Plant Mol Biol
DOI 10.1007/s11103-008-9315-7

Characterisation of 3′ transgene insertion site and derived mRNAs in MON810 YieldGard® maize

Alessio Rosati · Patrizia Bogani · Alisa Santarlasci · Marcello Buiatti

Received: 31 July 2007 / Accepted: 16 February 2008

Abstract The construct inserted in YieldGard® MON810 maize, produced by Monsanto, contains the CaMV 35S promoter, the *hsp70* intron of maize, the *cryI(A)b* gene for resistance to lepidopterans and the NOS terminator. In a previous work a truncation event at the 3′ end of the *cryI(A)b* gene leading to the complete loss of the NOS terminator was demonstrated. The 3′ maize genome junction region was isolated in the same experiment not showing any homology with known sequences. The aim of the experiments here reported was therefore to isolate and characterize a larger portion of the 3′ integration junction from genomic DNA of two commercial MON810 maize lines. Specific primers were designed on the 3′ integration junction sequence for the amplification of a 476 bp fragment downstream of the sequence previously detected. In silico analysis identified the whole isolated 3′ genomic region as a gene putatively coding for the HECT E3 ubiquitin ligase. RT-PCR performed in this region produced cDNA variants of different length. In silico translation of these transcripts identified 2 and 18 putative additional aminoacids in different variants, all derived from the adjacent host genomic sequences, added to the truncated CRY1A protein. These putative recombinant proteins did not show homology with any known protein domains. Our data gave new insights on the genomic organization of MON810 in the YieldGard® maize and confirmed the previous suggestion that the integration in the genome of maize caused a complex recombination event without, apparently, interfering with the activity of the partial CRY1A endotoxin and both the vigor and yield of the YieldGard® maize.

Keywords Cry-hect recombinant mRNAs · 3′ insertion site · HECT protein ligase · YieldGard® MON810 maize

EMBL Accession #: AM749995; AM749996; AM749997; AM749998; AM750007.AM749999; AM750000; AM750001; AM750002; AM750003; AM750004; AM750005; AM750006.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-008-9315-7) contains supplementary material, which is available to authorized users.

A. Rosati · P. Bogani (✉) · A. Santarlasci · M. Buiatti
Dipartimento di Biologia animale e Genetica, Università degli Studi di Firenze, Via Romana, 17-19, 50125 Firenze, Italy
e-mail: patrizia.bogani@unifi.it

Introduction

In modern plant transformation biotechniques, one of most relevant putative sources of uncertainty and unpredictability in the production of new transgenic crop plants is the vector integration site. In fact, the transgene insertion site cannot be predetermined (Pawlowski and Somers 1996, 1998; Tinland 1996; Somers and Makarevitch 2004; Tzfira et al. 2004). For this reason, transgenes may be inserted in functional genomic regions, disrupting the structure and/or altering the regulation patterns of genes from the plant host genome. Results obtained from large-scale T-DNA tagging experiments in *Arabidopsis thaliana* (Szabados et al. 2002; Alonso et al. 2003; Forsbach et al. 2003; Qin et al. 2003) and *Oryza sativa* (Jeong et al. 2002, 2006; Chen et al. 2003; Sha et al. 2004; Zhang et al. 2007) showed a non-random distribution of T-DNA insertion sites, more frequent in genic sequences endowed with different functions such as metabolism, signal transduction and transcription, disease processes and defence mechanisms, intracellular traffic. Experiments in other organisms, namely the legume *Medicago truncatula* (Scholte et al.

2002), barley (Salvo-Garrido et al. 2004), potato and tobacco (Koncz et al. 1989; Lindsey et al. 1993) also suggested that T-DNA insertion occurred generally into coding sequences. Particle bombardment is another plant transformation technique broadly applied in genetically modified plants production. Many studies have shown that it normally creates extremely complex insertion events with the presence of multiple copies of intact or fragmented transgene DNA (Maqbool and Christou 1999; Kohli et al. 1999, 2003; Fu et al. 2000; Mehlo et al. 2000; Svitashev et al. 2000; Svitashev and Somers 2001, 2002; Breitler et al. 2002; Loc et al. 2002; Vain et al. 2002). However it should be stressed that genome rearrangements also often occur in non transgenic maize plants and have been thoroughly studied by several authors (see for instance Dunning Hotopp et al. 2007; Morgante et al. 2005; Brunner et al. 2005). In many instances whole loci have been spontaneously modified also due to the high amount of mobile elements in maize genome. So even in commercially approved genetically modified cultivars, selected for single insertion events, genome rearrangements and the presence of foreign DNA sequences have been detected (Fitch et al. 1992; Windels et al. 2001; Hernandez et al. 2003).

The interest for the analysis of the junction regions in transgenic plants is also due to their event-specific nature, a critical feature for the GM products traceability.

The *Zea mays* event MON810 (US 2004/0180373 A1, 16 September 2004), commercialized by Monsanto with the trade name YieldGard®, is one of the major maize GM products approved for commercialization in the European Union. It contains a transgene cassette carrying the cauliflower mosaic virus 35S promoter, the *hsp70* intron and the *cryIA*(*b*) gene for resistance to lepidopterans, particularly the European corn borer. The first report by Monsanto suggested the presence in MON810 of a unique transgene element of 3.6 kb corresponding to a truncated version of the *cryIA*(*b*) gene inserted into the maize genome. Further studies (Hernandez et al. 2003) confirmed the truncation event of *cryIA*(*b*) gene between positions 2235 and 2571, with the complete loss of the NOS terminator.

The aim of our work was: (i) the isolation and identification of the 3′ transgene junction region in the MON810 maize genome to precisely characterize the genomic organization at this locus and (ii) the study of its transcriptional activity to investigate if the truncation of the expression cassette (Hernandez et al. 2003) could eventually determine read-through transcription of the plant genomic region downstream the truncated cry sequence. For these purposes, two MON810 maize lines, PR33P67 and Elgina and the corresponding isogenic untrasformed lines, respectively PR33P66 and Cecilia, were used and a genome walking approach was applied to isolate the maize genomic sequences flanking the 3′ site of the MON810 event integration locus. Furthermore we performed a detailed in silico analysis of the isolated sequences for structure and function predictions. PCR experiments were finally carried out to confirm in silico predictions.

Materials and methods

Plant materials and DNA extraction

Seeds from PR33P67 and Elgina MON810 maize lines (Piooneer Hi-Bred), and their untransformed parent lines, respectively PR33P66 and Cecilia (www.oecd.org/dataoecd/1/48/33999570.PDF), were kindly provided by Dr. G. Monastra (Istituto Nazionale di Ricerca per gli Alimenti e la Nutrizione, I.N.R.A.N., Rome, Italy). Seeds were sown on a soil-vermiculite mixture and grown in a greenhouse at 24 ± 1°C *T*, under long day light conditions for 2 weeks after germination. DNA was extracted from frozen leaves of each germinated plant as described in Bogani et al. (1995).

