Plant Molecular Biology **46:** 215–227, 2001.

Early and multiple *Ac* transpositions 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, P.O. Box 16, 6700 AA, Wageningen, Netherlands (*author for correspondence; e-mail: A.Pereira@plant.wag-ur.nl);* [2]*Institute of Molecular Plant Sciences, Leiden University, P.O. Box 9505, 2300 RA Leiden, Netherlands;* [3]*Department of Biology, University of Milan, Via Celoria 26, 20133 Milano, Italy;* [4]*Institut de Biologia Molecular de Barcelona, CID-CSIC, Jordi Girona 18, 08034 Barcelona, Spain*

Received 10 July 2000; accepted in revised form 19 February 2001

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

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 characterization 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.

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 the contemporary genomics revolution, reverse genetics strategies are being developed with 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 with appropriate insertional mutagenesis tools.

The nucleotide sequence data reported will appear in the GenBank database under the accession numbers from AZ 92 3008 to AZ 923053.

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.

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 into 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 utilized 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 by Pereira, 1998). Transposon excision can be selected for by resistance to an antibiotic such as kanamycin (Baker *et al.*, 1987) or visualized 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, 1997; 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* has 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 multipoint 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 CaMV 35S 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 CaMV 35S 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 humanized 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, Netherlands) which contains a chimaeric CaMV 35S-hygromycin phosphotranferase-tNos for selection during transformation.

The 35S-smGFP:*Ac* construct was assembled by ligating the CaMV 35S promoter (*Hin*dIII-*Sal*I frag-

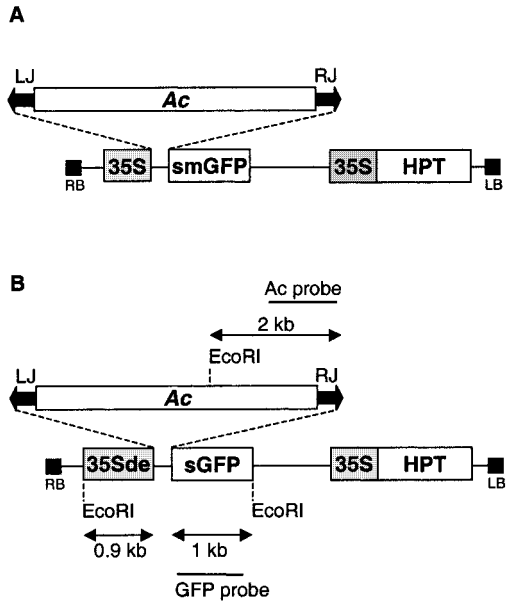

Figure 1. Schematic representation of 35S-smGFP:*Ac* (A) and 35Sde-sGFP:*Ac* (B). In both constructs the autonomous *Ac* element is inserted between the CaMV 35S promoter and the GFP gene used as excision markers so that expression of GFP is restored upon *Ac* excision. The orientation of *Ac* is such that the endogenous promoter is located near the CaMV 35S 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 phosphotransferase gene; 35S, CaMV 35S promoter; 35Sde, doubly enhanced CaMV 35S promoter; Rj, Lj, *Ac* right and left junction.

ment), 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.

Agrobacterium *transformation of rice*

Embryogenic calluses 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 calluses 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 h light, 28 °C, 85% relative humidity (RH) and 12 h dark, 21 °C, 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 calluses 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 with a colour video camera (Sony-iris with integration unit, Sony DKR700) and processed with Adobe Photoshop.

Molecular analysis of Ac *transposition*

Leaf samples (1–2 per plant) were collected either *in vitro* from 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 with 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) with 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′-GAACAAAAATACCGGTT CCCGTCC) 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′-ACGACTCCATTCCTCAGATG ACG; and *Ac*1395-reverse: 5′-CTTGACTCGGATCT GTAGCTGTACC). All PCR reactions were performed with 50 ng of genomic DNA and standard amplification conditions.

