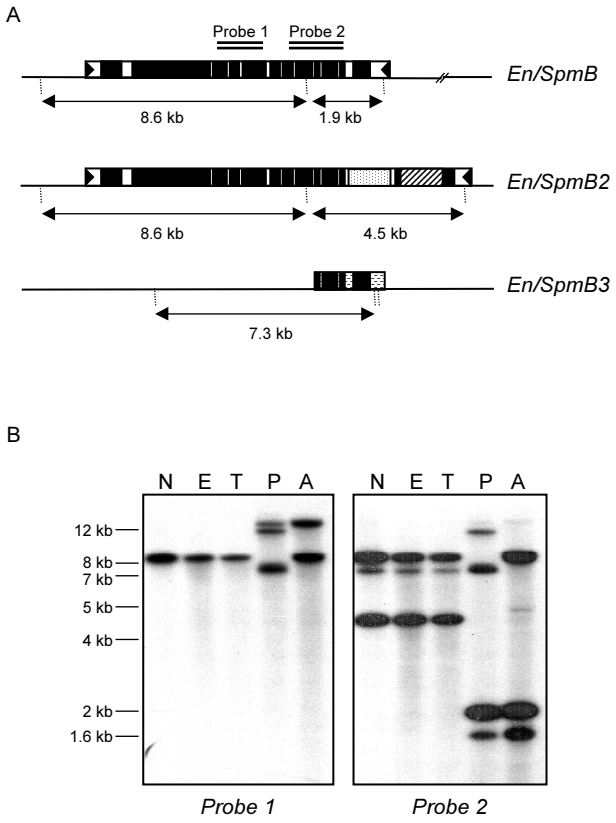


Figure 5. Southern blot hybridization of wild-type rice cultivars. A) Diagram of the hybridization patterns expected for the different *En/SpmB*-like elements in Nipponbare when genomic DNA is digested with *Bam*HI and blots hybridized with a probe homologous to the 5' (probe 1) or to the 3' region of *En/SpmB* (probe 2). Expected *Bam*HI restriction sites, shown as small vertical lines below, in *En/SpmB* (AF172282) and *En/SpmB2* (AC123523) are at positions -2540, 6093, 8038. As the complete sequence of *hipa* is not available, predicted presence/absence of a site could only be confirmed upon hybridization. In *En/SpmB3* (AC0733991), *Bam*HI is expected to cut at positions 32686, 40006, and 40071. B) Southern blot containing three *japonica* (N=Nipponbare, E=Eyi105, T=Taipei309) and two *indica* (A=A19, P=Pusa Basmati) varieties digested with *Bam*HI and hybridized with the above-mentioned probes.

Alternative transcripts of En/SpmB in Nipponbare due to helitron and hipa insertions

In view of the evidence on the structure of the *En/SpmB* transposons in the Nipponbare genome, it would seem that the two transcript forms detected in RT-PCR and RACE analysis originated from alternative splicing of the same element rather than from different elements. Moreover, the presence of the two transposon insertions may influence the process by favoring the premature termination of transcription with prevalent formation of the shorter form. On the other hand, as the longer transcript was shown to not contain the *helitron* and *hipa*-like insertions, it means that these sequences need to be removed to obtain this final product. Since the *helitron* is inserted in the last intron, it would seem reasonable to assume its removal as a consequence of a splicing event.

The *hipa*-like transposon, however, is inserted in the last exon and could be eliminated as a result of a transposition event. The estimated size of this element suggests it is defective, therefore unable to transpose by itself. Nevertheless, the *hipa* sequence described by Panaud et

al. (2002) has a size compatible with it being an autonomous element (10,539 bp) and might be active to promote transposition of this defective partner. This hypothesis is supported by the sequence identity revealed in the TIRs (14/16 bp) and the sub-terminal regions (mainly at the level of the 16 bp repetitive motifs) by the two elements, indicating that they belong to the same family (Fig. 6A).

Hipa was demonstrated to be widespread among several species of the genus *Oryza* (Panaud et al. 2002). Similarity searches in public databases with the 10,539 bp-element and the 1.3 kb-element revealed the occurrence of several *hipa*-homologs in the Nipponbare genome, in contrast to the low copy number of *En/SpmB*. GENSCAN prediction of splicing sites on the 10,539 bp genomic sequence (Fig. 6B) detected two ORFs coding for putative peptides of 1221 and 1022 amino acids, with homology to TnpD-like transposases (ORF1) or to putative transposases and hydroxyproline-rich glycoproteins (ORF2). These two genes appeared to be arranged in tandem in most other related elements in the Nipponbare genome.

As a preliminary indication of activity, we assessed whether this *hipa* element was transcribed. We therefore designed primer pairs complementary to the first and second predicted ORFs and tested them in RT-PCR reactions on the same cDNAs used before. Transcripts could be detected in both cases, although at low amount and more abundant for the first ORF (Fig. 6C). Since products of equal size were amplified from cDNA and genomic DNA, this would indicate that the splicing did not occur as predicted.

To monitor the occurrence of excision of the *hipa*-like transposon from *En/SpmB2*, we tested several primer combinations around the *hipa* insertion site at position 7082 of *En/SpmB* in the attempt to amplify an empty donor site (see Material and Methods; Table 2). When using primers L1 and L or L2, faint bands of 130 or 480 bp could be amplified that upon sequencing were confirmed to be derived from amplification of *En/SpmB3*. Additional primers in the flanking *adh1-adh2* genomic region were therefore employed to specifically target *En/SpmB2*. Expecting the occurrence of excision events to be rare, nested PCR reactions were also performed and genomic DNA from wild-type and transgenic rice containing the maize *En/Spm* was used. DNA was isolated from rice seedlings or developing calli, since mobility of transposable elements could be triggered by stress conditions such as tissue culture, as demonstrated for other transposable elements like *Ac/Ds* and retrotransposons (e.g. in rice Hirochika 1996; Izawa et al. 1997; Ki et al. 2002). However, we were unable to detect any excision products, nor any other PCR products containing the *hipa*-like insertion still residing at the donor place with any of the appropriate primer combinations used.

These results suggest that elimination of the *hipa*-like transposon must involve a different mechanism than excision at the DNA level. Moreover, the expression of the maize *En/Spm*

transposase may suppress the formation of the *En/SpmB2* long transcript product made by removal of the *hipa*-like element.

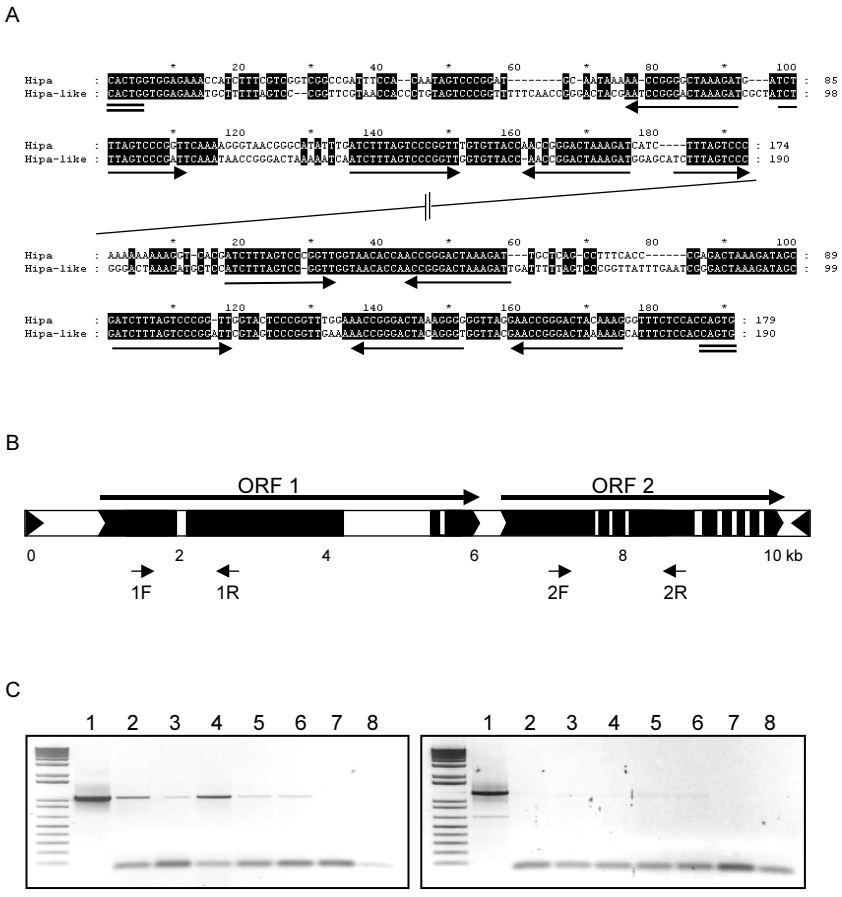


Figure 6. Structure and transcription analysis of the *En/Spm*-related *hipa* element. A) Sequence alignment between the sub-terminal regions of *hipa* (Panaud et al. 2002; Genbank accession AP002484) and the *hipa*-like insertion in *En/SpmB2*. B) GENSCAN splicing prediction of *hipa* (position 1 and 10,539 correspond to position 15,798 and 26,337 in the BAC clone AP002484). Initial, internal and terminal exons are represented as tailed block arrows, boxes or headed block arrows. C) RT-PCR amplification of *hipa* using primers in ORF1 and ORF2. Lanes 1: wild-type Nipponbare control genomic DNA; lanes 2-6: cDNA from wild-type and four *En*/*Spm* transgenic Nipponbare lines (enhancer trap lines 4.2 and 13.7, transposase lines 1.4 and 7.9); lane 7: cDNA from transgenic *En*/*Spm* Arabidopsis (Aarts et al. 1995); lane 8: no DNA.

In some indica genotypes En/SpmB does not contain the nonsense mutation or the transposon insertions and reveals recent activity

Database searches for *En*/*SpmB* homologous sequences in the draft genome sequence of the 93-11 *indica* cultivar at the Beijing Genomics Institute (http://btn.genomics.org.cn/rice) revealed the presence of at least 2-3 intact copies of the element, none of them containing the

helitron and *hipa*-like insertions (contigs 24542-15451, 7968-22749 and 15165). PCR amplification using primer L of genomic DNA from two other *indica* varieties A19 and Pusa Basmati revealed a clear product for all *En/SpmB* primer combinations (data not shown), indicating this as a common feature of *indica* genotypes.