Isolation of 3′ junction flanking sequence

3′ plant-vector junction sequence was isolated by GenomeWalker™ technique (Siebert et al. 1995). Enzymes, adapters and adapters primers (AP1, AP2) were provided by GenomeWalker Universal Kit™ (Clontech, BD Bioscience, Becton, Dickinson and Co., Europe). Specific primers were designed on the sequence identified by Hernandez et al. (2003) and deposited in GenBank with the Accession # AF490398, using Primer 3 program (Rozen and Skaletsky 2000). Primer sequences are listed in Table 1. A first PCR amplification was performed with the adapter-specific primer AP1 and MONGW1fwd. The nested PCR amplification was then performed using the adapter-specific primer AP2 and MONGW2fwd primer, using as template 1 µl of 1:50 dilution of the first amplification reaction, according to the manufacturer's instructions. The PCR product was then ligated in the TOPO-TA cloning vector (pCR2.1-TOPO®, Invitrogen), and cloned into *E. coli* competent cells. Sequencing of inserts (both strands from three clones) was done by MWG-Biotech, Münich (www.mwg-biotech.com). The sequence obtained appeared in the EMBL databank under the Accession #: AM749995; AM749996; AM749997; AM749998; AM750007.

PCR conditions

Conventional PCR reactions were performed in a final volume of 25 µl, containing 100 ng genomic DNA as template, 5 µmol dNTPs, 10 pmol of each primer (Table 2), 1U Taq DNA polymerase (Amersham Biosciences, Italy), 1.5 mM $MgCl_2$ and 1× PCR buffer

Table 1 Primers used for PCR amplification of DNA and cDNA fragments

Primer	Sequence	Primer position with respect to 3′ junction site (+1)	Localization within BAC clone ZMMBBc0409B05
CRYfwd	TCT TCA CGT CCA GCA ATC AG	−570/−550	–
3′GENrev	TTC TCA CGT CCA GCA ATC AG	+231/+251	130143/130123
HECTupRev	TTT GAG GTG AGC CTT GCC GAT G	+1992/+2012	131903/131883
HECTupRev2	ATT TGT GAG GGA AGG TGT CG	+1245/+1265	131157/131137
HECTupRev3	TTT GGG AAG GAA AAG GTA TC	+656/+676	130567/130547
HECTExfwd	TCA ATC ATC AAA GCA TCA TCG	−826/−806	129056/129076
MONGW1fwd	CTA CTT GTA CCA GAA GAT CGA TGA GTC	−254/−227	–
MONGW2fwd	ATA CCA ATC CTA CTA GAC TGG CTG AG	+467/+493	130358/130384

(Amersham). PCR was carried out in a PTC-100™ P Thermal Cycler (MJ Research, INC.) with the following cycling parameters: 95°C for 10 min, 30 cycles at 95°C for 1 min, *T* annealing (T_a) for 1 min and 72°C for different extension times depending on the fragment size, followed by a final extension at 72°C for 8 min. T_a and extension times are reported in Table 2.

RNA isolation and purification

Leaves from 15-days old plants were harvested and immediately frozen in liquid nitrogen prior RNA isolation. 500 mg of fresh tissue were then homogenized with a pestle and mortar and suspended in the extraction buffer provided by the Plant RNA extraction kit (Macherey–Nagel, M-Medical, Italy) following instructions of the protocol supplied. RNA obtained was treated with 0.1 U/µl RNase DNase I free (Roche Diagnostics, Italy) overnight at 37°C. Total RNA was then precipitated with 0.1 V LiCl 4 M and 3 V absolute ethanol, washed twice in 70% ethanol and resuspended in sterile distilled water. RNA was then spectrophotometrically quantified at 260 nm and stored at −80°C until further use. To check for DNA contamination, 1 µg total RNA was used as template in a PCR amplification reaction.

Reverse-transcription PCR

cDNAs were reverse transcribed from 1 µg total RNA from MON810 and isogenic untransformed maize lines according to the First Strand cDNA Synthesis Kit for RT-PCR (AMV) (Roche). Primers used for retrotranscription reaction were HECTupREV, or HECTupREV2 or HECTupREV3 as reverse primers (Table 2; Fig. 4). The same reverse primers were then used in combination with CRYfwd as forward in cDNA PCR amplification. PCR fragments obtained were cut out from the agarose gel, purified using the PCR purification kit (Qiagen, Italy), ligated in pCR®2.1 TOPO-TA kit (Invitrogen) and sequenced as described above. Sequence obtained were deposited in the EMBL databank with the following Accession #: AM749999; AM750000; AM750001; AM750002; AM750003; AM750004; AM750005; AM750006.

Computational data

Sequence comparison and alignment were performed by ClustalW (Thompson et al. 1994). Homology search analysis was done by using the BLAST algorithm (Altschul et al. 1990), (BLASTN version 2.2.15, Oct-15-2006) against *Z. mays* Plant Genome Database (http://www.plantgdb.org/ZmGDB/index.php) for DNA sequences, and against SWISSPROT (release of Jan-15-2008) (http://expasy.org/sprot/), PROSITE (Hulo et al. 2006), (release 20.25 of Jan-17-2008) (http://expasy.org/prosite/), and SMART version 5.1 (Schultz et al. 1998; Letunic et al. 2006) (http://smart.embl-heidelberg.de/index2.cgi) databases, for aminoacid sequences. Putative coding regions

Table 2 Primer pairs and PCR conditions

Primer pair		Annealing temperature (°C)	Extension time	Expected fragment size (bp)
1	CRYfwd/ HECTupRev	54	3′	2587
2	CRYfwd/ HECTupRev2	54	2′	1835
3	CRYfwd/ HECTupRev3	54	1′30″	1246
4	HECTEXfwd/3′GENrev	56	1′30″	1086
5	HECTEXfwd/ HECTupREV3	56	2′	1502

search was performed using GETORF software from EMBOSS GUI on-line suite (release of November 2002) (Rice et al. 2000), and TWINSCAN software (Korf et al. 2001) (http://mblab.wustl.edu/query.html). Splice site prediction analysis was performed with NetPlantGene Server (release of November 2003), a software producing neural network prediction of splice sites in *Arabidopsis thaliana* DNA (http://www.cbs.dtu.dk/services/NetPGene/).