DNA blot hybridization

3 μg of genomic DNA was restricted with *Eco*RI, separated overnight on a 0.8% agarose gel in 1× TAE at 4 °C and vacuum-blotted onto Hybond-N^+membranes. Hybridization 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 to 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). Genomic DNA (200–400 ng) was restricted separately with *Hae*III or *Nla*III, self-ligated and linearized again with *Bam*HI. Flanking DNA was then recovered by PCR amplification using two *Ac*-specific primers (*Ac*44-reverse: 5′-GATAACGGTCGGTACGGGAT; *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 with 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 calluses examined for GFP fluorescence. In the first transformation experiment, using the 35S-smGFP:*Ac* construct, no GFP activity was visible among the hygromycin-selected calluses. 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 calluses were obtained, which were selected for GFP activity (Figure 2A). Plants were regenerated from these calluses and showed GFP expression in most tissues examined. Due to the selection procedure, we did not expect to see variegation for GFP activity. To visualize 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 calluses yielded multiple regenerants which were grown in the greenhouse until 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.

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 hybridization 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

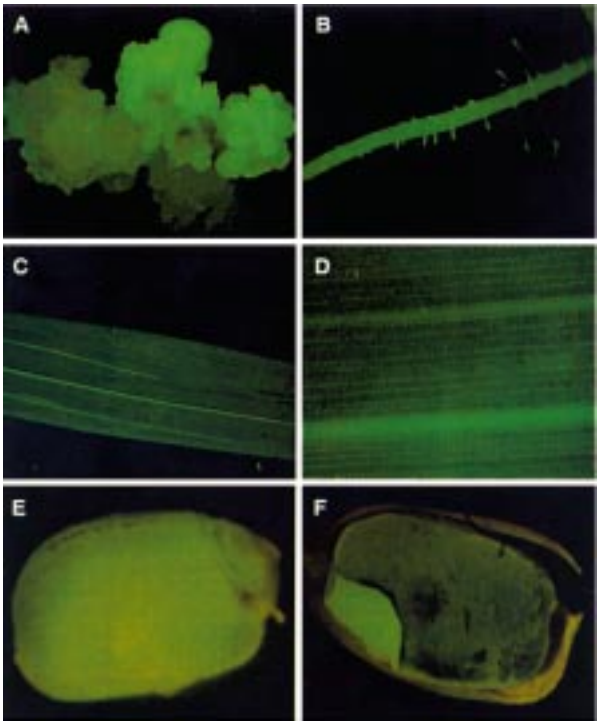

Figure 2. sGFP expression after *Ac* excision. GFP fluorescence was detected in regenerating calluses (A), roots (B) and leaves of regenerated plants (C, D), T_1 seeds (E) and embryos (F).

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 Materials 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.

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 hybridization (data not shown), suggesting a 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.

The second transformation experiment (with 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 the presence of an *Ac* element by PCR with internal primers. The absence of an FDS in all regenerants could be accounted due to the calluses being selected for GFP activity.

A more detailed analysis was performed on the regenerants not selected for GFP activity from the third transformation experiment. DNA blot hybridization 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 hybridization, shown in Figure 4, revealed only the 1.9 kb EDS expected