Moreover, the three *En/SpmB* copies in 93-11 are inserted at three loci distinct from the *adh1-adh2* locus on chromosome 11. Interestingly, no insertion is revealed at this locus (contig 30106). The absence of a footprint or any other molecular alteration at the target site suggests that *En*/*SpmB* did not insert at this position in this cultivar. However, PCR amplification on A19 and Pusa Basmati genomic DNA using primers in the last exon of *En/SpmB* and in the flanking *adh1-adh2* region revealed an amplification product in case of A19 (data not shown). This shows that the *indica* cultivar A19 possesses an *En/SpmB*-containing *adh1-adh2* allele, such as found for the *japonica* cultivars. The actual insertions sites of the three *En/SpmB* copies identified in 93-11 could however not be mapped on the rice genome, as Blast searches against public databases did not give significant results.

Southern analysis carried out on A19 and Pusa Basmati revealed internal fragments of the expected size (i.e. 1.9 kb) predicted for intact copies of *En*/*SpmB*, confirming the absence of the two insertions also in these two *indica* varieties, as for 93-11 (Fig. 5). In particular, the pattern observed would suggest the presence of three *En*/*SpmB* copies in Pusa Basmati and two in A19. Furthermore, A19 seems indeed to possess an *En/SpmB* element inserted at the *adh1-adh2* locus like in the *japonica* cultivars, as revealed by the presence of the same 8.6 kb fragment originated by a *Bam*HI restriction site in the flanking genomic sequence. The different patterns displayed by the two *indica* genotypes would point to a recent activity of these elements, in contrast to the apparent stability revealed by the invariable *japonica* hybridization profiles.

Interestingly, analysis of the genomic sequences of two *En/SpmB* copies in the *indica* cultivar 93-11 revealed the absence of the stop codon at position 5069 in the sixth exon. As described earlier, this nonsense mutation causes an interruption in the ORF coding for the putative TnpA transposase, with a premature termination of the synthesis of the peptide after a few amino acids from the translation start. As TnpA was demonstrated to be necessary for transposition activity of *En/Spm* in maize, together with TnpD (Frey et al. 1990; Masson et al. 1991), the absence of this TnpA-like protein may presumably cause lack of mobility of *En/SpmB*. In agreement with the presumed inactivity, this mutation exists in the *japonica* YT14 and was demonstrated to be present in Nipponbare (see previous section), as well as in Taipei 309 and Eyi 105 (data not shown). However, sequencing of *En/SpmB* genomic amplification products in *indica* A19 and Pusa Basmati displayed absence of the stop codon only in the latter (data not shown), showing polymorphism for the active transposase allele within *indica*.

DISCUSSION

En/Spm homologous elements in rice

There exists a paradox on how so many transposable elements are present in a genome, and yet so few are active. In the larger genome of maize, cut and paste transposons are very active, while in the smaller genome of rice it seems not to be so. This would at first suggest that a selection pressure is there to stabilize transposons and maintain a smaller genome. We were therefore interested on the interaction between active transposons introduced from maize and the resident homologous elements in rice.

Our initial experiments began with the introduction of the maize *En/Spm* transposon system into rice for transposon tagging and the observation of a lower than expected mobility of the transposon system in this cereal host compared to that in the dicot host Arabidopsis (R. Greco and A. Pereira, unpublished data). At first we addressed the role of an incorrect processing of the maize element in rice or a sub-optimal choice of the transposon components used in the constructs. However, there remained the possibility that potentially active endogenous transposons could influence mobility of similar heterologous elements. We therefore searched the rice genome that was being sequenced for potential *En/Spm* homologous candidates to test the inter-element interactions.

Many CACTA-like transposable elements have been identified in the rice genome but none of them was so far demonstrated to be active. Detection of *En/Spm*-like transposons is often hampered by low sequence conservation. A recent computational approach combining a search for sequence and structural conservation in the terminal regions proved to be quite efficient in the identification of several autonomous and defective *En/Spm* homologous elements in the Triticeae genomes (Wicker et al. 2003). Alternatively, detection of autonomous elements in different species can be performed exploiting the homology throughout the conserved TnpD transposase protein (e.g. Staginnus et al. 2001; Langdon et al. 2003). Even the low conserved TnpA sequence could detect *En/Spm* related elements in rice (Wicker et al. 2003).

In our search for autonomous elements in the sequenced rice genome, we found only four putative genomic regions encoding both the transposase products, which in analogy with maize are supposed to identify them as a class of *En/Spm* homologous elements. We could thus focus on a set of few defined *En/Spm* homologous copies in rice and draw our conclusions based on them. However, several other autonomous elements might be present in the rice genome that we missed, due to the incompleteness of the database, annotation flaws or the fact that transposase proteins with the same function may have lost homology during evolution.

En/SpmB has lost transpositional activity in japonica while still mobile in indica lineages

Active endogenous autonomous elements might be expected to influence the transposition behavior of the maize transposon and vice versa. We found a few putative *En/Spm*-like elements and tested one that seemed to fit the criteria of a homologous autonomous element. *En/SpmB* turned out to be present at a very low copy number in all the rice cultivars tested, with 2-3 copies per genome, in sharp contrast to the high abundance of the related *hipa*-like transposable elements that show little homology beyond the TIRs with the maize element (Panaud et al. 2002). Indeed, copy number of CACTA-like transposons can vary significantly among plant species, ranging from less than five as in the *CAC1* in Arabidopsis (Miura et al. 2001) to several thousands in grasses with large genomes, such as Lolium, barley (Langdon et al. 2003) and wheat (Wicker et al. 2003). Remarkably in rice, it seems that variation in copy number of *En/Spm*-like elements occur among different sub-families within the same species as well. It was intriguing to discover what factors determined whether a certain element would be successfully propagated and maintained in a given host.

Remarkably, polymorphism was observed between *indica* and *japonica* varieties in terms of transposition activity of *En/SpmB* and nucleotide genomic sequence, which raised some interesting considerations about the evolutionary history of this transposable element. *Indica* genotypes contain mainly 2-3 copies of what we could define as an "intact" *En/SpmB* transposon, encoding two putative transposase products of 1147 and 529 aa with homology to the maize TnpD and TnpA, respectively. The different hybridization patterns displayed by the *indica* varieties in Southern analysis would suggest recent mobility of these copies in the *indica* genomes. Conversely, the *japonica* cultivars Nipponbare, Eyi 105 and Taipei 309 seem to contain 2 distinct and "altered" copies of *En/SpmB*, both apparently inactive. The truncated copy on chromosome 10 (*En/SpmB3*) is expected to be immobile, since it lacks one of the termini. The other copy residing at the *adh1-adh2* locus on chromosome 11 (*En/SpmB2*) contains a nonsense mutation and two transposon insertions in the ORF of the TnpA-like transposase, potentially affecting the production of a functional protein and therefore transposition ability. Accordingly, all the three *japonica* varieties displayed the same hybridization pattern in Southern experiments, suggesting that both the *En/SpmB* copies were already present in the progenitor almost certainly in an inactive form, as they were inherited as such during evolution.

Interestingly, one *indica* variety (A19) also contained a copy of *En/SpmB* inserted at the same *adh1-adh2* locus of chromosome 11, in addition to other intact copies. Moreover, insertion of *En/SpmB* at this locus always correlates with the presence of the nonsense mutation in the ORF of TnpA, strongly suggesting that this mutation is responsible for the loss of transpositional activity. According to this observation, a plausible hypothesis would imply the presence of an ancient *En/SpmB* insertion at this allele in the common progenitor of the *indica* and *japonica*

subspecies, consistent with their evolution. After the insertion, yet before the divergence between *indica* and *japonica* subspecies, a point mutation creating a termination codon in the ORF of TnpA might have occurred, making this element immobile. However, the two other *indica* varieties employed in the analysis (Pusa Basmati and 93-11) lack this mutated *En/SpmB* copy at the *adh1-adh2* locus. This would mean that either *En/SpmB* excised from this locus before it became immobile, or it did not insert in the first place. In the first hypothesis, the absence of molecular footprints at the site of insertion in these two *indica* genotypes would imply preferential occurrence of perfect excisions. This phenomenon has been demonstrated to occur occasionally in case of excision of *Ds* and *Mutator* in maize and *dTph1* in petunia (e.g. Dennis et al. 1986; Britt and Walbot 1991; Doseff et al. 1991; Chen et al. 1992; Scott et al. 1996; Koes et al. 1999), apparently via a separate pathway involving a novel yet undisclosed mechanism. However, precise excisions in plants are known to be exceptional events compared to footprint-leaving excisions (Saedler and Nevers, 1985; Coen et al. 1989) and therefore unlikely to happen in more than one genotype. Evidence would thus support the hypothesis that *En/SpmB* never visited the *adh1-adh2* locus in Pusa Basmati and 93-11. In this case, we could envisage two alleles in the common progenitor, differing in the presence of the *En/SpmB* insertion at the *adh1-adh2* locus. The evolution of *japonica* might have resulted in the selection of the immobile *En/SpmB* allele, while the *indica* lineages inherited both alleles.

On the other hand, the insertion of the *helitron* and *hipa*-like transposons in the *En/SpmB* copy at the *adh1-adh2* locus of chromosome 11 seem to have taken place exclusively in *japonica* genomes. Their absence in the double haploid line YT14 indicate they may have been restricted to some lineages, though a more thorough analysis of additional *japonica* genotypes would be required to assess their incidence and relative timing. Undoubtedly, both insertion events must have occurred after the divergence from the *indica* subspecies and subsequently to the nonsense mutation, as the two traits always co-segregate. If we assume that immobility arose as a consequence of the nonsense mutation, the two insertions probably accumulated in the transposon that had lost transpositional activity and got fixed.