Results

Isolation and molecular characterization of the 3′ junction flanking region

Hernandez et al. (2003) identified a 1222 bp fragment (AF490398) spanning the 3′ junction region in MON810 maize, composed by the last 633 bp of the truncated *cryIA*(*b*) gene and 598 bp corresponding to plant junction DNA. This sequence did not show significant identity with any known plant DNA sequences. To further characterize the 3′ insertion site of MON810 transgenic maize, we analyzed a 345 bp junction flanking region downstream of the previously isolated one. For this aim the genome walking technique was applied using two different primer pairs. Primers were anchored, respectively, one to the *cryIA*(*b*) gene (MONGW1fwd) and the other to the annotated AF490398 3′ junction (MONGW2fwd) (see Sect. 'Materials and methods'). The nested 476 bp long PCR product, amplified from PR33P67 and Elgina transgenic lines genomic DNA, were cloned and sequenced. No fragments were obtained from isogenic untrasformed lines. The isolated 476 bp fragment was shown to overlap with the last 130 base pairs of the known 3′ junction region and to extend downstream of it for further 345 base pairs. Sequence alignment showed, moreover, a lower identity between the two sequences from nucleotide position 100 to 130 (Fig. 1).

The whole isolated 943 bp plant DNA fragment (345 bp isolated in this work and the annotated 598 bp), flanking the 3′ end of MON810 insert was then used as query in the search of homology against the *Z. mays* Plant Genome Database using the BLASTN program. Search results showed a 99% identity with the chromosome 5 BAC clone ZMMBBc0409B05 (Zm-BAC gi:110625153; Accession # AC185641) from nucleotide positions 129935 to 130834. To confirm the identity of the isolated sequence, one forward primer annealing within the *cryIA*(*b*) gene and three different reverse primers annealing in different position with the sequence of the BAC clone ZMMBBc0409B05, were used to amplify in transgenic lines the corresponding regions. PCR experiments with the following primer

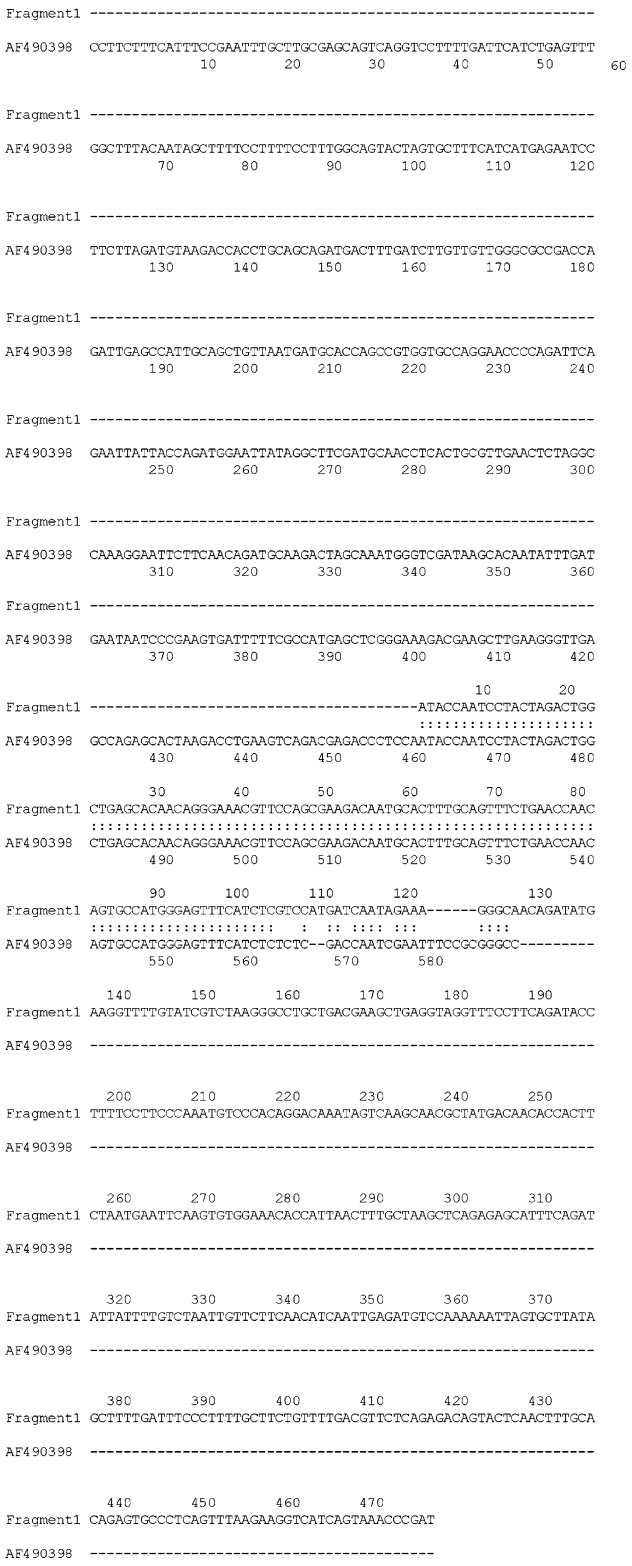

Fig. 1 Sequence alignment between the 476 bp fragment obtained through PCR amplification with the AP2 and MONGW2fw primer pair, and the annotated maize MON810 3′ flanking sequence (AF490398)

combinations, respectively CRYfwd/HECTupREV (1), CRYfwd/HECTupREV2 (2) and CRYfwd/HECTupREV3 (3) (Fig. 2), produced the amplification of products of expected length, respectively 2,585, 1,835, 1,246 bp on PR33P67 and Elgina genomic DNA, no amplified bands being detected in isogenic untransformed lines. All three PCR fragments were cloned and sequenced. Alignments shown in Supplement A confirmed the identity with chromosome 5 BAC clone ZMMBBc0409B05.

In silico analysis of the structure of the 3′ insertion site

The in silico 3′–5′ translation of the 943 bp fragment produced a sequence of 314 aminoacids. Homology search (BLASTp) against SWISSPROT database showed a good level of identity (81%) with a putative HECT ubiquitin ligase from *Oryza sativa* (Accession # Q6YU89) (Supplement B). To determine the general structure of the putative gene in the maize genome we analysed a 9.2 Kbp portion from the BAC clone ZMMBBc0409B05 encompassing the region from 124972 to 134148, containing the 943 bp fragment, with the GETORF software from EMBOSS GUI and TWINSCAN softwares (see Sect. 'Materials and methods'). Results obtained with the two programmes were in agreement and evidenced 17 putative exons (Fig. 2a) coding for a 1711 aminoacid sequence showing 82% identity with the *O. sativa* HECT E3 ubiquitin ligase (Supplement B). The putative gene structure was partially confirmed by splice-site analysis with NetPlantGene Server, recognizing 13 out of 17 potential exon–intron junctions, and 7 out of 17 potential intron–exon junctions with a confidence level ranging from 0.74 to 1.00 (Table 3). Finally, the search for the presence of protein domains in PROSITE (Hulo et al. 2006) and SMART (Schultz et al. 1998; Letunic et al. 2006) databases revealed the presence of the HECT domain at the carboxyl terminus (bold letters in Supplement B). Our data therefore suggested, as shown in Fig. 2b, that the MON810 transgene cassette is inserted in the genome of maize beside exon 8 of a putative HECT gene not reported so far, to our knowledge, in *Z. mays*.