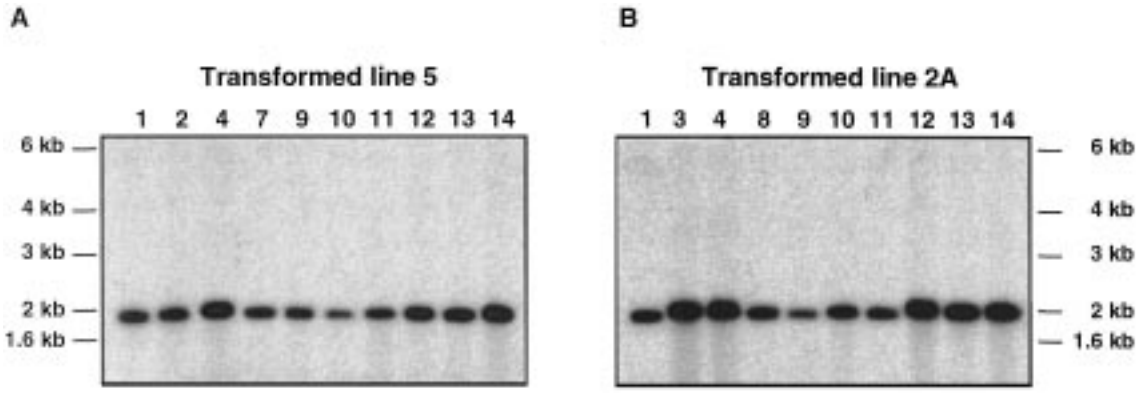

Figure 4. Excision analysis in regenerants from construct 35Sde-sGFP:*Ac*. DNA blot hybridization of 10 regenerants of line 5 (A) and line 2A (B) digested with *Eco*RI and probed with the sGFP gene. *Ac* excision reveals an EDS fragment of 1.9 kb, between *Eco*RI sites upstream of the 35S promoter and 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*.

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

Transformed line	Regenerants			Progeny		
	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. In parenthesis: the number present in the parental plant.
[b] The ITF is calculated as the percentage of unique inserts amongst the total number of inserts.

fragment and no FDS. *Ac* was still present in 70.7% (29/41) of the regenerants demonstrating the complete transposition process.

To characterize 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.

Multiple transpositions in regenerants

The regenerants from the third transformation experiment, not selected for GFP activity, were analysed further by DNA blot hybridization 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*.

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 (Table 1) varies between 49% and 64%. The regenerants from line 5 with a single *Ac* element had 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 calluses 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 hybridization 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.

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 hybridization with an *Ac* probe (Figure 5). Hybridization of the same blot with a T-DNA probe revealed plants carrying the T-DNA and enabled co-segregation 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* hybridizing fragments and 27% with only the strongly hybridizing *Ac* fragments, potentially inherited from the parent as germinal inserts. A 3.4 kb *Ac* homologous fragment co-segregated with the T-DNA, while two other inherited *Ac* elements were present in most progeny and some plants carried extra 1–3 additional 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. The sequences, their origin and homologues in databases are summarized 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 number 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 among a number of transposed *Ac* inserts was revealed by a set of ITSs homologous to a contiguous genomic sequence in a PAC clone (accession number 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, summarized 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 the basis of homology to proteins or ESTs. ITSs in predicted genes like *S*-adenosyl-L-methionine:salicylic methyltransferase or Cyt-5 DNA methyltransferase can provide information on the unique function of these genes in rice. Interesting mutants are being followed up in

Table 2. Insertion tagged sites with similarity to sequences in public databases.