The maize En/Spm transposase suppresses transcription reading through the hipa insertion in En/SpmB

In spite of its apparent immobility and the presence of the two transposon insertions, *En/SpmB* still happens to be transcribed, at least in the Nipponbare genome. The two insertions appear though to affect the splicing, promoting alternative transcript processing that seems to be suppressed by the presence of the maize *En/Spm* transposase. However, the mechanisms

leading to the production of the two alternative transcripts have still to be clarified, and so also the dynamics of the interaction with the exogenous maize transposase source.

Formation of the long transcript

The preliminary hypothesis that the two transcripts could be originating from two different copies of *En/SpmB*, one with the insertions and the other without, could be ruled out from the analysis of the genomic clones in the database and from Southern hybridization. Therefore, they must be formed as alternative products from the same template.

The presence of the *helitron*-like insertion seems to prevent splicing of the last intron in the majority of the cases, inducing premature polyadenylation before the site of insertion itself and resulting in the formation of the short transcript. Examples of insertions in transcribed regions influencing pre-mRNA processing were previously reported. Among them, the presence of a *helitron* insertion in the eleventh intron of the *shrunken2* gene of maize was shown to prevent splicing of the ninth intron and transcription to proceed after the insertion itself (Lal et al. 2003). Similarly, retrotransposons insertions in introns of three different alleles of the *waxy* gene could promote exon skipping upstream or downstream the site of insertion (Varagona et al. 1992).

In some cases though, splicing can occur leading to the removal of the *helitron* and transcription through the last exon. Concurrently, the *hipa*-like insertion has to be removed, since it is also not present in the long transcript form produced. The possibility that the *hipa*-less transcript could be generated during reverse transcription as a result of template switching triggered by the presence of the direct repeats of the target site duplication (Zaphiropoulos 2002 and references therein) could be ruled out, as its presence was confirmed by Northern analysis.

The most obvious mechanism to eliminate the *hipa*-like transposon would then be excision promoted by a trans-activating transposase from an autonomous partner. The low abundance of the long transcript seems indeed to support this hypothesis. However, attempts to demonstrate the occurrence of excision proved unsuccessful, even when appropriate conditions to reveal rare events were used. Whether the absence of excision is due to lack of mobility of the *hipa*-like transposon or to the absence of active transposase sources in the genome is not known. Certain is that, although several putative autonomous *hipa* elements in the rice genome have been revealed that are potentially transcribed (Panaud et al. 2002; this work), compelling evidence of their transpositional activity still has to be supplied. Moreover, the absence of a footprint at the site of *hipa* insertion in the long transcript would require the occurrence of perfect excision events that, as previously mentioned, rarely happen in plants.

Alternatively, removal of the *hipa*-like transposon could be a phenomenon occurring at RNA level rather than DNA, not involving transposition. In this case, transcription read through the

hipa-element occurs first and then the element is removed. Splicing of transposable elements from pre-mRNA has been previously described for defective members of the *Ac/Ds* and *En/Spm* families in maize, suggesting that they may function as novel introns (reviewed by Wessler 1989; Weil and Wessler 1990; Menssen et al. 1990). Most relevant in many cases, splicing was suppressed by the presence of the autonomous partner, presumably as a consequence of the binding of the trans-acting transposases to the termini of the defective element. In all cases, canonical donor and acceptor sites from the host DNA or (cryptic sites) from the element itself were employed, which never led to the complete deletion of transposon sequences. While elimination of the defective *hipa*-like element seems indeed inhibited by the maize transposase, the process leads to the perfect restoration of the original donor site. A very similar situation was reported by Giroux et al. (1994) in which the exact removal of a *Ds* transposon and one copy of the 8-nucleotide target site duplication occurred occasionally (less than 5% frequency) as a result of RNA splicing from an exon of the maize *sh2-m1* allele. Also in this case however, the random incidence of consensus GT/AG splicing sites (Goodall and Filipowicz 1991) within the repeats of the target site duplication could allow the alternative splicing. In this respect, elimination of the defective *hipa*-like element does not resemble a “true” splicing event, as no common splice recognition sequences are present in the surrounding region that could generate the final product. A similar example of atypical mRNA processing was described for the removal of a 144-bp exonic portion of an opioid receptor in human melanomas (Mayer et al. 2000), although the underlying mechanism remains unclear. Atypical splicing in the absence of canonical splice sites was also revealed for the 13th intron of the FCA gene (McKnight et al 1997), suggesting that this alternative mechanism of intron excision might exist in plants.

Another possibility may involve trans-splicing between the donor site of the last intron and the acceptor site of the last exon of a transcript derived from an intact *En/SpmB* element at an ectopic position. Different forms of trans-splicing were described in several lower eukaryotes such as trypanosomes and nematodes, and in mammalians as well (reviewed by Bonen 1993; Finta et al. 2002). In plants, mRNA trans-splicing occurs commonly in organelles (Bonen 1993; Malek and Knoop 1998) and has recently been described in nuclear genes in rice (Kawasaki et al. 1999). Indeed, the truncated *En/SpmB3* copy on chromosome 10 could serve as a template, as it does not contain the two insertions. However, sequence alignment rather disproves this hypothesis, since mismatches at specific positions are selectively present in this truncated copy, while absent in the 1.4 kb RT-PCR product. Moreover, there is no evidence that this truncated copy is transcribed.

Eventually, strand invasion of a small segment of the ectopic *En/SpmB3* site may occur during replication at the *hipa* target site, induced by double-stranded breaks near the transposon insertion. This could occur in a transient unstable intermediate that forms a substrate for

transcription, but not represented at the DNA at a high concentration. This form of gene conversion from an ectopic copy might alternatively explain the absence of a "footprint". In conclusion, the results reveal the occurrence of insertion removal, yet more experiments are needed to determine the exact mechanisms involved.

Transcript 'suppression' caused by the En/Spm transposase

In transgenic plants containing the maize transposase, the production of the long transcript is suppressed. This is reminiscent of the phenomenon of 'suppression' invoked by the *En/Spm* transposase in maize that has been originally documented at the phenotype level in maize kernel pigmentation (McClintock 1954). A possible hypothesis for the transcript suppression would imply the binding of the maize TnpD transposase to the termini of the related *hipa* element, with consequent inhibition of transcript read-through. A similar explanation was first suggested for the *En/Spm*-mediated suppression of the *Spm-I8* read through transcript in the *wx-m8* allele in maize (Gierl et al, 1985). In the example we document here, the function of suppression can be attributed to the expression of the maize TnpD, as the *hipa* element contains the corresponding CACTA-like binding sites. However, the influence of the maize TnpA, with or without TnpD, on the suppression effect cannot be ruled out.

Concluding Remarks

The maize *En/Spm* element encoded products were shown to interact with homologous elements in rice displaying the 'Suppressor' effect, to which this transposon system owes its name (McClintock, 1954). This effect on influencing the expression of genes, was what Barbara McClintock foresaw as one of the primary roles of these 'controlling elements'. This example is probably the tip of the iceberg, revealed by genomics methods, providing a regulatory role for transposable elements that comprise the bulk of complex plant genomes.

Acknowledgments

We wish to thank Anke J.C. Taal for generation of the transgenic rice lines and Renato Tarchini for help in some bioinformatic analyses. This research was funded by a European Union project BIO4 CT 972132 on "Transposon mutagenesis in rice". The work has been carried out in compliance with the current laws governing genetic experimentation in The Netherlands.

LITERATURE CITED

Aarts MGM, Corzaan P, Stiekema W, Pereira A (1995) A two-element enhancer-inhibitor transposon system in *Arabidopsis thaliana*. Mol Gen Genet 247: 555-564

Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402

Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42: 251-269

Bonen L (1993) Trans-splicing of pre-mRNA in plants, animals, and protists. FASEB J 7: 40-46

Britt A, Walbot V (1991) Germinal and somatic products of *Mu1* excision from the *Bronze1* gene of *Zea mays*. Mol Gen Genet 227: 267-276

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268: 78-94

Chen J, Greenblatt IM, Dellaporta SL (1992) Molecular analysis of *Ac* transposition and DNA replication. Genetics 130: 665-676

Chopra S, Brendel V, Zhang J, Axtell J, Peterson T (1999) Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*. Proc Natl Acad Sci 96: 15330-15335.

Coen E, Robbins T, Almeida J, Hudson A, Carpenter R (1989) Consequences and mechanisms of transposition in *Antirrhinum majus*. In: Mobile DNA. Edited by D. E. Berg and M. M. How. American Society for Microbiology, Washington, DC, pp. 413-436

Dennis E, Gerlach W, Peacock W (1986) Excision of the *Ds* controlling element from the *Adhl* gene of maize. Maydica 31: 47-57

Doseff A, Martienssen R, Sundaresan V (1991) Somatic excision of the *Mu1* transposable element of maize. Nucleic Acids Res 19: 579-584

Finta C, Warner SC, Zaphiropoulos PG (2002) Intergenic mRNAs. Minor gene products or tools of diversity? Histol Histopathol 17: 677-682

Frey M, Reinecke J, Grant S, Saedler H, Gierl A (1990) Excision of the *En/Spm* transposable element of *Zea mays* requires two element-encoded proteins. EMBO J 9: 4037-4044

Gierl A, Lutticke S, Saedler H (1988) TnpA product encoded by the transposable element En-1 of *Zea mays* is a DNA binding protein. EMBO J 7: 4045-4053

Gierl A, Schwarz-Sommer Z, Saedler H (1985) Molecular interactions between the components of the *En-I* transposable element system of *Zea mays*. EMBO J 4: 579-583

Giroux MJ, Clancy M, Baier J, Ingham L, McCarty D, Hannah LC (1994) De novo synthesis of an intron by the maize transposable element *Dissociation*. Proc Natl Acad Sci 91: 12150-12154

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

Goodall GJ, Filipowicz W (1991) Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. EMBO J 10: 2635-2644

Grant SR, Gierl A, Saedler H (1990) *En/Spm* encoded TNPA protein requires a specific target sequence for suppression. EMBO J 9: 2029-2035

Grant SR, Hardenack S, Trentmann S, Saedler H (1993) Functional cis-element sequence requirements for suppression of gene expression by the TNPA protein of the *Zea mays* transposon *En/Spm*. Mol Gen Genet 241: 153-160

Greco R, Ouwerkerk PBF, Taal AJC, Favalli C, Beguiristain T, Puigdomenech P, Colombo L, Hoge JHC, Pereira A (2001) Early and multiple *Ac* transpositions in rice suitable for efficient insertional mutagenesis. Plant Mol Biol 46: 215-227.