Characterization of the 3′ undisrupted site in isogenic untrasformed maize

PCR reactions performed on DNA extracted from PR33P66 and Cecilia maize isogenic lines using HECTEXfwd/3′GENrev primer pair led to the amplification of a single DNA fragment. No bands were obtained in MON810 transgenic lines. The PCR product was cloned and sequenced, and the alignment between this sequence and *Z. mays* Plant Genome Database identified a 1,086 bp region (Fig. 2a) identical to the BAC clone ZMMBBc0409B05 from nucleotide position 129056 to 131908 (Fig. 3). The alignment evidenced 100% identity between the fragment and the region comprising the last 289 bp of exon 8, the 742 bp of intron 8, and the first 55 bp of exon 9 of the predicted HECT protein ligase gene. Sequence comparison between undisrupted site from isogenic maize and MON810 3′ flanking region, showed that the insertion of MON810 expression cassette occurred within the exon 8 of the putative maize HECT E3 ligase gene, and that the insertion event caused rearrangements also at the first 44 bp of MON810 3′ flanking region (Figs. 2b and 3).

Analysis of the expression of the 3′ junction region

The analysis of the expression of the 3′ MON810-maize genome junction was performed by extracting total RNA

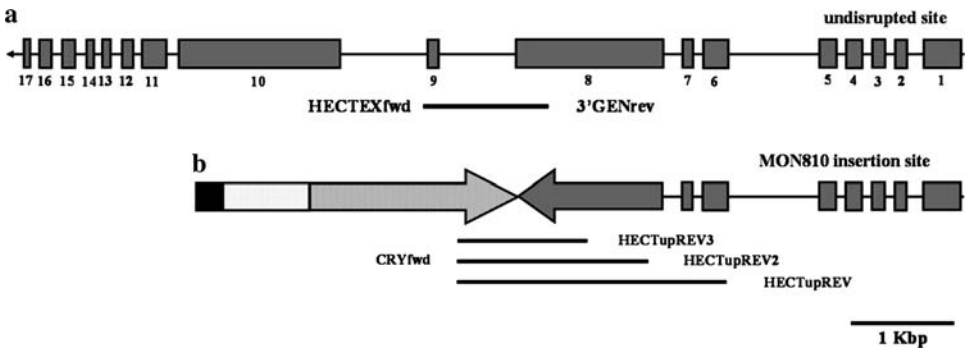

Fig. 2 (**a**) Scheme for the structure of the putative HECT E3 ubiquitin-ligase gene of maize, within the BAC clone ZMMBBc0409B05 from nucleotides 124972 to 134148. Numbers refer to exons of the putative HECT gene; (**b**) insertion site of MON810 vector in the maize genome (*black box*: CaMV35S promoter; *dotted box*: hsp70 intron; *light grey arrow*: *cryIA*(*b*) truncated gene; *dark grey arrow*: truncated exon 8 of putative HECT E3 ubiquitin-ligase gene). *Arrows* indicate the direction of transcription for both the MON810 insert and the endogenous putative maize HECT gene. PCR fragments obtained in this study with the corresponding forward and reverse primers are shown by *black bars*

Table 3 Putative exon boundaries and predicted splice-sites in the BAC clone ZMMBBc0409B05 region from 124972 to 134148 nucleotide positions

Exons[a]				Acceptor splice-sites[b]		Donor splice sites[b]	
	Nucleotide position on BAC clone	Length (bp)	Number of aminoacids	Confidence	Intron–exon junction	Confidence	Exon–intron junction
1	134139–133779	360	120				
2	133618–133507	111	37				
3	133414–133291	123	41			0.92	TGATTCCAAG^GTACTTTGCA
4	133190–133037	153	51			0.90	CACATATACA^GTATGGGCCA
5	132946–132784	162	54	0.94	CGCTTTTCAG^GGCTTAATCC	1.00	TGCAGATCAG^GTGAATATTT
6	131903–131663	240	80	1.00	AATATTGCAG^ATGTTTGAGA	1.00	AATGACACAG^GTTAGTGCTG
7	131566–131467	99	33	0.69	GCTTGTTCAG^GTGTATGGTT		
8	131276–129851	1425	475			1.00	TAGCTCTGAG^GTATTTTTTT
9	129110–129005	105	35	1.00	GTATATGTAG^GATGAAGAGC	1.00	TGATCATGAG^GTAACTGTTT
10	128164–126604	1560	520	0.95	AAATTCTTAG^GTTCTCCAAG	1.00	ACTTGGACAG^GTATGTACAG
11	126485–126251	234	78			0.96	AGGCACAATT^GTAAGTGATG
12	126168–126063	105	35			1.00	GTTTAATCAG^GTACAGCCTC
13	125952–125874	78	26			0.93	AATTTGGGAG^GTAATGTTCT
14	125801–125732	69	23	0.96	GAATCTGCAG^TCGGAATCCC	1.00	GATCGTGAAT^GTAAGCTCAT
15	125637–125508	129	43			1.00	AGTTAGGAAG^GTACCTTTTC
16	125401–125281	120	40			0.74	CTCCACAAAA^GTACAGTGTC
17	125193–125133	60	20	0.87	GTTTGGACAG^GAAATCATGC		
	Total	5133	1711				