Plant	ITSs[a]	Blast homology	Organism	Accession numbers	Insertion	Identity (bp)	Similarity (%)
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]	5[th] 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*	AU056910 BF430546/ AAD21461[f]	5′-UTR	185/187 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	[74%]
						112/134 bp	[83%]
		Genomic clone/Hypothetical protein[g]	*O. sativa*	AQ690576AP002537/ BAB16867[g]	5′-UTR	266/281 bp	[94%]
						245/270 bp	
5-7	3	Unknown protein/Genomic clone	*A. thaliana/O. sativa*	AAC77868/AQ074191	ORF	66/97 aa	[67%]
						291/303 bp	[96%]
		S-adenosyl-L-methionine:salicylic methyltransferase	*A. thaliana*	AAF16558BAB08594	ORF	5371/90 aa	[78%]
5-11	23	Genomic clone/Ribosome apurinic lyase[g]lyase[h]	*O. sativa/T. aestivum*	AQ691029/ BAA87875[g]BAA87875[h]	3′-UTR 3′-UTR?	268/280 bp	[95%]
5-12	3	EF-hand cont. protein/EST	*A. thaliana/O. sativa*	AAD48968 BAB02809/	ORF	111/125 aa	[88%]
				C98851		210/226 bp	[92%]
		EST	*O. sativa*	C99498	–	171/176 bp	[97%]
7-1	6	Cyt-5 DNA methyltransferase (ZMET1)/EST	*Z. mays/ T. aestivum*	AAC16389 AAG15406/	ORF	42/42 aa	[100%]
				AW352710BE400695		110/126 bp	[827%]
		EST/Ribosomal protein[d,h,i]	*O. sativa/A. thaliana*	AU031084/	3′-UTR		
				AAB71459[h]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,h,i]	*O. sativa/ A. thaliana*	AU031084/	3′-UTR		
			AAB71459[h]AAB71459[i]	3′-UTR?		62/63 bp	[98%]
		Genomic clone/Hypothetical protein[i] protein[l]	*O. sativa/A. thaliana*	AQ688452/	ORF	59/59 bp	[100%]
				AAD32789[i] AAD32789[l]			
		Genomic clone/Hypothetical protein[m]	*O. sativa*	AP002843/BAB17166[m]	ORF	52/52 bp	[100%]
7-3	6	Hypothetical protein/Genomic clone Genomic clone	*O. sativa*	BAB17166/AP002843	ORF	14/16 aa	[98%]
				AQ860963		65/68 bp	[95%]
7-5	4	Genomic clone	*O. sativa*	AQ687097AZ046446	–	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 (PAC clone P0514G12, accession AP000616).
[d]Same *Ac* insertion.
[e,g,m]Homology derived from AP000616 (chr. 6). AP002537 (chr. 1), AP002843 (chr. 1) respectively.
[f]Homology derived from BlastX of AU056910 BF430546 (4279/10061 aa – 10048%).
[g,h]Homology derived from BlastX of AQ691029 (67/109 aa – 60%).
[h,i]Homology derived from BlastX of AU031084 (32/33 aa – 96%).
[i,l]Homology derived from BlastX of AQ688452 (40/76 aa – 52%).