Hirochika H (2001) Contribution of the *Tos17* retrotransposon to rice functional genomics. Curr Opin Plant Biol 4: 118-122.

Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci 93: 7783-7788

Inagaki Y, Hisatomi Y, Suzuki T, Kasahara K, Iida S (1994) Isolation of a suppressor-mutator/enhancer-like transposable element, *Tpn1*, from Japanese morning glory bearing variegated flowers. Plant Cell 6: 375-383

Izawa T, Ohnishi T, Nakano T, Ishida N, Enoki H, Hashimoto H, Itoh K, Terada R, Wu C, Miiyazaki C, Endo T, Iida S, Shimamoto K (1997) Transposon tagging in rice. Plant Mol Biol 35: 219-229

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA transposon family in rice. Nature 421: 163-167

Kapitonov V, Jurka J (2001) Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci 98: 8714-8719

Kawasaki T, Okumura S, Kishimoto N, Shimada H, Higo K, Ichikawa N (1999) RNA maturation of the rice *SPK* gene may involve trans-splicing. Plant J 18: 625-632

Ki C, Je B, Piao H, Par S, Kim M, Park S, Park J, Park S, Lee E, Chon N, Won Y, Lee G, Nam M, Yun D, Lee M, Cha Y, Le K, Eun M, Han C (2002) Reprogramming of the activity of the *Activator*/*Dissociation* transposon family during plant regeneration in rice. Mol Cells 14: 231-237

Kikuchi K, Terauchi K, Wada M, Hirano HY (2003) The plant MITE *mPing* is mobilized in anther culture. Nature 421: 167-170

Kunze R, Weil C (2002) The hAT and CACTA superfamilies of plant transposons. In: Mobile DNA II. Edited by N. Craig, R. Craigie, M. Gellert and A.M. Lambowitz. American Society for Microbiology, Washington, DC, pp. 565-610

Lal SK, Giroux MJ, Brendel V, Vallejos CE, Hannah LC (2003) The maize genome contains a *helitron* insertion. Plant Cell 15: 381-391

Langdon T, Jenkins G, Hasterok R, Jones R, King I (2003) A high-copy-number CACTA family transposon in temperate grasses and cereals. Genetics 163: 1097-1108

Macknight R, Bancroft I, Page T, Lister C, Schmidt R, Love K, Westphal L, Murphy G, Sherson S, Cobbett C, Dean C (1997) *FCA*, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. Cell 89: 737-745

Malek O, Knoop V (1998) Trans-splicing group II introns in plant mitochondria: The complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. RNA 4: 1599-1609

Mao L, Wood TC, Yu Y, Budiman MA, Tomkins J, Woo S, Sasinowski M, Presting G, Frisch D, Goff S, Dean RA, Wing RA, Mao L, Yu YS, Woo SS (2000) Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. Genome Res 10: 982-990

Masson P, Surosky R, Kingsbury JA, Fedoroff NV (1987) Genetic and molecular analysis of the *Spm*-dependent *a-m2* alleles of the maize *a* locus. Genetics 117: 117-137

Masson P, Strem M, Fedoroff N (1991) The *tnpA* and *tnpD* gene products of the *Spm* element are required for transposition in tobacco. Plant Cell 3: 73-85

Mayer P, Tischmeyer H, Jayasinghe M, Bonnekoh B, Gollnick H, Teschemacher H, Hollt V (2000) A delta opioid receptor lacking the third cytoplasmic loop is generated by atypical mRNA processing in human malignomas. FEBS Lett 480: 156-160

McClintock B (1951) Chromosome organization and genic expression. Cold Spring Harbor Symp Quant Biol 16: 13-47

McClintock B (1954) Mutations in maize and chromosomal aberrations in *Neurospora*. Carnegie Inst Wash Year Book 56: 254-260

Menssen A, Hohmann S, Martin W, Schnable PS, Peterson PA, Saedler H, Gierl A (1990) The *En/Spm* transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a *dSpm* element in the *A2* gene. EMBO J 9: 3051-3057

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature 411: 212-214

Motohashi R, Ohtsubo E, Ohtsubo H (1996) Identification of *Tnr3*, a suppressor-mutator/enhancer-like transposable element from rice. Mol Gen Genet 250: 148-152

Nacken WKF, Piotrowiak R, Saedler H, Sommer H (1991) The transposable element *Tam1* from *Antirrhinum majus* shows structural homology to the maize transposon *En/Spm* and has no sequence specificity of insertion. Mol Gen Genet 228: 1-2

Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H, Tanisaka T (2003) Mobilization of a transposon in the rice genome. Nature 421: 170-172

Nicholas K, Nicholas HJ, Deerfield DI (1997) GeneDoc: analysis and visualization of genetic variation. EMBNET NEWS 4: 1-4

Panaud O, Vitte C, Hivert J, Muzlak S, Talag J, Brar D, Sarr A (2002) Characterization of transposable elements in the genome of rice (*Oryza sativa* L.) using Representational Difference Analysis (RDA). Mol Gen Genom 268: 113-121.

Pereira A, Schwarz-Sommer Z, Gierl A, Bertram I, Peterson P, Saedler H (1985) Genetic and molecular analysis of the *Enhancer* (*En*) transposable element system of *Zea mays*. EMBO J 4: 17-23

Pereira A, Cuypers H, Gierl A, Schwarz-Sommer Z, Saedler H (1986) Molecular analysis of the *En/Spm* transposable element system of *Zea mays*. EMBO J 5: 835-841

Peterson P (1953) A mutable *pale green* locus in maize. Genetics 45: 115-133

Raina R, Schlappi M, Karunanandaa B, Elhofy A, Fedoroff N (1998) Concerted formation of macromolecular suppressor-mutator transposition complexes. Proc Natl Acad Sci 95: 8526-8531

Saedler H, Nevers P (1985) Transposition in plants: a molecular model. EMBO J 4: 585-590

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765-768

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20: 43-45

Scott L, LaFoe D, Weil CF (1996) Adjacent sequences influence DNA repair accompanying transposon excision in maize. Genetics 142: 237-246

Snowden K, Napoli C (1998) Psl: A novel *Spm*-like transposable element from *Petunia hybrida*. Plant J 14: 43-54

Staginnus C, Huettel B, Desel C, Schmidt T, Kahl G (2001) A PCR-based assay to detect *En/Spm*-like transposon sequences in plants. Chromosome Res 9: 591-605.

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice *Adh1-adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12: 381-391

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876-4882

Trentmann S, Saedler H, Gierl A (1993) The transposable element *En/Spm*-encoded TNPA protein contains a DNA binding and a dimerization domain. Mol Gen Genet 238: 201-208

van Houwelingen A, Souer E, Mol J, Koes R (1999) Epigenetic interactions among three *dTph1* transposons in two homologous chromosomes activate a new excision-repair mechanism in petunia. Plant Cell 11: 1319-1336.

Varagona MJ, Purugganan M, Wessler SR (1992) Alternative splicing induced by insertion of retrotransposons into the maize *waxy* gene. Plant Cell 4: 811-820

Verwoerd T, Dekker B, Hoekema A (1989) A small-scale procedure for the rapid isolation of plant RNAs. Nucleic Acids Res 17: 2362

Weil CF, Wessler SR (1990) The effect of plant transposable element insertion on transcription initiation and RNA processing. Annu Rev Plant Physiol Plant Mol Biol 41: 527-552

Wessler SR (1989) The splicing of maize transposable elements from pre-mRNA--a minireview. Gene 82: 127-133

Wicker T, Guyot R, Yahiaoui N, Keller B (2003) CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. Plant Physiol 132: 52-63

Yu J, Hu S, Wang J, Wong GK, Li S et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296: 79-92.

Zaphiropoulos PG (2002) Template switching generated during reverse transcription? FEBS Lett 527: 326

Chapter 9

Summary, concluding remarks and future prospects.

SUMMARY

Rice has evolved with humans to be a major staple food crop and has recently emerged to be a monocot model plant from which gene function discovery is projected to contribute to improvements in a variety of cereals like wheat and maize. The recent release of rough drafts of the rice genome sequence for public research provides a vast resource of gene sequences whose functions need to be determined by reverse genetics methods.

Based on the assumption that the mutant phenotype associated with a gene function is often the best clue to understand its role in the plant, insertional mutagenesis with transposable elements was chosen as a strategy in the research work described in this thesis to address gene function in rice (Chapter 2). Transposon insertions can cause a knockout mutation by blocking the expression of a gene and display a mutant phenotype. The mutant gene tagged by the insertion can then be isolated by recovering DNA flanking the insert and subsequently lead to the isolation of the wild-type gene. However, the scope in the use of knockout mutations is limited, as the majority of genes display no obvious phenotype. This may be due to functional redundancy where one or more other homologous loci can substitute for the same function, subtle or conditional phenotypes, or early lethality. Gene detection strategies have therefore been developed in addition to classical knockout mutagenesis to address the function of genes that do not directly reveal an obvious phenotype when disrupted (Chapter 5).

The well characterized maize *Ac/Ds* and *En/Spm* transposon systems were employed as insertional mutagens based on their known ability to transpose in heterologous species. The use of both transposon systems could theoretically enhance genome coverage by offering the advantage of combining transposition to closely linked sites characteristic of the *Ac/Ds* system with the random transposition typical of *En/Spm*. Constructs for knockout mutagenesis were developed employing the autonomous *Ac* element and tested in *japonica* (Chapter 3) and *indica* (Chapter 4) genotypes. To incorporate gene detection, advanced two-component enhancer traps vectors were assembled, comprising of a mobile transposon element (*Ds* or *I/dSpm*) and the corresponding stable transposase (*Ac* or *En/Spm*) source under control of the CaMV 35S promoter. The mobile transposons contained in this case a GUS marker gene driven by a CaMV 35S minimal promoter that could display the pattern of expression of the adjacent trapped gene (Chapters 6 and 7). A large number of rice transformants were produced to test the different transposon constructs and molecularly analyzed for activity, with the final aim of identifying optimal “starter” lines for the development of tagging populations. Among the parameters evaluated were the continuous transposition through successive generations, the ability to generate large numbers of independent inserts in progeny plants and the target site specificity of

insertion. The usefulness of the selectable markers incorporated in the constructs was also assessed.