[a] Predicted by the GET-ORF and TWINSCAN softwares

[b] Predicted by the NetPlantGene Server software

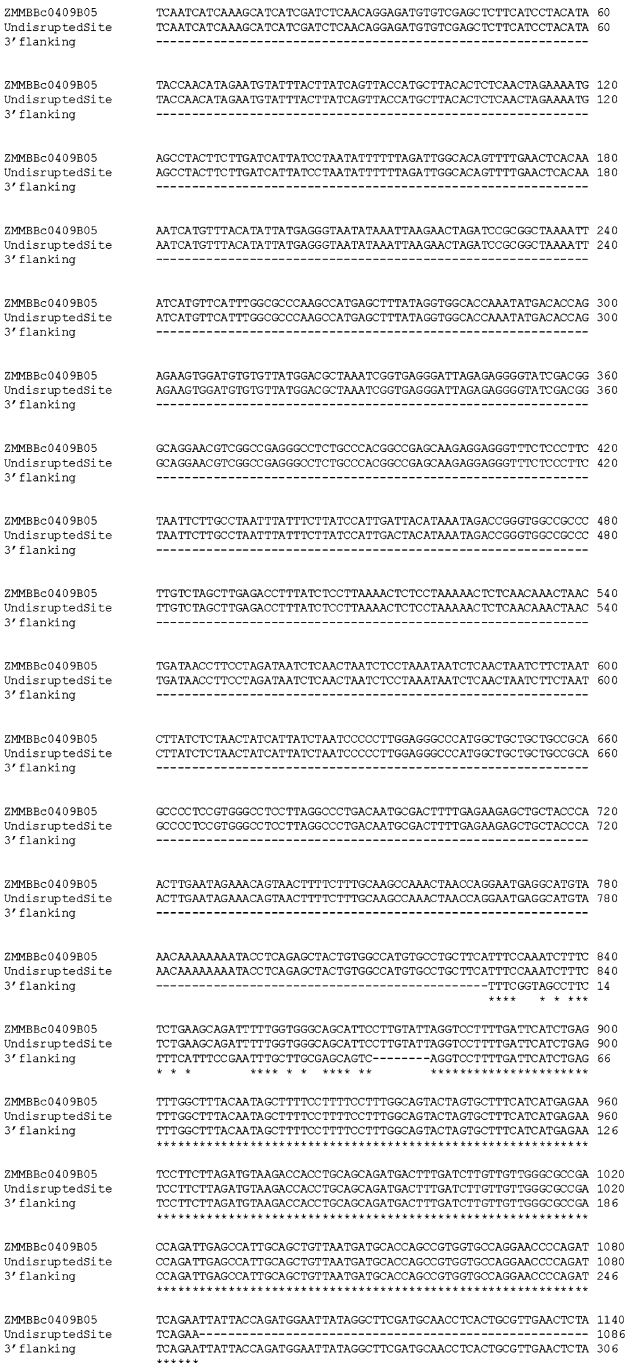

Fig. 3 Comparison of undisrupted site sequence from isogenic untransformed maize lines and MON810 3′ flanking region from transgenic ones obtained in this study, and the BAC clone ZMMBBc0409B05 sequence from nucleotides 129056 to 130195

from transgenic tissues. Total RNA was then retrotranscribed in three different retrotranscription reactions using respectively HECTupREV, HECTupREV2 and HECTupREV3 as reverse primers as described in Sect. 'Materials and methods'. The resulting cDNAs were then used for PCR amplifications. PCR reactions were performed with CRYfwd as forward primer, and with the same reverse primers used in retrotranscription reactions. The amplification of cDNA with CRYfwd/HECTupREV (1), CRYfwd/HECTupREV2 (2) and CRYfwd/HECTupREV3 (3) primers combinations led to the production of a series of fragments with different lengths suggesting the synthesis of different RNA variants. As shown in Table 4, the use of primer combination 3 produced only one fragment of the same length both when cDNA or genomic DNA (gDNA) were used as templates. Primer combination 2 gave four cDNA fragments, one of which had the same length when gDNA and cDNA were amplified. The other three were all different and shorter. Finally, primer combination 1 gave three amplification products all showing smaller sizes than the fragment obtained in gDNA amplification; the lack of this fragment in cDNA amplification may be due to PCR competition with shorter cDNA molecules. All cDNA-PCR products extracted from the agarose gel and purified were ligated in the pCR2.1-TOPO® vector (Invitrogen), cloned in *E. coli* competent cells and sequenced.

Sequencing of each different insert (both strands from three different *E. coli* clones) always showed 100% identity with 569 bp of *cryI(A)b* gene while the remaining region of different length matched with chromosome 5 BAC clone ZMMBBc0409B05. Shorter cDNA products evidenced deletions located in different positions along the sequence but all within the maize plant flanking region (Table 4; Fig. 4; Supplement C). When analyzed with the NetPlantGene Server for splice-site prediction (see sect. 'Materials and methods'), deletion events were found to be similar to spurious exon–intron or intron–exon junctions (Table 5). The in silico translation of different RNA variants in all fragments lacking the specific deletion from nucleotide 5 to nucleotide 102 showed two putative additional aminoacids downstream of the truncated CryI(A)b protein with a stop codon encompassing nucleotide positions 576, 577 and 578 (corresponding to nucleotide positions 280–282 in SEQ ID NO:4 Monsanto Patent US2004/0180373 A1, Pub. Date sp. 16, 2004) (http://www.agbios.com/). On the other hand, mRNA variants showing the deletion from nucleotide 5 to 102, showed 18 putative additional aminoacids added to the truncated CryI(A) protein (Table 4; Fig. 5). The other deletions identified were all localized downstream of the termination codon and would not influence the putative recombinant ORFs.

Discussion

In the last 10 years there has been a growing interest in the detailed analysis of the molecular structure of transgenic loci in GM crops in order to develop markers for traceability of crops and derived products. In several instances

Table 4 Length (bp) of the sequenced PCR and RT-PCR fragments obtained with the same primer pairs

Primer pairs	gDNA amplified fragment (bp)	cDNA amplified fragment (bp)	Deletion nucleotide positions					
			5–102	643–882	643–905	1070–1161	1415–1975	1416–1550
3 CRYfwd/HECTupREV3	1246	1246	−					
2 CRYfwd/HECTupREV2	1835	1835	−	−	−	−		
		1572	−	−	+	−		
		1497	+	+	−	−		
		1405	+	+	−	+		
1 CRYfwd/HECTupREV	2587	2120	−	+	−	+	−	+
		2022	+	+	−	+	−	+
		1688	+	+	−	−	+	−

Positions of deletions, obtained from the comparison of the cDNA and gDNA corresponding sequences, are indicated respectively to the 3′ transgene junction site, considered as nucleotide position 0