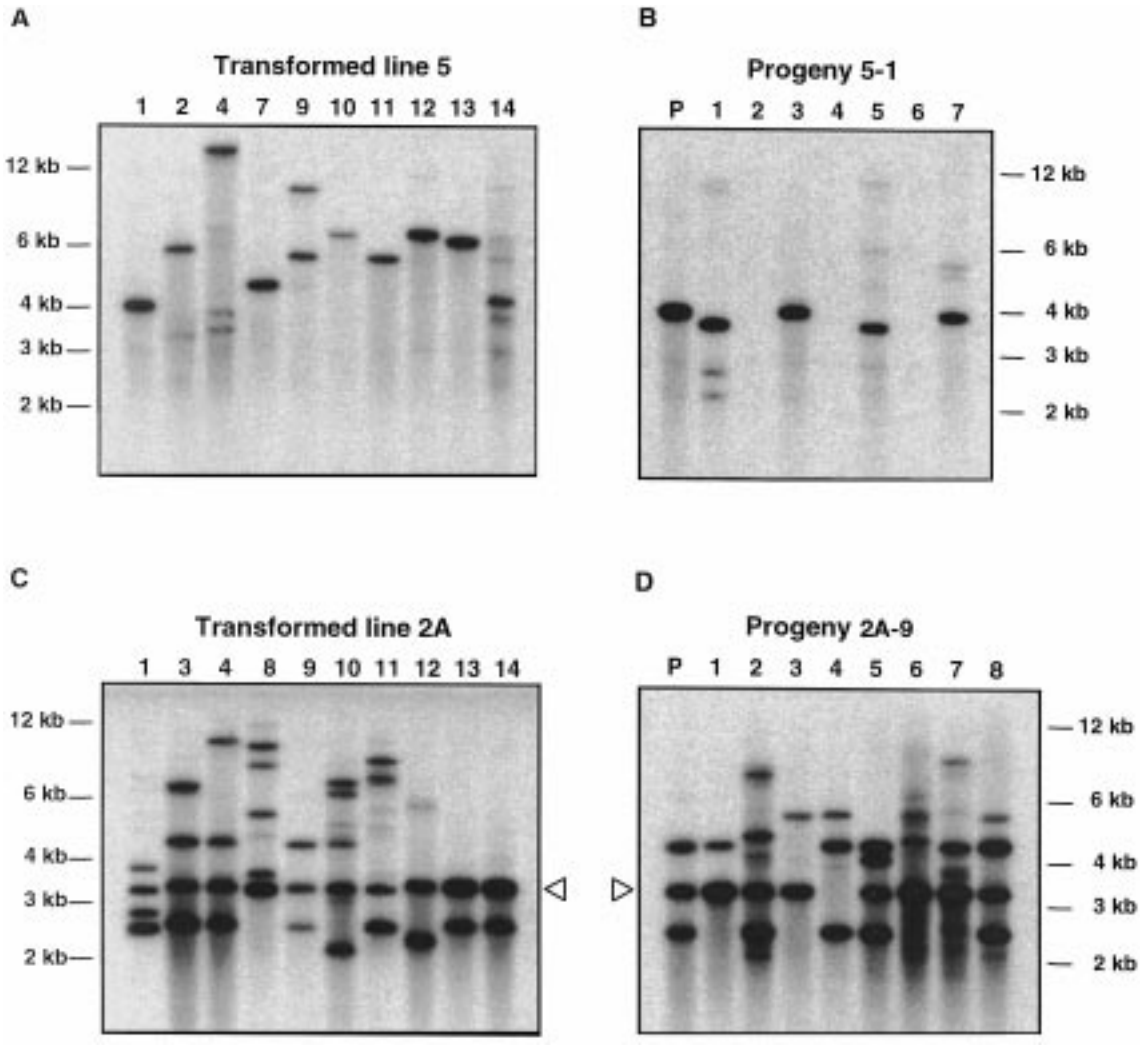

Figure 5. Transposition analysis in different regenerants and progeny plants transformed with construct 35Sde-sGFP:*Ac*. A, C. Blots from regenerants of line 5 and 2A shown in Figure 4 were hybridized with a 1.1 kb *Ac* 3′ region probe. A predicted FDS of 3 kb is not visible but other fragments (larger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

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] In brackets is indicated the number of ITSs longer than 40 bp.
[c] Only the number of identical ESTs is reported (bp identity>95%).

the next generation for phenotypes to validate this reverse genetics strategy.

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 with 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 the effectiveness 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), but 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 after 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 the 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. Re-insertion 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 upon 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 after 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 characterized 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 to 50% in different progeny, 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 is 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 on the basis of 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 (Arabidopsis Genome Initiative, 2000). Of the about 70 000 rice ESTs in the database, about 26 000 are different based on the Gene Index calculations (Quackenbush *et al.*, 2000). 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 a size of 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 on the basis of the two methods described above reveal a consistent bias of about three times more insertions in genes and confirm 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-thirds 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 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 programmes world-wide involved in rice functional genomics.

Acknowledgements

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

References

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. Nucl. Acids Res. 25: 3389–3402.

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

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: 325–330.

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. Nucl. 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: 561–750.

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: 761–876.

Mayer, K., Schuller, C., Wambutt, R. *et al.*, 1999. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. Nature 402: 769–777.

McClintock, B. 1947. Cytogenetic studies of maize and *Neurospora*. Carnegie Inst. Wash. Yearb. 46: 146–152.

Murai, N., Li, Z.J., Kawagoe, Y. and Hayashimoto, A. 1991. Transposition of the maize *activator* element in transgenic rice plants. Nucl. Acids Res. 19: 617–626.

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., 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 Totowa, NJ, 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. Nucl. 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. Nucl. 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 Tissue Organ Cult. 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: 354–360.

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. Bio/technology 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.

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.