In general the *Ac/Ds* system revealed high mobility in rice, although inactivation of *Ds* was observed in later generations (Chapter 6) that confirmed earlier reports. The autonomous *Ac* transposon, in contrast, does not seem to lose mobility (Chapter 3) and was shown to efficiently transpose in *indica* genotypes as well (Chapter 4), supporting its further use in the establishment of a tagging system in this economically important subspecies.

Both *Ac* and *Ds* displayed amplification of copy number that enabled the generation of lines containing multiple transposons. The phenomenon, resulting from transposition during replication, was more pronounced in case of *Ac* and could be attributed to the high level of transposition activity induced in the T_0 regenerants by the presence of a strong enhancer adjacent to the promoter (either endogenous or constitutive) driving the *Ac* transposase gene in the donor T-DNA (Chapters 2, 3 and 6). High transposition activity generally correlated with the induction of early excisions as well, which nevertheless did not prevent later remobilization of transposons in sectors of the plants giving rise to the germ-line. Indeed, the relatively high rate of late transposition events ensured the generation of T_1 progeny plants with independent insertions even within the same family, both for *Ac* (Chapters 3 and 4) and for *Ds* (Chapter 6). In addition, these germinal insertions were observed to be preferentially propagated through the next generations.

The high frequency of T_1 progeny plants bearing independent insertions allowed the effective use of the two-component *Ac*/*Ds* lines for gene tagging, despite the reduced mobility of *Ds* documented in later generations. *Ds* inactivation shown to occur in a large part of the T_2 families (as well as in T_3; Chapter 6) can in fact be exploited as a natural system for the attainment of stable insertion lines, with no further need of selection against the transposase source. As the diversity in the T_2 insertion pool appeared to be directly correlated to the initial number of different T_0 regenerants and therefore T_1 families propagated, large numbers of lines in earlier generations are therefore being propagated in order to obtain the maximum number of independent stabilized insertions. The T_2 and T_3 lines generated will thus provide a collection of stable insertions that can directly be used for reverse genetics screenings.

The difference observed in the mobility of *Ac* and *Ds* through successive generations may reflect a reduced sensitivity of the active autonomous *Ac* promoter and termini to methylation, which is usually the cause of inactivation of transposable elements. Based on this assumption, two-component vectors in which the endogenous *Ac* promoter drives the transposase source in place of the generally employed CaMV 35S promoter could be utilized to overcome *Ds* inactivation and are currently being tested. To ensure high levels of transposition activity, the *Ac* promoter is being flanked by a strong enhancer.

The Ac/Ds enhancer trap lines were also assessed for their ability to function as “detectors” of gene activity (Chapter 6). However, a low frequency of GUS staining patterns in somatic sectors could be revealed, likely due to the fact that only subsets of unstable T_1 lines with ongoing active somatic transpositions were tested. More thorough screenings are therefore required to fully evaluate the functionality of the system. Nevertheless, inefficient expression of the *GUS* reporter gene might be expected as a result of the presence of foreign sequences (*lox* site) between the minimal CaMV 35S promoter and the *GUS* translation start.

Pilot sequencing of genomic sites flanking the *Ac* and *Ds* inserts (Chapter 3, 4 and 6) revealed a preferential insertion of these transposons into genes or gene-rich regions, supporting their use as effective mutagenic agents. In addition, their tendency to transpose to linked sites was confirmed, which makes them suitable for targeted tagging, thereby saturating local chromosomal regions with insertions. In view of the need to generate databases of insertion tagged sites for reverse genetics screenings, a number of techniques were assessed for their potential use in systematic isolation of flanking sequences. Particularly efficient turned out to be TAIL-PCR and an adapted version of Transposon Insertion Display (Chapter 6), with the latter being preferable for the isolation of sequences from multiple insertion lines. As an alternative reverse genetics method, the use of the “tagged transcriptome display” (Chapter 4) was shown to further implement the direct identification of insertions in transcribed sequences, and offers the opportunity to isolate insertions in genes specifically expressed in certain developmental stages or experimental conditions.

On the other hand, the *En/Spm* system displayed a surprisingly low transposition activity in rice, restricted to somatic events that were not transmitted to the next generation, in spite of being a well-established mutagenic system in heterologous dicot species such as Arabidopsis (Chapter 7). Among the hypotheses invoked to explain this behaviour, splicing deficiency of the *En/Spm* maize element in rice may cause a reduced availability of the transposase products. Interestingly, correct splicing of the element does not seem to correlate with transposition ability, as all the expected products were shown to be produced in rice similarly to maize, while differences were revealed with the dicot Arabidopsis (similar to tobacco). Rather, the ratio at which the different transposase products necessary for transposition are produced seems to be critical and influenced by host factors. In addition, transposition efficiency might be further reduced by the lack of essential *cis*-required sequences in the modified *I/dSpm* version utilised, although similar constructs were successfully employed in Arabidopsis. Eventually, cross-talk with endogenous related transposable elements may affect the mobility of the maize transposon in rice and this possibility was explored in Chapter 8. Indeed, interaction was revealed among the maize transposases and a rice *En/Spm*-homologous element that results in the specific suppression of an alternative transcript in the latter. This finding, although does not explain the poor transposition

ability of *En/Spm* in rice, on the other hand demonstrates that interference is possible and trans-activation potentially could occur among elements belonging to the same family in different species.

Effective tagging relies in the availability of a consistent number of lines bearing independent transposon insertions and this, in turn, relies on the possibility of performing fast and efficient selections for independent transposants. Thus, the choice of appropriate selectable markers is a critical step in the design of the constructs to be employed. The use of heterologous transposons over endogenous transposons offer the great advantage of being able to modify the transposon itself to carry ideal markers that will aid during the screening procedure. In case of the simple knockout constructs developed in this work, only excision of the autonomous *Ac* element from the donor T-DNA could be monitored and a *GFP*-excision assay was developed (Chapters 2, 3 and 4). However, sectorial silencing of GFP expression in developing plants and later generations hampers its efficient use as a reliable indicator of excision. In case of two-component constructs, a double selection scheme for greenhouse screenings of stable insertions was initially devised based on Tissier et al. (1999), with the *BAR* gene as positive selection marker on the mobile transposon (*Ds* or *I/dSpm*) and the *SU1* gene as negative selection marker on the donor T-DNA against the transposase source (Chapters 6 and 7). However, *SU1* did not seem to work effectively in rice, contrary to reports describing its successful use in the related cereal barley (Koprek et al. 1999). As an alternative, constructs were made using the *GFP* gene as screenable negative selection marker and are being used in forthcoming experiments.

Based on the results of these analyses, a core collection of knockout and gene detection *Ac/Ds* lines with active transposition could be selected as a basis to develop populations for (forward and) reverse genetic screenings. The propagation of lines containing multiple transposons and the preferential insertion into gene-rich regions will reduce the number of plants that would have to be produced in order to saturate the genome with insertions. Segregation of the additional transposon copies can then be carried out once a mutant is identified. Knockout *Ac* lines have the disadvantage that they cannot be stabilised, but mutants obtained that display a variegated phenotype can easily be analyzed. Other transposons mutagenesis systems employing endogenous multi-copy transposons have also been proven to be effectively useful for gene tagging. A ready example is coming directly from rice where the endogenous retrotransposon *Tos17*, which is activated by tissue culture and can usually reach an average copy number of 5-30 per plant, has been widely used for producing large insertional populations and is to date the only system that has yielded a tagged gene (Agrawal et al. 2001; Hirochika 2001).

CONCLUDING REMARKS AND FUTURE PROSPECTS

At the time this thesis work started efforts toward systematic functional analysis of the rice genome were beginning to rise, driven by the imminent completion of the genome sequencing. Now, several approaches are being undertaken, such as transcript profiling using microarrays, proteomics and large-scale mutant generation mainly via insertional mutagenesis, which will collectively contribute to shed light on the biological role of the many uncovered rice genes. A summary of the main functional genomic resources that are being developed for rice in terms of insertional mutagenesis is reported in Table 1.

Table 1. Insertional mutant collections.

Group	Features	Cultivar	Lines	ITS	References
NIAS, Japan (Hirochika)	Tos17 / knockout	Nipponbare	50,000	18,000	Hirochika 2001, Miyao and Hirochika 2003
NIAS, Japan (Kikuchi)	T-DNA / Activation tag	Nipponbare	10,000		Mori et al. 2003
NIG, Japan (Kurata)	Ac/Ds / Enhancer trap	Nipponbare	18,000		Ito et al. 2003
CRFGP, China	T-DNA / Enhancer trap	Zhonghua 11 and 15 *, Nipponbare	30,000		Wu et al. 2003, Xue et al. 2003
	Ac/Ds / Knockout		5,000		
CIRAD, France (Guiderdoni)	T-DNA / knockout	Nipponbare	40,000	8000	Sallaud et al. 2003a,b
University of Cambridge, UK (Tester) - CIRAD, France (Guiderdoni)	T-DNA / Enhancer trap	Nipponbare	8,000		Johnson et al. 2003
CSIRO, Australia (Upadhyaya)	T-DNA / Gene trap + *Ds*	Nipponbare			Eamens et al. 2002
	Ac/Ds / Gene and Enhancer trap		5,000		Upadhyaya et al. 2002a, b
Postech Plant Functional Genomics Laboratory, Korea (An)	T-DNA / Gene trap + activation tag	Dongjing *	100,000	7000	Jeon et al. 2000, Jeong et al. 2002
GSNU, Korea (Han)	Ac/Ds / Gene trap	Dongjing *	>30,000	3000	Chin et al. 1999, Han 2003
Temasek Life Sciences Laboratory, Singapore (Ramachandran)	Ac/Ds / Gene trap	Nipponbare	4,200	2000	Kolesnik et al. 2003, Szeverenyi et al. 2003
US NSF group (Leach)	Ac/Ds / Activation tag	Nipponbare			Qu et al. 2003

* *Japonica* cultivars.