+/−, presence/absence of deletion event

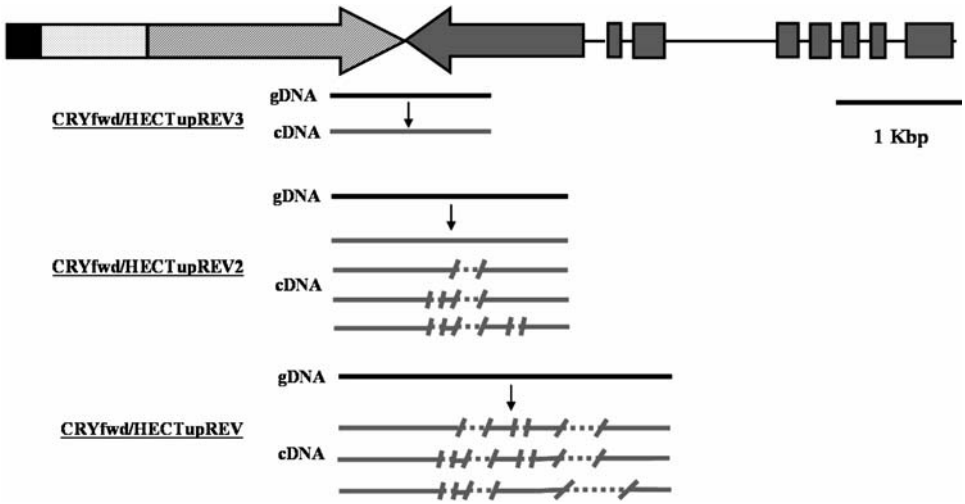

Fig. 4 Scheme for the MON810 3′ insertion site. PCR fragments obtained from both genomic DNA (gDNA, black bars) and RNA (cDNA, grey bars), with the primer pair used are shown. Black box: CaMV35S promoter; dotted box: hsp70 intron; light grey arrow: *cryIA(b)* truncated gene; dark gray arrow: truncated exon 8 of the putative HECT E3 ubiquitin-ligase gene

Table 5 Identification of deletion events in mRNA variants as exon–intron/intron–exon junctions by NetPlantGene server for splice-site prediction

Deletion	Donor splice sites		Acceptor splice sites	
	Confidence	Exon–intron	Confidence	Intron–exon
5–102	0.74	AGGACTTTCG^GTAGCCTTCT	1.00	CCTTTGGCAG^TACTAGTGCT
643–882			0.98	ACGTTCTCAG^AGACAGTATT
1070–1161	1.00	TAACACCAAG^GTTGGAACTA		
1415–1975	0.96	TGGAAAAGAA^GTATGTACAT		
1416–1550			0.85	TCCAGGTGAG^AGAGAGAGAA

the molecular characterization of commercial transgenic crops showed the presence of vector fragments and modifications of the insertion site in the host genome. (Wilson et al. 2006; Fitch et al. 1992). Moreover Roundup Ready® Soybean 40-3-2 showed an intact copy of the CP4 EPSPS transgene followed by a 254 bp CP4 EPSPS fragment (Windels et al. 2001). In this case the transcription of the transgene locus gave rise to four distinct mRNA variants, probably generated by unknown post-transcriptional mechanisms. All these RNA variants coded for putative CP4 EPSPS fusion proteins, each with a redundancy of at least 56 aminoacids from CP4 EPSPS at the N-terminus, and a further region of 24 aminoacids derived from the host genome flanking sequence (Rang et al. 2005).

As mentioned before, in MON810 maize the 3′ junction region, isolated by Hernandez et al. (2003), showed a

```
Rev2-d          ------------------FTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEKKELS 41
Rev-b           ------------------FTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEKKELS 41
Rev-c           ------------------FTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEKKELS 41
Rev2-c          ------------------FTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEKKELS 41
Rev-a           ------------------FTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEKKELS 41
Rev2-b          ------------------FTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEKKELS 41
Rev3-a          ------------------FTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEKKELS 41
Rev2-a          ------------------FTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEKKELS 41
AF490398        VTFEAEYDLERAQKAVNELFTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEKKELS 60
                                   ****************************************

Rev2-d          EKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITIQGGDDVFKENYVTLLGTFDE 101
Rev-b           EKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITIQGGDDVFKENYVTLLGTFDE 101
Rev-c           EKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITIQGGDDVFKENYVTLLGTFDE 101
Rev2-c          EKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITIQGGDDVFKENYVTLLGTFDE 101
Rev-a           EKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITIQGGDDVFKENYVTLLGTFDE 101
Rev2-b          EKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITIQGGDDVFKENYVTLLGTFDE 101
Rev3-a          EKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITIQGGDDVFKENYVTLLGTFDE 101
Rev2-a          EKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITIQGGDDVFKENYVTLLGTFDE 101
AF490398        EKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITIQGGDDVFKENYVTLLGTFDE 120
                ************************************************************

Rev2-d          CYPTYLYQKIDESKLKAYTRYQLRGYIEDSQDLEIYLIRYNAKHETVNVPGTGSLWPLSA 161
Rev-b           CYPTYLYQKIDESKLKAYTRYQLRGYIEDSQDLEIYLIRYNAKHETVNVPGTGSLWPLSA 161
Rev-c           CYPTYLYQKIDESKLKAYTRYQLRGYIEDSQDLEIYLIRYNAKHETVNVPGTGSLWPLSA 161
Rev2-c          CYPTYLYQKIDESKLKAYTRYQLRGYIEDSQDLEIYLIRYNAKHETVNVPGTGSLWPLSA 161
Rev-a           CYPTYLYQKIDESKLKAYTRYQLRGYIEDSQDLEIYLIRYNAKHETVNVPGTGSLWPLSA 161
Rev2-b          CYPTYLYQKIDESKLKAYTRYQLRGYIEDSQDLEIYLIRYNAKHETVNVPGTGSLWPLSA 161
Rev3-a          CYPTYLYQKIDESKLKAYTRYQLRGYIEDSQDLEIYLIRYNAKHETVNVPGTGSLWPLSA 161
Rev2-a          CYPTYLYQKIDESKLKAYTRYQLRGYIEDSQDLEIYLIRYNAKHETVNVPGTGSLWPLSA 161
AF490398        CYPTYLYQKIDESKLKAYTRYQLRGYIEDSQDLEIYLIRYNAKHETVNVPGTGSLWPLSA 180
                ************************************************************

Rev2-d          PSPIGKCAHHSHHFSLDIDVGCTDLNEDFRTSAFIMRILLRCKTTCSR 209
Rev-b           PSPIGKCAHHSHHFSLDIDVGCTDLNEDFRTSAFIMRILLRCKTTCSR 209
Rev-c           PSPIGKCAHHSHHFSLDIDVGCTDLNEDFRTSAFIMRILLRCKTTCSR 209
Rev2-c          PSPIGKCAHHSHHFSLDIDVGCTDLNEDFRTSAFIMRILLRCKTTCSR 209
Rev-a           PSPIGKCAHHSHHFSLDIDVGCTDLNEDFR------------------ 191
Rev2-b          PSPIGKCAHHSHHFSLDIDVGCTDLNEDFR------------------ 191
Rev3-a          PSPIGKCAHHSHHFSLDIDVGCTDLNEDFR------------------ 191
Rev2-a          PSPIGKCAHHSHHFSLDIDVGCTDLNEDFR------------------ 191
AF490398        PSPIGKCAHHSHHFSLDIDVGCTDLNEDFR------------------ 210
                ******************************
```

Fig. 5 Protein sequence alignment of in silico translated PCR Rev3-a, Rev2-a,b,c,d, Rev-a,b,c fragments, amplified from cDNA respectively with CRYfwd/HECTupREV3, CRYfwd/HECTupREV2, CRYfwd/HECTupREV primer combinations and aminoacid sequence, deduced from in silico translation of the annotated AF490398 3′ junction region

truncation event of *cryIA*(*b*) gene between positions 2,235 and 2,571, with the complete loss of the NOS terminator. Furthermore these authors were unable to amplify in non-transgenic maize the undisrupted insertion site using primers designed in the 3′ junction, and in the 5′ junction (AF090447), this latter being 88% homologous with LTR sequences of the *Z. mays* 22 kDa alpha zein gene cluster (Holck et al. 2002). These results suggested that the integration of the MON810 vector has probably caused a complex recombination event.

To further characterize the 3′ insertion site of MON810 in the genome of maize, in our work a 2 Kbp region downstream the MON810 truncated cassette was analysed in PR33P67 and Elgina transgenic lines, and compared with the corresponding control site in isogenic untransformed PR33P66 and Cecilia. For this purpose we sequenced 345 additional bp in the flanking region downstream of that previously isolated by Hernandez et al. (2003). In silico analysis against the *Z. mays* database of the whole 3′ genomic sequences (345 bp isolated in this work and the annotated 598 bp) showed a 99% identity with positions 129,935–130,834 of a maize chromosome 5 BAC clone. This sequence did not show any identity with the 5′ annotated junction (Holck et al. 2002), thus suggesting that the two MON810 flanking regions do not belong to the same DNA locus. On the other hand, the same sequence showed an 82% identity with *O. sativa* locus coding for a putative HECT E3 ubiquitin ligase. E3 ubiquitin ligase proteins are important regulators of different physiological processes both in animals and plants, involving all aspects of the life cycle including cell division, embryogenesis, photomorphogenesis, senescence, disease resistance and hormone responses (Hellmann and Estelle 2002; Vierstra 2003; Downes et al. 2003). To our knowledge, no HECT E3 ligase gene has been so far identified in the genome of *Z. mays*. Our analysis allowed to localize this gene in the chromosome 5 BAC clone ZMMBBc0409B05. The search for protein domains in PROSITE and SMART databases revealed the presence of a HECT domain in the deduced aminoacid sequence. Thus, the characterization of the 3′ insertion site in MON810 maize transgenic lines suggested that the transformation process may have involved the truncation of the 3′ end of putative HECT endogenous gene leading to the partial loss

off the inserted fragment and to a deletion of adjacent genomic sequences without, apparently, interfering with the activity of the partial CRY1A endotoxin and both the vigor and yield of the YieldGard® maize. Further experiments were then carried out on the expression of the analysed sequences using forward primers on the *cry1A*(*b*) gene and reverse primers on the genomic region downstream of the 3′ vector end. The experiments showed the synthesis, due to the known loss of the NOS terminator, of fusion RNA variants coding for putative CRY fusion protein showing 2 or 18 putative additional aminoacids, composed of the 3′ end of truncated *cry1A*(*b*) gene and the putative HECT 3′ sequences transcribed in antisense orientation. Different transcripts showed a series of deletions some of which in common to all observed RNAs (Fig. 4; Table 4). The mRNA rearrangements observed could be probably due to the presence of spurious exon–intron and intron–exon junctions as evidenced by results reported in Table 5. In silico translation of putative fusion RNAs did not show significant identities with known protein domains. Our data however are partially consistent with Monsanto's Patent US 2004/0180373 A1 (Pub. Date sp. 16, 2004) where Monsanto claimed that the first nine nucleotides of the 3′ flanking plant DNA immediately adjacent to the *cry1A*(*b*) truncation event coded for two aminoacids and a stop codon. Taken together, our data, while suggesting the insertion of the transgenic sequence in a putative HECT gene, show the transcription of new fusion RNAs, a result similar to that obtained by Rang and et al. (2005) in Roundup Ready® Soybean 40-3-2.

Acknowledgments We thank Dr. G. Monastra for providing seeds of MON810 and isogenic control maize. This work was supported by a grant from MIPAF (Ministero delle Politiche Agricole, Alimentari e Forestali), Project: "OGM in Agricoltura".

References

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseeuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301(5633):653–657

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Bogani P, Simoni A, Bettini P, Mugnai M, Pellegrini MG, Buiatti M (1995) Genome flux in tomato auto- and auxotrophic cell clones cultured in different auxin/cytokinin equilibria. I. DNA multiplicity and methylation levels. Genome 38:902–912

Breitler JC, Labeyrie A, Meynard D, Legavre T, Guiderdoni E (2002) Efficient microprojectile bombardment-mediated transformation of rice using gene cassettes. Theor Appl Genet 104:709–719

Brunner S, Pea G, Rafalski A (2005) Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize. Plant J 43(6):799–810

Chen S, Jin W, Wang M, Zhang F, Zhou J, Jia Q, Wu Y, Liu F, Wu P (2003) Distribution and characterisation of over 1000 T-DNA tags in rice genome. Plant J 36:105–113

Downes BP, Stupar RM, Gingerich DJ, Vierstra RD (2003) The HECT ubiquitin-protein ligase (UPL) family in *Arabidopsis*: UPL3 has a specific role in trichome development. Plant J 35:729–742

Dunning Hotopp JC, Clark ME, Oliveira DCS, Foster JM, Fischer P, Muñoz Torres MC, Giebel JD, Kumar N, Ishmael N, Wang S, Ingram J, Nene RV, Shepard J, Tomkins J, Richards S, Spiro DJ, Ghedin E, Slatko BE, Tettelin H, Werren JH (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317:1753–1756

Fitch MM, Manshardt RM, Gonsalves D, Slightom JL, Sanford JC (1992) Virus resistant papaya plants derived from tissues bombarded with the coat protein gene of papaya ringspot virus. Biotechnology 10:1466–1472

Forsbach A, Schubert D, Lechtenberg B, Gils M, Schmidt R (2003) A comprehensive characterisation of single-copy T-DNA insertions into the *Arabidopsis thaliana* genome. Plant Mol Biol 52:161–176

Fu X, Duc LT, Fontana S, Bong BB, Tinjuangjun P, Sudhakar D, Twyman RM, Christou P, Kohli A (2000) Linear transgene constructs lacking vector backbone sequences generate low-copy-number transgenic plants with simple integration patterns. Transgenic Res 9:11–19

Hellmann H, Estelle M (2002) Plant development: regulation by protein degradation. Science 297:793–797

Hernandez M, Pla M, Esteve T, Prat S, Puigdomenech P, Ferrando A (2003) A specific real-time quantitative PCR detection system for event MON810 in maize YieldGard® based on the 3′-transgene integration sequence. Transgenic Res 12:179–189