The large number of predicted genes and the average gene density indicate that a minimum of 400,000 insertions would be required in order to tag each gene in the 430 Mb of the rice genome. As this effort is undoubtedly not feasible for a single lab, an international collaboration among 18 institutions from 10 countries and two international agricultural research centers has resulted in the formation of an International Rice Functional Genomics Consortium (http://www.iris.irri.org/IRFGC/default.shtml). Among the projected consortium achievements is the tagging of 90% of the rice genes by the year 2005 and the functional characterisation of 60% by 2010. In view of this ambitious goal, the development of saturated populations using different types of tags, each having different target-site specificity, will ensure a non-biased collection of insertions increasing the tagging efficiency. The work reported in this thesis will contribute towards the effort by the utilization of active transposon lines. At present over 10,000 stabilised T_2 *Ac/Ds* transposon lines are being analyzed in 5 EU labs for transposon flanking sequences that by comparison to the complete and annotated rice sequences will reveal tagged genes of interest that can be used for reverse genetics. In parallel, a gain-of-function activation tagging strategy is being pursued to complement gene detection for gene function discovery, especially in case of genes with redundant function (Marsch Martinez et al 2002).

As preferential insertion in gene rich regions emerged to be a general feature shared by other insertion tags utilised in rice besides *Ac* (Enoki et al. 1999, Chapters 3 and 4) and *Ds* (Chapter 6), such as *Tos17* (Yamazaki et al. 2001), as well as T-DNA (Barakat et al. 2000), it is reasonable to predict that the estimated number of insertions needed to saturate the genome would indeed be smaller. On the other hand, it now starts to become evident that rice is a difficult species for the identification of tagged mutants. Results from several groups endeavouring functional genomics strategies using insertional mutagenesis reveal a very low percentage of forward mutations that are eventually tagged by the insertion element used and accordingly, the amount of genes that were cloned or identified is surprisingly low if compared to similar approaches employed in the other model species Arabidopsis. To date, the only examples of successful gene tagging in rice refer to the cloning of two genes involved in conferring the *viviparous* mutant phenotype caused by *Tos17* insertions (Agrawal et al. 2001), while tagging of other genes has only been mentioned without further reporting (Hirochika 2001; Zhu et al. 2002; Xue et al. 2003). The low tagging frequency is likely to be attributed to the extensive use of tissue culture regeneration procedures for the isolation of the insertions, known to create a high background of somaclonal untagged mutations. The use of advanced generation transposon lines should circumvent this tissue culture background and reveal more tagged mutants in the future.

Transposable elements were first described as ‘controlling elements’ by Barbara McClintock suggesting they had roles to play in regulating the expression of genes. Since their molecular elucidation and use in transposon tagging for the isolation of genes, studies on them

have often adapted a utilitarian role. However their utilisation in heterologous plant species meant that one had to study their behaviour anew and they often offered many surprises. In the course of this thesis, beginning with transposon engineering and ending up studying the interaction of transposons in the genome, one is finally confronted with the truth that transposable elements are ‘controlling elements’ of gene regulation.

LITERATURE CITED

Agrawal GK, Yamazaki M, Kobayashi M, Hirochika R, Miyao A, Hirochika H (2001) 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: 1248-1257

Barakat A, Gallois P, Raynal M, Mestre Ortega D, Sallaud C, Guiderdoni E, Delseny M, Bernardi G (2000) The distribution of T-DNA in the genomes of transgenic Arabidopsis and rice. FEBS Lett 471: 161-164

Chin HG, Choe MS, Lee SH, Park SH, Park SH, Koo JC, Kim NY, Lee JJ, Oh BG, Yi GH, Kim SC, Choi HC, Cho MJ, Han CD (1999) Molecular analysis of rice plants harboring an *Ac/Ds* transposable element-mediated gene trapping system. Plant J 19: 615-623

Eamens A, Blanchard C, Ramm K, Dennis ES, Upadhyaya NM (2002) A duel orientation T-DNA-cum-*Ac/Ds* gene trap system. Abstract of the International Worhshop "Towards building a global rice gene machine". CSIRO Plant Industry, Canberra, Australia

Enoki H, Izawa T, Kawahara M, Komatsu M, Koh S, Kyozuka J, Shimamoto K (1999) *Ac* as a tool for the functional genomics of rice. Plant J 19: 605-613

Han CD (2003) Mass production and analysis of *Ds* transposant lines in rice. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Hirochika H (2001) Contribution of the *Tos17* retrotransposon to rice functional genomics. Curr Opin Plant Biol 4: 118-122

Ito Y, Thirumurugan T, Miyoshi T, Eiguchi M, Kurata N (2003) Generation and screening of enhancer trap lines of rice. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Jeon JS, Lee S, Jung Ki H, Jun Sung H, Jeong DH, Lee J, Kim C, Jang S, Lee S, Yang K, Nam J, An K, Han Min J, Sung Ryo J, Choi Hyun S, Yu Jung H, Choi Jung H, Cho SY, Cha Sang S, Kim Shi I, An G (2000) T-DNA insertional mutagenesis for functional genomics in rice. Plant J. 22: 561-570.

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

Johnson A, Tester M, Hibberd J, Gay C, Sallaud C, Guiderdoni E (2003) Generation of enhancer trap lines of rice (*Oryza sativa* L.) Expressing GAL4 and GFP in specific cell types. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Kolesnik T, Szeverenyi I, Bachmann D, Chellian S, Shuye J, Rengasamy R, Cai N, Gang M, Venkatesan S, Ramachandran S (2003) Establishing an efficient *Ac/Ds* tagging system in rice: large-scale analysis of *Ds* flanking sequences. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Koprek T, McElroy D, Louwerse J, Williams Carrier R, Lemaux PG (1999) Negative selection systems for transgenic barley (*Hordeum vulgare* L.): comparison of bacterial *codA*- and cytochrome P450 gene-mediated selection. Plant J 19: 719-726

Marsch Martinez N, Greco R, Van Arkel G, Herrera Estrella L, Pereira A (2002) Activation tagging using the *En-I* maize transposon system in Arabidopsis. Plant Physiol 129: 1544-1556

Miyao A, Hirochika H (2003) Retrotransposon insertion lines of rice for functional genomics. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Mori M, Tomita C, Sugimoto K, Ooka H, Onodera H, et al. (2003) Large scale production of rice activation tagging lines and characterization of several mutants. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Qu S, Jeon J, Richter T, Leach J, Ronald P, Wang G (2003) Development of chemical inducible *Ac/Ds* elements for gene mutagenesis in rice. Abstract of Plant and Animal Genome XI Conference. San Diego, CA

Sallaud C, Gay, Larmande P, Bourgeois E, Piegu B, Regad F, Bes M, Piffanelli P, Perin C, Ghesquiere A, Delseny M, Guiderdoni E (2003b) A genome wide library of insertion lines in rice: distribution of T-DNA inserts over the rice genome. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Sallaud C, Meynard D, Van Boxtel J, Gay C, Bes M, Brizard J, Larmande P, Ortega D, Raynal M, Portefaix M, Ouwerkerk P, Rueb S, Delseny M, Guiderdoni E (2003a) Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics. Theor Appl Genet 106: 1396-1408

Szeverenyi I, Bachmann D, Kolesnik T, Cai M, Lau P, Luan H, Sundaresan S, Ramachandran S (2003) Transposon tagging in rice using *Ac/Ds* gene trap system. Abstract of the 7th International Congress of Plant Molecular Biology. Barcelona, Spain

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JD (1999) Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. Plant Cell 11: 1841-1852

Upadhyaya NM, Zhou XR, Zhu QH, Ramm K, Wu L, Eamens A, Sivakumar R, Kato T, Yun DW, Santhoshkumar C, Narayanan KK, Peacock JW, Dennis ES (2002a) An i*Ac/Ds* gene and enhancer trapping system for insertional mutagenesis in rice. Funct Plant Biol 29: 547-559

Upadhyaya N, Zhou X, Zhu Q, Eamens A, Ramm K, et al. (2002b) Can we produce a substantial "Rice Gene Machine" using the *Ac/Ds* system? Abstract of the International Worhshop "Towards building a global rice gene machine". CSIRO Plant Industry, Canberra, Australia

Wu C, Li X, Yuan W, Chen G, Kilian A, Li J, Xu C, Li X, Zhou D, Wang S, Zhang Q (2003) Development of enhancer trap lines for functional analysis of the rice genome. Plant J 35: 418-427

Xue Y, Li J, Xu Z (2003) Recent highlights of the China Rice Functional Genomics Program. Trends Genet 19: 390-394

Yamazaki M, Tsugawa H, Miyao A, Yano M, Wu J, Yamamoto S, Matsumoto T, Sasaki T, Hirochika H (2001) The rice retrotransposon *Tos17* prefers low-copy-number sequences as integration targets. Mol Genet Genom 265: 336-344

Zhu Q, Ramm K, Shivakkumar R, Dennis E, Upadhyaya N (2002) A *Ds* tagged mutant defective in anther dehiscence. Abstract of the International Worhshop "Towards building a global rice gene machine". CSIRO Plant Industry, Canberra, Australia

SAMENVATTING (Dutch Summary)

Rijst is een belangrijk voedingsgewas en een erkende monocotyle modelplant, waarvan de ontdekking van de functies van genen kan bijdragen aan de verbetering van graangewassen zoals tarwe en maïs. Het recente beschikbaar worden van een eerste versie van de rijstgenoomsequentie voor onderzoek in de publieke sector levert een bron van gensequenties waarvan de functies bepaald moeten worden met genetische technieken.

Karakterisering van het fenotype van een mutant is een van de meest geschikte technieken voor het leggen van een verband tussen een gen en zijn funktie. Hier vanuit gaande werd mutagenese door middel van mobiele ("transposable") DNA elementen gekozen voor het onderzoek dat in dit proefschrift beschreven wordt, om genfunctie in rijst te bestuderen (Hoofdstuk 2). De goed gekarakteriseerde *Ac/Ds* and *En/Spm* transposon systemen uit maïs werden gebruikt als insertiemutagenen, mede gebaseerd op hun reeds bekende vermogen om ook in andere soorten te transposeren. Transposon inserties kunnen een "knockout" mutatie veroorzaken door het blokkeren van de correcte expressie van een gen, wat dan kan leiden tot een mutant fenotype. Het gemuteerde gen, dat zo moleculair gemarkeerd wordt door het ingevoegde transposon, kan dan worden gevonden en gezuiverd door het isoleren van het DNA aan weerszijden van het ingevoegde element. De flankerende DNA sequenties kunnen dan gebruikt worden voor isolatie van het wild-type gen.

Constructen voor "knockout" mutagenese met het autonome *Ac* element werden gemaakt en getest in de rijstvariëteiten *japonica* (Hoofdstuk 3) en *indica* (Hoofdstuk 4). Het nut van "knockout" mutaties wordt beperkt door het feit dat de meerderheid geen herkenbaar fenotype laat zien. Dit kan veroorzaakt worden door functionele redundantie, waar een of meerdere genen dezelfde funktie kunnen vervullen, door subtiele of conditionele fenotypen, of door letaliteit van de mutatie. Om de funktie te bestuderen van genen, waarvan uitschakeling niet direct tot een duidelijk fenotype leidt, werden behalve klassieke "knockout" mutagenese ook gendetectie-strategiën ontwikkeld (Hoofdstuk 5). Voor gendetectie werd *japonica* rijst getransformeerd met nieuwe twee-componenten "enhancer trap" vectoren, bestaande uit een mobiel transposon element (*Ds* of *I/dSpm*), en het corresponderende, stabiele transposase gen (*Ac* of *En/Spm*) gecontroleerd door de CaMV 35S-promoter. In dit geval bevatte het mobiele transposon een *GUS* markergen onder controle van een minimale CaMV 35S-promoter, welke het expressiepatroon van het aangrenzende gen kan laten zien en zo aanwijzingen voor de funktie ervan oplevert (Hoofdstukken 6 en 7). Een groot aantal rijsttransformanten werd gemaakt om de activiteit van de verschillende transposonconstructen te testen, met het uiteindelijke doel om de optimale "startlijnen" voor ontwikkeling van gemarkeerde populaties te identificeren. Eigenschappen die beoordeeld werden waren o.a. het vermogen om transpositie te blijven vertonen in

opeenvolgende generaties, het vermogen om grote aantallen onafhankelijke inserties in nakomelingen te produceren, en de doelwitlocatie-specificiteit van de insertie. De bruikbaarheid van de in de constructen ingebouwde selecteerbare markers werd ook beoordeeld.

De resultaten toonden een hoge mobiliteit van het *Ac/Ds*-systeem in rijst aan (Hoofdstukken 3, 4 en 6), hoewel inactivering van *Ds* in latere generaties werd waargenomen (vanaf T_2). Niettemin maakte de hoge frequentie van onafhankelijke transpositie in de eerste generaties (T_0 en T_1) de productie mogelijk van T_2 en T_3 lijnen met onafhankelijke, gestabiliseerde inserties, welke gebruikt konden worden voor “reverse genetics” screening zonder dat verder tegen de oorspronkelijke bron van transposase geselecteerd behoefde te worden. Het autonome *Ac* transposon daarentegen leek zijn mobiliteit niet te verliezen en efficiënte transpositie in *japonica* èn *indica* variëteiten kon ook aangetoond worden, wat de verdere bruikbaarheid voor het opzetten van een markeringssysteem in laatstgenoemde, economisch belangrijke ondersoort bevestigd. Zowel *Ac* als *Ds* lieten een toename van het aantal kopieën zien, wat het mogelijk maakte om lijnen met meerdere transposons te maken. Een eerste sequentiebepaling van het DNA aan weerszijden van de *Ac* en *Ds* inserties liet een voorkeur voor insertie van deze transposons in genen en genrijke omgevingen zien en bevestigde hun voorkeur voor transpositie naar nabij liggende locaties, hetgeen ze geschikt maakt voor doelgericht markeren van genen. Eerste testen van de *Ac/Ds* “enhancer-trap” lijnen op hun vermogen om te functioneren als “detectoren” van genactiviteit lieten een lage frequentie van GUS-kleuring in somatische sectoren zien. Meer uitgebreide testen om het functioneren van dit systeem te onderzoeken zijn momenteel aan de gang.

In tegenstelling tot *Ac/DS* liet het *En/Spm*-systeem in rijst een verrassend lage transpositie activiteit zien (Hoofdstuk 7), beperkt tot somatische gebeurtenissen die niet naar de volgende generatie werden doorgegeven, hoewel het een met succes gebruikt mutagenese systeem in heterologe dicotyle gastheren als Arabidopsis is. Transcriptie analyse van het *En/Spm*-element in rijst liet zien dat correcte splicing van het element plaats vindt, maar dat het niet genoeg is voor het vermogen om te springen. In plaats daarvan lijken de relatieve hoeveelheden van de verschillen transposase producten, noodzakelijk voor transpositie, die gevormd worden bepalend en beïnvloed door gastheerfactoren. Daarnaast kan de efficiëntie van transpositie verder afnemen door een gebrek aan essentiële in *cis*-werkende sequenties in de gemodificeerde *I/dSpm* versie die in dit onderzoek werd gebruikt, hoewel vergelijkbare constructen met succes in Arabidopsis gebruikt werden. Tenslotte, “cross-talk” met verwante endogene mobiele elementen zou de mobiliteit van het mais transposon in rijst kunnen beïnvloeden. Interactie van de maïs transposases met een *En/Spm*-homoloog element, hetgeen resulteerde in the specifieke suppressie van een alternatief transcript in de laatstgenoemde, werd aangetoond (Hoofdstuk 8). Dit resultaat laat zien dat interferentie mogelijk is en dat *trans*-activatie waarschijnlijk zou kunnen

plaatswinden tussen elementen uit dezelfde transposon familie, maar afkomstig uit verschillende soorten.

Met de resultaten van deze analyse kon een kerncollectie van “knockout” en gendetectie *Ac/Ds*-lijnen met actieve transpositie worden geselecteerd als basis voor de ontwikkeling van populaties voor (“forward” en) “reverse genetics” selecties. Het doorkweken van lijnen met meerdere transposons en de voorkeur van transposons voor inserties in gen-rijke omgevingen zullen bijdragen aan een verlaging van het aantal planten dat geproduceerd zou moeten worden om het genoom met inserties te verzadigen. Momenteel worden in 5 EU laboratoria meer dan 10.000 gestabiliseerde T_2 *Ac/Ds*-transposonlijnen geanalyseerd voor transposon-flankerende sequenties. Door vergelijking van die sequenties met het complete en geannoteerde rijstgenoom zullen die leiden tot interessante, gemarkeerde genen, die gebruikt kunnen worden voor “reverse genetics”.

NAWOORD (ACKNOWLEDGEMENTS)

In 1997, when I came to Wageningen to the lab of Dr. Pereira, I would never imagine that the originally planned three months visit would turn into six years and even more and that instead of just using transposon populations I would start generating them…from the very beginning. The major outcome of these years lays in the completion of this booklet, to which many people - in different ways - gave their special contribution.

Andy! Thanks for introducing me to the fascinating world of transposons and for your guidance during the years, your help and prompt availability whenever needed. Your knowledge and enthusiasm have been an incentive to always try to achieve the most possible achievable and to never give up. I enjoyed and learned a lot from the long evening discussions about the "how" and "why" of transposons behaviour. And last but not least, I enjoyed the many "borrels", dinner and parties, of which you were always a very active participant. My thanks also to Isolde, for her kind hospitality in many occasions.

Maarten, I would like to thank you for the care and support you gave me especially in the final steps of my thesis. I appreciated very much your advices and the example of a pure, straightforward scientific approach, as well as your extreme kindness.

The work presented in this thesis was carried out as part of an EU project involving five other partner labs. Pieter, Harry (Rest In Peace), Annemarie, Christophe, Emmanuel, Ajay, Paul, Cristina, Lucia, Enrico, Thierry and Pere. I enjoyed both the collaboration and not less the meetings in those beautiful locations! Harry, it was a pleasure working with you and I missed your enthusiasm and "pragmatic" approach. Pieter, thanks for your help, discussion and fruitful cooperation during these years. I sincerely appreciate your constant motivation, your will to always progress and help whenever needed. And as you say…may the reviewers be with us!! Saskia, thanks for your advices as "rice expert" and your guidance during the rice transformation phase (together with Anke), and to all of you thanks for the nice atmosphere I always found when I came to work in Leiden. Christophe, in spite of the bad news (!) I enjoyed interacting with you and wish you good luck with your future career. Cristina, Enrico and Thierry, thanks for your help with the analysis of many "transposon" rice plants in the first part of the project, before the "high throughput era" started in our lab. Cristina, I enjoyed very much working with you (you were unbelievably tireless!) and having you around as a friend…I'll never forget you trying to fit in the boxes in Gamma!!

From our group at PRI, Gert and Wim are "hartelijk" thanked for setting up the HTP facilities (that changed my life … and many others) and for the help in the isolation of the (many!) transposon flanking sequences, but also and especially for their friendly company in the lab and around. Bedankt, jonges! I also wish to acknowledge Marleen and Paul for the sequencing

support, and Hans (Rest In Peace) and Joost for providing me with ready-to-analyse Blast results (which saved me much time). Renato, I appreciated your bioinformatic support and, not less, your company. John, thanks for your kind help and availability whenever needed with microscopes and cytological stuffs, and not only (Robert, thanks to you as well!). Unfortunately, that tiny "blue spot" that we photographed from all directions was too tiny to be true! Peter and Jan, thanks for the prompt assistance in the many computer emergencies I encountered during those years. Growing rice in the Dutch climate is not an easy task. Dick, Gerrit ! thanks for taking good care of the rice plants and for the enthusiasm in learning how to do it in the best way.