Holck A, Va M, Didierjean L, Rudi K (2002) 5′-nuclease PCR for quantitative event-specific detection of the genetically modified Mon810 MaisGard maize. Eur Food Res Technol 214:449–453

Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJA (2006) The PROSITE database. Nucleic Acids Res 34:D227–D230

Jeong DH, An S, Kang HG, Moon S, Han JJ, Park S, Lee HS, An K, An G (2002) T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol 130(4):1636–1644

Jeong DH, An S, Kang HG, Park GG, Kim SR, Sim J, Kim YO, Kim MK, Kim SR, Kim J, Shin M, Jung M, An G (2006) Generation of a flanking sequence-tag database for activation-tagging lines in *Japonica* rice. Plant J 45(1):123–132

Kohli A, Griffiths S, Palacios N, Twyman RM, Vain P, Laurie DA, Christou P (1999) Molecular characterisation of transforming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. Plant J 17(6):591–601

Kohli A, Twyman RM, Abranches R, Wegel E, Stoger E, Christou P (2003) Transgene integration, organization and interaction in plants. Plant Mol Biol 52:247–258

Koncz C, Martini N, Mayerhofer R, Koncz-Kalman Z, Korber H, Redei GP, Schell J (1989) High-frequency T-DNA-mediated gene tagging in plants. PNAS USA 86(21):8467–8471

Korf I, Flicek P, Duan D, Brent MR (2001) Integrating genomic homology into gene structure prediction. Bioinformatics 17(90001):S140–S148

Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P (2006) SMART 5: domains in the context of genomes and networks. Nucleic Acids Res 34:D257–D260. doi:10.1093/nar/gkj079

Lindsey K, Wei W, Clarke MC, Mcardle HF, Rooke LM, Topping JF (1993) Tagging genomic sequences that direct transgene expression by activation of a promoter trap in plants. Transgenic Res 2(1):33–47

Loc TN, Tinjuangjun P, Gatehouse AMR, Christou P, Gatehouse JA (2002) Linear transgene constructs lacking vector backbone sequences generate transgenic rice plants which accumulate higher levels of proteins conferring insect resistance. Mol Breed 9:231–244

Maqbool SB, Christou P (1999) Multiple traits of agronomic importance in indica rice plants: analysis of transgene integration patterns, expression levels and stability. Mol Breed 5:471–480

Mehlo L, Mazithulela G, Twyman RM, Boulton MI, Davies JW, Christou P (2000) Structural analysis of transgene rearrangements and effects on expression in transgenic maize plants generated by particle bombardment. Maydica 45:277–287

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002

Pawlowski WP, Somers DA (1996) Transgene inheritance in plants genetically engineered by microprojectile bombardment. Mol Biotechnol 6(1):17–30

Pawlowski WP, Somers DA (1998) Transgenic DNA integrated into the oat genome is frequently interspersed by host DNA. PNAS USA 95(21):12106–12110

Qin G, Kang D, Dong Y, Shen Y, Zhang L, Deng X, Zhang Y, Li S, Chen N, Niu W, Chen C, Liu P, Chen H, Li J, Ren Y, Gu H, Deng X, Qu LJ, Chen Z (2003) Obtaining and analysis of flanking sequences from T-DNA transformants of *Arabidopsis*. Plant Sci 165:941–949

Rang A, Linke B, Jansen B (2005) Detection of RNA variants transcribed from the transgene in Roundup Ready soybean. Eur Food Res Technol 220:438–443

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16(6): 276–277

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Salvo-Garrido H, Travella S, Bilham LJ, Harwood WA, Snape JW (2004) The distribution of transgene insertion sites in barley determined by physical and genetic mapping. Genetics 167(3):1371–1379

Scholte M, D′erfurth I, Rippa S, Mondy S, Cosson V, Durand P, Breda C, Trinh H, Rodriguez-Lorente I, Kondorosi E, Schultze M, Kondorosi A, Ratet P (2002) T-DNA tagging in the model legume *Medicago truncatula* allows efficient gene discovery. Mol Breed 10:203–215

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. PNAS 95(11):5857–5864

Sha Y, Li S, Pei Z, Luo L, Tian Y, He C (2004) Generation and flanking sequence analysis of a rice T-DNA tagged population. Theor Appl Genet 108(2):306–314

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

Somers DA, Makarevitch I (2004) Transgene integration in plants: poking or patching holes in promiscuous genomes? Curr Opin Biotechnol 15(2):126–131

Svitashev SK, Somers DA (2001) Genomic interspersions determine the size and complexity of transgene loci in transgenic plants produced by microprojectile bombardment. Genome 44:691–697

Svitashev SK, Somers DA (2002) Characterization of transgene loci in plants using FISH: a picture is worth a thousand words. Plant Cell Tissue Organ Cult 69:205–214

Svitashev SK, Ananiev E, Pawlowski WP, Somers DA (2000) Association of transgene integration sites with chromosome rearrangements in hexaploid oat. Theor Appl Genet 100:872–880

Szabados L, Kovacs I, Obershall A, Abrahám E, Kerekes I, Zsigmond L, Nagy R, Alvarado M, Krasovskaja I, Gál M, Berente A, Rédei GP, Haim AB, Koncz C (2002) Distribution of 1000 sequenced T-DNA tags in the *Arabidopsis* genome. Plant J 32:233–242

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673–4680

Tinland B (1996) The integration of T-DNA into plant genomes. Trends Plant Sci 1(6):178–184

Tzfira T, Li J, Lacroix B, Citovsky V (2004). *Agrobacterium* T-DNA integration: molecules and models. Trends Genet 20(8):375–383

Vain P, James VA, Worland B, Snape JW (2002) Transgene behaviour across two generations in a large random population of transgenic rice plants produced by particle bombardment. Theor Appl Genet 105:878–889

Vierstra RD (2003) The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. Trends Plant Sci 8:135–142

Wilson A, Latham J, Steinbrecher R (2006) Transformation-induced mutations in transgenic plants: analysis and biosafety implications. Biotechnol Genet Eng Rev 23:209–234

Windels P, Taverniers I, Depicker A, Van Bockstaele E, De Loose M (2001) Characterisation of the roundup ready soybean insert. Eur Food Res Technol 213:107–111

Zhang J, Guo D, Chang Y, You C, Li X, Dai X, Weng Q, Zhang J, Chen G, Li X, Liu H, Han B, Zhang Q, Wu C (2007) Non-random distribution of T-DNA insertions at various levels of the genome hierarchy as revealed by analyzing 13 804 T-DNA flanking sequences from an enhancer-trap mutant library. Plant J 49(5):947–959