To my past and present colleagues of the cluster "Genoomanalyse" with whom I shared labs, offices and daily life at PRI: Mark, Pieter, Bas (hoe laat is het?), Elly, Petra, Miriam, Marianne, Marga (Rest In Peace), Ludmilla (thanks for your advices on any RNA or expression concerning matter), Doret, Annelies, Marjon, Ellen (you inherited the rice work…I wish you good luck with your progenies, at this moment especially with your second F1!), Edwin (always friendly and cheerful), Carolien (I appreciate your several attempts to make me speaking Dutch, although the results were not so promising…), Eric (will you still join us for the ERA birthday party??), Jan Peter; And to the rest of the "Genomics" colleagues. Thank you all for help, explanations, suggestions and nice atmosphere along the years! Nayelli, a special thanks for your help, your friendship and your support during this long time. Having you around was very nice; I knew I always had someone to discuss with and to count on. I am glad that our unsuccessful Arabidopsis transformation of the very beginning turned out later on to be very successful for you, not only professionally! I wish you much success with your future career ☺

Among the colleagues from the former Molecular Biology department (now back in Bioscience!), I wish to recall Hans and Luud as pleasant room-mates for a long while, Jules (for being Jules!), Maarten (thanks to you and Cory for teaching me how to make home-made pasta!). Ruud, the translator of the Samenvatting…thanks very much for that and for many years of friendship. It was a pleasure knowing you and your "critical" approach to any aspect of life has been always a stimulus towards improvements. To my former and present BU managers Willem and Raoul, thanks for your support. Raoul and Madelon, I sincerely appreciated your warm introduction to some of the Jewish traditions. Harro, I was honored to be a "terpenoid" for one evening! Willem (Rest In Peace), Hans, thanks for your kind patience whenever I was late for the closing time.

For the work on the Arabidopsis gain of function mutants (not included in this thesis) I am grateful to Isabelle, Elise, Meriem and Shital for their help; to Gerrit, Dan and Piet for taking care of the Arabidopsis plants in the greenhouse; to Stefan and Antonio for help and useful discussions. Stefan, thanks a lot for your friendship and kind help in many occasions and… sorry if your beloved aquarium lost a few "inhabitants" after my care!

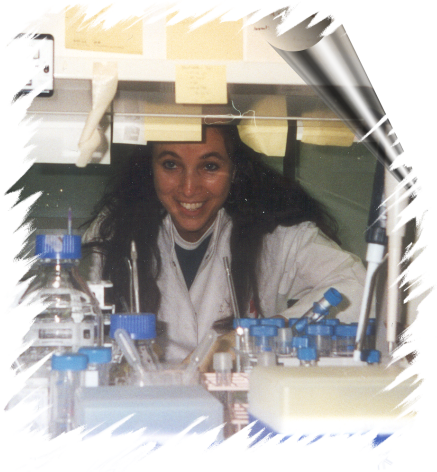
During the time, I had the opportunity to meet many people from many different countries and cultures, with whom I sometimes built special friendships. In the beginning, I found this extremely exciting. After almost seven years, I must admit this is indeed still exciting, yet emotionally very tiring. Very few people remained from the beginning, most have left, new have come, and eventually I'll leave as well. Good-bye parties became kind of "painful" with the time… Maria Carmen, Nikola, Boris, Seetharam, Ela (one of the most gentle souls I have ever met), Gyuri (Rudi…ogje vagje?), Armini, Mariateresa (no words for the unforgettable laughs and your exceptional friendship!), Katja (I still did not lose hope to come to Ljubljana and visit you one day), Gaby, Nikolay and Rumyana (and the rest of the Bulgarian community, Samir, Elena&Arnoud, Marina, Violeta, Tzanco, Pavlina&Serghiey) with whom I had always very nice time, Daniele, Alex, Rajesh and Sapna (no more Indian lollypops!), Ashok (what kind of this thing is??), Rajan, Elena (is there hope for a deep frozen fish??), Natasha, Rolando (how about the temperatures in Cuba?), Prof. Yang (it was a pleasure, Yehua!), Wang, Vagner (the goal is almost reached!!), Paulo (thanks for the "last" crucial advices) Tetty (24!!), Kim.Engelke (the woman from the thousands surprises), Markus and Christian and Willi (we know this Germans …they eat all our spaghetti!), Jana, Radka (what about the labucce?), Salva (please…"drei" more songs!), Cinzietta (Santa Cinzia da San Secondo di Pinerolo…buona fortuna con le specule!), Ale (my reliable fitness partner e non solo…), Isabella, Stefano (..Fabio?!!), Hairpin-Teun and many others…. I am very happy I had the chance to meet all of you and I really enjoyed your company, the nice talks, parties, etc. of which the memories will always be with me.

To my "old" friends that during this long time shared with me joys and disappointments, from far away or from very close: Lucia (l'artefice di tutto cio', nonche' mia paraninfa "storica"!) and Martin (should I say Professor Kater?!), Diana (spero che rimarremo nonostante tutto "vicine") and Jeroen (J), Lisa, Sara. Thanks a lot for your support, for being there whenever I needed you and for the fun time we had together! Unforgettable will remain also the special atmosphere of the very beginning, for which I'd like to remember Silvia (in spite of everything I'll always remember our countless happy and sad chats and our shared feelings about many things) and Marcel (your friendship and your help was of great support in the difficult start).

Ruthi and Faiko (and the rest of the Aharoni family), I appreciated your warm support and …thanks for giving us the opportunity to "roam around Holland" every now and then!

For a strange law of life, the people you care the most are always the ones you make bearing the most difficult tasks. And so, Mamma e Vince, for you most difficult was to accept the choice to leave so far for so long, and to come home so little and for so short. Still you supported me all these years and to know that you were there was of great comfort to me during this long period abroad. Mammina, grazie per essermi stata sempre vicina nonostante ti abbia lasciato sola negli anni forse piu' difficili. Fratellino, saresti stato un paraninfo eccezionale!

And….in these years of very hard working, high pressure and much stress, I owe my immense gratitude to the one person that more than any other supported and encouraged me from the beginning and all the way, through many difficult moments, till the final step resulting in the completion of this thesis. Thanks Asaph!! Your “press ahead, press ahead” was successful after all. I hope in the future we could see other fields than strawberry and rice… Greece is still waiting!



Raffa

CURRICULUM VITAE

Raffaella Greco was born in Milano (Italy) on the 26th of September 1964. She started her studies in Biological Sciences in 1983 at the University of Milano, where she obtained her M.Sc. degree *cum laude* in 1991 presenting a thesis entitled “Environmental influences on the estimate of the genetic variability of the photosynthetic rate in maize: A biometric approach”. The experimental work was carried out under the supervision of Prof. Alessandro Camussi and Prof. Ercole Ottaviano at the Department of Genetics of the University of Milano, where she continued to work after graduation as a fellow of the National Research Council with Dr. Enrico Pe’ and Prof. Mirella Sari Gorla. The research area involved mapping of resistance genes to the fungal pathogen *G. zeae* in maize (1991-1993) and flower development in cereals (1993-1995). In January 1996 she joined the group of Prof. Chiara Tonelli to work on a EU project concerning the functional analysis of the *myb* family of transcription factors in Arabidopsis. Between 1995 and 1997 she spent eight months at Plant Research International (Wageningen, NL) for research training in the groups of Dr. Gerco Angenent and Dr. Andy Pereira. Since August 1997 she is working at Plant Research International in the Business Unit Bioscience, with the supervision of Dr. Andy Pereira. The work carried out on the development of transposon mutagenesis systems for functional genomics in the model species rice and Arabidopsis is partly presented in this thesis.

PUBLICATIONS

The work carried out during the course of this research has resulted in the following publications:

Greco R, Ouwerkerk PBF, Sallaud C, Kohli A, Favalli C, Beguiristain T, Colombo L, Pè E, Puigdomènech P, Guiderdoni E, Christou P, Hoge JHC, and Pereira A (2001)
Transposons and functional genomics in rice.
In: Rice Genetics IV. Edited by Khush GS, Brar D and Hardy B. IRRI and Science Publishers, Inc. pp. 263-277

Greco R, Ouwerkerk PBF, Sallaud C, Kohli A, Colombo L, Puigdomènech P, Guiderdoni E, Christou P, Hoge JHC, and Pereira A (2001)
Transposon Insertional Mutagenesis in rice.
Plant Physiology 125: 1175-1177

Greco R, Ouwerkerk PBF, Taal AJC, Favalli C, Beguiristain T, Puigdomènech P, Colombo L, Hoge JHC and Pereira A (2001)
Early and multiple *Ac* transpositions in rice suitable for efficient insertional mutagenesis.
Plant Molecular Biology 46: 215-227

Kohli A., Xiong J, Greco R, Christou P and Pereira A (2001)
Tagged Transcriptome Display (TTD) in indica rice using *Ac* transposition.
Molecular Genetics and Genomics 266: 1-11

Marsch Martinez N, Greco R, Van Arkel G, Herrera Estrella L, Pereira A (2002)
Activation tagging using the *En-I* maize transposon system in Arabidopsis.
Plant Physiology 129(4):1544-1556

Cotsaftis O, Sallaud C, Breitler JC, Meynard D, Greco R, Pereira A, Guiderdoni E (2002)
Transposon-mediated generation of T-DNA- and marker-free rice plants expressing a Bt endotoxin gene.
Molecular Breeding 10(3): 165-180

Greco R, Ouwerkerk PBF, De Kam RJ, Sallaud C, Favalli C, Colombo C, Guiderdoni E, Meijer AH, Hoge JHC and Pereira A (2003)
Transpositional behaviour of an *Ac/Ds* system for reverse genetics in rice.
Theoretical and Applied Genetics - DOI: 10.1007/s00122-003-1416-8

Greco R, Ouwerkerk PBF, Taal AJC, Sallaud C, Guiderdoni E, Meijer AH, Hoge JHC and Pereira A (2003)
Transcription and transposition of the maize *En/Spm* transposon system in rice.
(submitted)

Greco R, Ouwerkerk PBF and Pereira A (2003)
Influence of maize *En/Spm* transposase expression in rice on endogenous homologous transposons.
(submitted)

Greco R, Marsch Martinez N, Dixit S, Aharoni A and Pereira A (2003)
Higher order transcriptional regulation conferred by the *bountiful* gain-of-function mutant.
(in preparation)

Marsch Martinez N, Greco R, de Folter S, Aharoni A and Pereira A (2003)
Gain-of-function mutations affecting leaf architecture reveal integrated pathways.
(in preparation)

COVER AND INVITATION DESIGN BY NAYELLI MARSCH MARTINEZ ☺